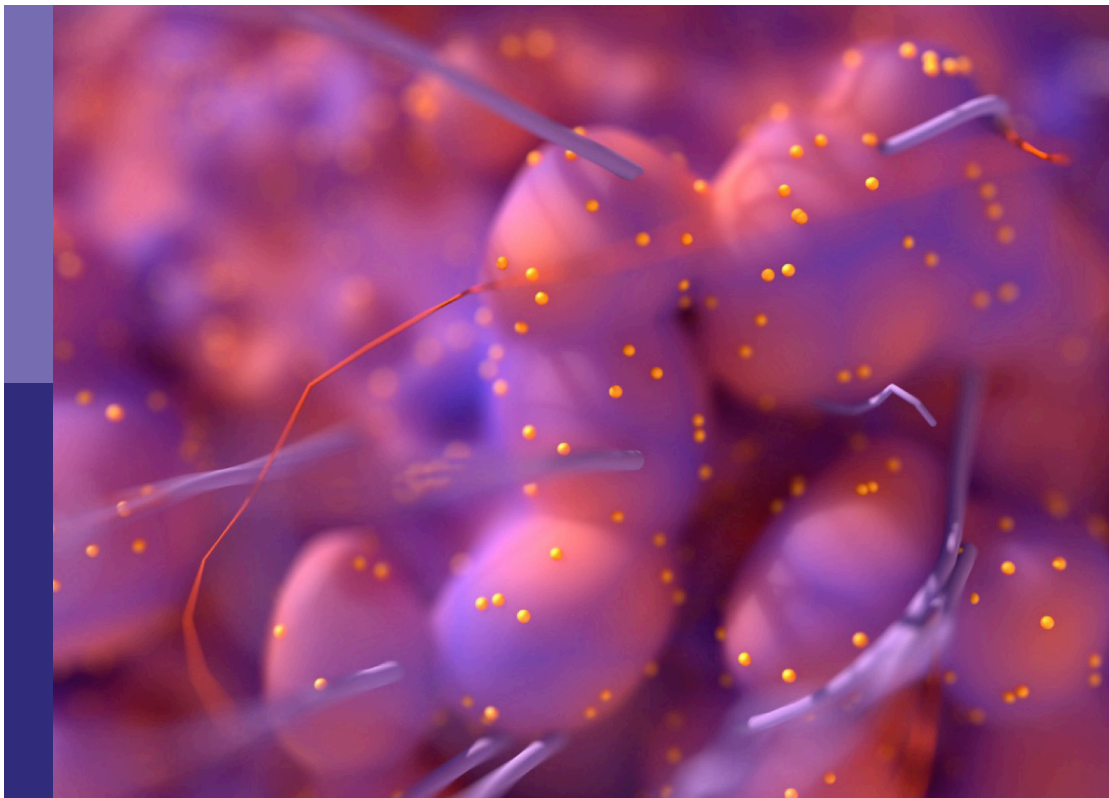
frontiers | Research Topics

Novel biomarkers for potential clinical applications in lung cancer

Edited by
Hongda Liu, Donghong Zhang, Ping Zhan, Qingqing Zhu and Miao Liu

Published in
Frontiers in Oncology



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 978-2-8325-5474-6
DOI 10.3389/978-2-8325-5474-6

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 for potential clinical applications in lung cancer

Topic editors

Hongda Liu — Nanjing Medical University, China
Donghong Zhang — Emory University, United States
Ping Zhan — Nanjing University School of Medicine, China
Qingqing Zhu — The First Affiliated Hospital of Soochow University, China
Miao Liu — Harvard Medical School, United States

Citation

Liu, H., Zhang, D., Zhan, P., Zhu, Q., Liu, M., eds. (2024). *Novel biomarkers for potential clinical applications in lung cancer*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-5474-6

Table of contents

08 Editorial: Novel Biomarkers for Potential Clinical Applications in Lung Cancer
Hongda Liu and Miao Liu

11 Development and validation of a screening model for lung cancer using machine learning: A large-scale, multi-center study of biomarkers in breath
Jing Li, Yuwei Zhang, Qing Chen, Zhenhua Pan, Jun Chen, Meixiu Sun, Junfeng Wang, Yingxin Li and Qing Ye

22 WDR5 is a prognostic biomarker of brain metastasis from non-small cell lung cancer
Zheng Li, Nan Liang, Na Wang, Yan Jia and Cui Tian

29 *MDFI* is a novel biomarker for poor prognosis in LUAD
Pengyu Chen, Zhen Quan, Xueyu Song, Zhaojia Gao and Kai Yuan

46 AI-based chest CT semantic segmentation algorithm enables semi-automated lung cancer surgery planning by recognizing anatomical variants of pulmonary vessels
Xiuyuan Chen, Hao Xu, Qingyi Qi, Chao Sun, Jian Jin, Heng Zhao, Xun Wang, Wenhan Weng, Shaodong Wang, Xizhao Sui, Zhenfan Wang, Chenyang Dai, Muyun Peng, Dawei Wang, Zenghao Hao, Yafen Huang, Xiang Wang, Liang Duan, Yuming Zhu, Nan Hong and Fan Yang

56 The diagnostic significance of blood-derived circRNAs in NSCLC: Systematic review and meta-analysis
Weijie Yu, Ruixue Liu, Zhoulin Miao, Liwei Zhang, Ilyar Sheyhidin and Julaiti Ainiwaer

66 Circulating tumor DNA detection in MRD assessment and diagnosis and treatment of non-small cell lung cancer
Xiaoxu Fang, Shaokun Yu, Yingying Jiang, Yan Xiang and Kaihua Lu

77 DNA methylation topology differentiates between normal and malignant in cell models, resected human tissues, and exfoliated sputum cells of lung epithelium
Jian Tajbakhsh, Fariborz Mortazavi and Nirdesh K. Gupta

87 A prognostic signature model for unveiling tumor progression in lung adenocarcinoma
Zijian Li, Tao Zeng, Chong Zhou, Yan Chen and Wu Yin

104 Prognostic signature of lipid metabolism associated LncRNAs predict prognosis and treatment of lung adenocarcinoma
Jie Zhao, Guangjian Li, Guangqiang Zhao, Wei Wang, Zhenghai Shen, Yantao Yang, Yunchao Huang and Lianhua Ye

123 Identification and validation of a novel phagocytosis regulators-related signature with potential prognostic and immunotherapeutic value in patients with lung adenocarcinoma
Jingyang Li, Qinyun Du, Jiayi Sun, Li Xiang and Shaohui Wang

141 Machine learning identifies characteristics molecules of cancer associated fibroblasts significantly correlated with the prognosis, immunotherapy response and immune microenvironment in lung adenocarcinoma
Qian Wang, Xunlang Zhang, Kangming Du, Xinhui Wu, Yexin Zhou, Diang Chen and Lin Zeng

157 Modeling extracellular matrix through histo-molecular gradient in NSCLC for clinical decisions
Camila Machado Baldavira, Tabatha Gutierrez Prieto, Juliana Machado-Rugolo, Jurandir Tomaz de Miranda, Lizandre Keren Ramos de Oliveira, Ana Paula Pereira Velosa, Walcy Rosolia Teodoro, Alexandre Ab'Saber, Teresa Takagaki and Vera Luiza Capelozzi

182 Characterizing microbiota and metabolomics analysis to identify candidate biomarkers in lung cancer
Bo Liu, Yige Li, Lijun Suo, Wei Zhang, Hongyun Cao, Ruicai Wang, Jiahui Luan, Xiaofeng Yu, Liang Dong, Wenjing Wang, Shiyang Xu, Shiyong Lu and Mei Shi

195 The combination of a seven-autoantibody panel with computed tomography scanning can enhance the diagnostic efficiency of non-small cell lung cancer
Tianyu He, Ziheng Wu, Pinghui Xia, Weidong Wang, Hua Sun, Li Yu, Wang Lv and Jian Hu

206 Investigation of M2 macrophage-related gene affecting patients prognosis and drug sensitivity in non-small cell lung cancer: Evidence from bioinformatic and experiments
Zhen Zeng, Jiachen Yu, Zhuo Yang, Kangming Du, Yuewei Chen and Lei Zhou

220 Minor histological components predict the recurrence of patients with resected stage I acinar- or papillary-predominant lung adenocarcinoma
Wei Liu, Qian Zhang, Tiantian Zhang, Li Li and Chunhua Xu

231 CCKAR is a biomarker for prognosis and asynchronous brain metastasis of non-small cell lung cancer
Nan Liang, Suohui Sun, Zheng Li, Tao Wu, Chunpu Zhang and Tao Xin

241 Postoperative ctDNA detection predicts relapse but has limited effects in guiding adjuvant therapy in resectable stage I NSCLC
Bolin Wang, Bing Zou, Shengnan Xu, Chao Zhao, Jinli Pei, Shijie Wang, Kunlong Zhao, Jinming Yu and Jie Liu

250 Integrated multi-dimensional deep neural network model improves prognosis prediction of advanced NSCLC patients receiving bevacizumab
Butuo Li, Linlin Yang, Chao Jiang, Yueyuan Yao, Haoqian Li, Shuping Cheng, Bing Zou, Bingjie Fan and Linlin Wang

261 Association between anlotinib trough plasma concentration and treatment outcomes in advanced non-small-cell lung cancer
Ling Chen, Hong Jiang, Jun-jie Rao, Liu-sheng Wang, Wei Yan, Jian Ye and Jiang Lou

272 Use of the prognostic nutrition index as a predictive biomarker in small-cell lung cancer patients undergoing immune checkpoint inhibitor treatment in the Chinese alpine region
Yunjiao Wu, Jing Yang, Xinyi Qiao, Yingjie Li, Rui Zhao, Tie Lin, Xiaoli Li and Meng Wang

283 CT radiomics model combined with clinical and radiographic features for discriminating peripheral small cell lung cancer from peripheral lung adenocarcinoma
Jingting Wang, Feiyang Zhong, Feng Xiao, Xinyang Dong, Yun Long, Tian Gan, Ting Li and Meiyan Liao

296 Preoperative levels of folate receptor-positive circulating tumor cells in different subtypes of early-stage lung adenocarcinoma: Predictive value for determining extent of surgical resection
Chao Zhou, Ran Zhao, Ruiying Zhao, Ansheng Wang and Wentao Li

307 Preoperative computed tomography-based tumoral radiomic features prediction for overall survival in resectable non-small cell lung cancer
Bo Peng, Kaiyu Wang, Ran Xu, Congying Guo, Tong Lu, Yongchao Li, Yiqiao Wang, Chenghao Wang, Xiaoyan Chang, Zhiping Shen, Jiaxin Shi, Chengyu Xu and Linyou Zhang

319 Clinicopathological characteristics and cancer-specific prognosis of primary pulmonary lymphoepithelioma-like carcinoma: a population study of the US SEER database and a Chinese hospital
Qun Zhang, Yuan Dai, Linling Jin, Shuangshuang Shi, Chang Liu, Rong Rong, Wenkui Sun, Shanlin Dai, Hui Kong and Weiping Xie

328 Diagnosis of primary lung cancer and benign pulmonary nodules: a comparison of the breath test and 18F-FDG PET-CT
Xiangxiang Ding, Guihu Lin, Peiyu Wang, Haibin Chen, Nan Li, Zhi Yang and Mantang Qiu

335 A nomogram combining thoracic CT and tumor markers to predict the malignant grade of pulmonary nodules ≤3 cm in diameter
Jianhao Qiu, Rongyang Li, Yukai Wang, Xiuyuan Ma, Chenghao Qu, Binyan Liu, Weiming Yue and Hui Tian

349 Therapeutic strategies to overcome EGFR mutations as acquired resistance mechanism in ALK-rearranged non-small-cell lung cancer: Case Reports
Lionel Michaux, Alexandre Perrier, Camille Mehlman, Hussa Alshehhi, Antonin Dubois, Roger Lacave, Florence Coulet, Jacques Cadranel and Vincent Fallet

354 A tumor mutational burden-derived immune computational framework selects sensitive immunotherapy/chemotherapy for lung adenocarcinoma populations with different prognoses
Wenlong Zhang, Chuzhong Wei, Fengyu Huang, Wencheng Huang, Xiaoxin Xu and Xiao Zhu

383 Deep learning-enhanced radiomics for histologic classification and grade stratification of stage IA lung adenocarcinoma: a multicenter study
Guotian Pei, Dawei Wang, Kunkun Sun, Yingshun Yang, Wen Tang, Yanfeng Sun, Siyuan Yin, Qiang Liu, Shuai Wang and Yuqing Huang

397 Generalization optimizing machine learning to improve CT scan radiomics and assess immune checkpoint inhibitors' response in non-small cell lung cancer: a multicenter cohort study
Marion Tonneau, Kim Phan, Venkata S. K. Manem, Cecile Low-Kam, Francis Dutil, Suzanne Kazandjian, Davy Vanderweyen, Justin Panasci, Julie Malo, François Coulombe, Andréanne Gagné, Arielle Elkrief, Wiam Belkaïd, Lisa Di Jorio, Michele Orain, Nicole Bouchard, Thierry Muanza, Frank J. Rybicki, Kam Kafi, David Huntsman, Philippe Joubert, Florent Chandelier and Bertrand Routy

407 Extracellular circulating miRNAs as potential non-invasive biomarkers in non-small cell lung cancer patients
Justyna Raczkowska, Agnieszka Bielska, Adam Krętowski and Magdalena Niemira

429 Clinical utility and diagnostic value of tumor-educated platelets in lung cancer: a systematic review and meta-analysis
Elvan Wiyarta, Darrin Ananda Nugraha, Muhammad Indera Ramadani, Gita Fajri Gustya, Muhammad Farrasy Ammar, Hana Dzakira Edwar, Nildza Kheirizzad, Mutiah Nurul Mukhlisah, Erlina Burhan and Elisna Syahruddin

440 Case Report: Durable therapy response to Osimertinib in rare EGFR Exon 18 mutated NSCLC
Michael Cekay, Philipp F. Arndt, Rio Dumitrascu, Rajkumar Savai, Andreas Braeuninger, Stefan Gattenloehner, Dagmar Steiner, Fritz Roller, Khodr Tello, Katja Hattar, Werner Seeger, Ulf Sibelius, Friedrich Grimminger and Bastian Eul

446 A proposed methodology for detecting the malignant potential of pulmonary nodules in sarcoma using computed tomographic imaging and artificial intelligence-based models
Esha Baidya Kayal, Shuvadeep Ganguly, Archana Sasi, Swetambri Sharma, Dheeksha DS, Manish Saini, Krithika Rangarajan, Devasenathipathy Kandasamy, Sameer Bakhshi and Amit Mehndiratta

459 In-depth single-cell and bulk-RNA sequencing developed a NETosis-related gene signature affects non-small-cell lung cancer prognosis and tumor microenvironment: results from over 3,000 patients
Liangyu Zhang, Xun Zhang, Maohao Guan, Fengqiang Yu and Fancai Lai

480 Case Report: Genetic profiling of small intestine metastasis from poorly differentiated non-small cell lung cancer: report of 2 cases and literature review of the past 5 years
Mengqin Wang, Gang Chen, Jiang Luo, Zhipeng Fan, Yuying Liu, Conghua Xie and Yan Gong

489 AI/ML advances in non-small cell lung cancer biomarker discovery
Minal Çalışkan and Koichi Tazaki

505 Plasma metabolomics reveals risk factors for lung adenocarcinoma
Mengjie Yu, Wei Wen, Yue Wang, Xia Shan, Xin Yi, Wei Zhu, Jiye Aa and Guangji Wang

522 Preoperative systemic immune-inflammation index-based nomogram for lung carcinoma following microwave ablation -a real world single center study
Jing Wang, Song-ping Cui, Qing Zhao, Ya Gao, Ying Ji, Yi Liu, Jin-bai Miao, Yi-li Fu and Bin Hu

frontiers | Frontiers in Oncology

TYPE Editorial
PUBLISHED 13 September 2024
DOI 10.3389/fonc.2024.1481799

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Lizza E. L. Hendriks,
Maastricht University Medical Centre,
Netherlands

*CORRESPONDENCE
Miao Liu
mliu0@bwh.harvard.edu

RECEIVED 16 August 2024
ACCEPTED 30 August 2024
PUBLISHED 13 September 2024

CITATION
Liu H and Liu M (2024) Editorial: Novel biomarkers for potential clinical applications in lung cancer.
Front. Oncol. 14:1481799.
doi: 10.3389/fonc.2024.1481799

COPYRIGHT
© 2024 Liu and Liu. 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: Novel biomarkers for potential clinical applications in lung cancer

Hongda Liu[1] and Miao Liu[2]*

[1]Department of General Surgery, First Affiliated Hospital, Nanjing Medical University, Nanjing, China, [2]Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, United States

KEYWORDS

lung cancer, prognosis, biomarker, clinical application, signal pathway

Editorial on the Research Topic
Novel biomarkers for potential clinical applications in lung cancer

Biomarkers play a critical role in modern medicine, significantly enhancing early diagnosis, prognostic evaluation, and personalized treatment across various diseases, including cardiovascular diseases, neurodegenerative disorders, infectious diseases, and cancers (1–4). Their application greatly improved diagnostic accuracy and increased opportunities for early intervention. In lung cancer, biomarkers such as EGFR, ALK, and ROS1 gene mutations were particularly important for guiding precision therapy and developing personalized treatment plans, resulting in significantly better clinical outcomes. The expanding use of biomarkers drove the evolution of medicine toward greater precision and personalization. The ongoing development of novel biomarkers for clinical applications in lung cancer further advanced this trend, improved treatment outcomes and survival rates, and made these biomarkers essential in lung cancer management.

Many biomarkers for various diseases were proven in numerous studies to be useful for predicting the prognosis of different diseases. Some studies have proved that the Systemic Immune-Inflammation Response Index (SIRI) reflected the prognostic significance in lung cancer patients undergoing microwave ablation (MWA). In a study involving 265 patients, the relationship between preoperative SIRI levels and long-term outcomes, such as overall survival (OS) and disease-free survival (DFS), was analyzed. The results indicated that higher SIRI levels were significantly associated with poorer long-term outcomes, including lower OS and DFS rates. Additionally, the study introduced nomograms based on SIRI and other independent factors to predict patient prognosis, demonstrating high accuracy. These findings highlighted the importance of SIRI as a prognostic tool, which might help clinicians identify high-risk patients and develop personalized treatment strategies for those undergoing MWA Wang et al. This also suggested the potential for cross-talk between signaling pathways across different diseases.

As new biomarkers continued to be identified, related biomarkers also emerged for some of the rarer types of non-small cell lung cancer (NSCLC). Recent research significantly enhanced our understanding of primary pulmonary lymphoepithelioma-like carcinoma (PPLELC), a rare subtype of NSCLC with distinct epidemiological and clinical features. Notably, PPLELC was more prevalent among Asian populations, a trend linked to

the Epstein-Barr virus (EBV), which was uncommon in other lung cancer types. This viral connection suggested PPLELC might share etiological features with other EBV-related malignancies, such as nasopharyngeal carcinoma. Studies also shed light on the genetic landscape of PPLELC, identifying specific mutations, like those in the TP53 and CYLD genes, which correlated with distinct clinical outcomes. These genetic markers could help stratify patients into different risk categories, facilitating more personalized treatment approaches—a crucial development given the rarity of PPLELC and the lack of established treatment protocols. The identification of these biomarkers not only aided in accurate diagnosis but also paved the way for targeted therapies. Tailoring treatments to the genetic and molecular profile of PPLELC could significantly improve patient outcomes, particularly for those in advanced stages where conventional therapies fell short. Integrating these biomarkers into clinical practice would mark a significant advancement in PPLELC management, shifting from a one-size-fits-all model to more individualized strategies Zhang et al.

While most biomarkers were protein-based, non-protein biomarkers became increasingly important in the prognosis and treatment planning for NSCLC patients. Among these, circulating tumor DNA (ctDNA) showed great potential as a biomarker for predicting disease recurrence in patients with resectable stage I NSCLC. ctDNA was proven to serve as a non-invasive tool to identify patients at high risk of relapse, which was crucial for early intervention and monitoring. However, the study also indicated that ctDNA's utility in guiding adjuvant therapy (ADT) decisions was limited, particularly in early-stage lung cancer. This suggested that while ctDNA held promise as a prognostic biomarker, further research and development were needed to improve its predictive accuracy and applicability in clinical decision-making, especially for treatment planning. The study contributed to the broader field of biomarker research by highlighting both the potential and the current limitations of ctDNA in lung cancer management, paving the way for more refined and effective use of biomarkers in the future Wang et al.

For predicting outcomes in lung cancer patients, various clinical and pathological factors were indispensable. In patients with resected stage I acinar- or papillary-predominant lung adenocarcinoma, the presence of non-predominant components, particularly micropapillary and solid components, was associated with poorer disease-free survival (DFS). When patients were categorized into risk subgroups based on these minor components, those with these high-grade patterns showed significantly worse prognoses. Conversely, the presence of lepidic components was associated with better outcomes Liu et al. Additionally, lymphovascular invasion, elevated carcinoembryonic antigen (CEA) levels, and a high platelet-to-lymphocyte ratio (PLR) were independent predictors of poor DFS. Comprehensive evaluation of these factors in the clinical management of early-stage lung adenocarcinoma was crucial, as they played a significant role in patient stratification and in guiding postoperative treatment strategies Bo Liu et al.

Some signaling pathways or mechanisms previously identified in tumors were also found to have new roles in lung cancer. A study, by analyzing tissue samples from patients, examined the expression of key proteins such as E-cadherin, beta-catenin, WNT proteins, and ECM components (such as chondroitin sulfate and various types of collagens) to explore the roles of epithelial-to-mesenchymal transition (EMT), the Wnt signaling pathway, and extracellular matrix (ECM) components in the progression and prognosis of NSCLC. The findings revealed that these markers were closely associated with tumor invasion and overall survival, highlighting their importance in NSCLC progression. Specifically, the study showed that WNT5A played a significant role in driving EMT and ECM remodeling, which were critical processes in tumor metastasis. Additionally, the study established correlations between these molecular markers and clinicopathological features, suggesting that they could serve as potential prognostic tools. The research enhanced our understanding of the molecular mechanisms driving NSCLC and indicated that these pathways and proteins might be viable targets for future therapeutic approaches Baldavira et al.

Novel biomarkers hold transformative potential in lung cancer management. They are central to the future of lung cancer care, playing a key role in everything from early detection and personalized treatment to ongoing disease monitoring. As research continues to advance, focusing on validating and implementing these novel biomarkers is crucial for driving personalized medicine in lung cancer. These efforts are essential for bringing the medical community closer to achieving more effective, individualized treatment strategies, ultimately improving patient outcomes and reducing the burden of lung cancer.

Author contributions

HL: Conceptualization, Writing – original draft. ML: Conceptualization, Data curation, Investigation, Supervision, Validation, Visualization, Writing – original draft, Writing – review & 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

1. Katrukha IA. Human cardiac troponin complex. Structure and functions. *Biochem (Mosc)*. (2013) 78:1447–65. doi: 10.1134/S0006297913130063

2. Ehrenberg AJ, Khatun A, Coomans E, Betts MJ, Capraro F, Thijssen EH, et al. Relevance of biomarkers across different neurodegenerative diseases. *Alzheimers Res Ther*. (2020) 12:56. doi: 10.1186/s13195-020-00601-w

3. Liu M, Chen Z, Dai MY, Yang JH, Chen XB, Chen D, et al. Lessons learned from early compassionate use of convalescent plasma on critically ill patients with Covid-19. *Transfusion*. (2020) 60:2210–6. doi: 10.1111/trf.15975

4. Kitamura T, Suzuki M, Nishimatsu H, Kurosaki T, Enomoto Y, Fukuhara H, et al. Final report on low-dose estramustine phosphate (EMP) monotherapy and very low-dose EMP therapy combined with LH-RH agonist for previously untreated advanced prostate cancer. *Aktuelle Urol*. (2010) 41 Suppl 1:S34–40. doi: 10.1055/s-0029-1224657

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 20 September 2022
DOI 10.3389/fonc.2022.975563

Check for updates

OPEN ACCESS

EDITED BY
Qingqing Zhu,
The First Affiliated Hospital of Soochow University, China

REVIEWED BY
Janaki Deepak,
University of Maryland, Baltimore, United States
Hongda Liu,
Nanjing Medical University, China

*CORRESPONDENCE
Meixiu Sun
sunmx@bme.cams.cn
Junfeng Wang
j.wang5@uu.nl

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

SPECIALTY SECTION
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

RECEIVED 22 June 2022
ACCEPTED 18 August 2022
PUBLISHED 20 September 2022

CITATION
Li J, Zhang Y, Chen Q, Pan Z, Chen J, Sun M, Wang J, Li Y and Ye Q (2022) Development and validation of a screening model for lung cancer using machine learning: A large-scale, multi-center study of biomarkers in breath. *Front. Oncol.* 12:975563. doi: 10.3389/fonc.2022.975563

COPYRIGHT
© 2022 Li, Zhang, Chen, Pan, Chen, Sun, Wang, Li and Ye. 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.

Development and validation of a screening model for lung cancer using machine learning: A large-scale, multi-center study of biomarkers in breath

Jing Li[1†], Yuwei Zhang[2†], Qing Chen[3], Zhenhua Pan[4], Jun Chen[4], Meixiu Sun[1*], Junfeng Wang[5*], Yingxin Li[1] and Qing Ye[2]

[1]Laser Medicine Laboratory, Institute of Biomedical Engineering, Chinese Academy of Medical Science and Peking Union Medical College, Tianjin, China, [2]Key Laboratory of Weak-Light Nonlinear Photonics, Ministry of Education, School of Physics and TEDA Applied Physics, Nankai University, Tianjin, China, [3]Departmentof Cardio-Pulmonary Function, National Clinical Research Center for Cancer, Cancer Institute and Hospital, Tianjin Medical University, Tianjin, China, [4]Tianjin Key Laboratory of Lung Cancer Metastasis and Tumor Microenvironment, Tianjin Lung Cancer Institute, Tianjin Medical University General Hospital, Tianjin, China, [5]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands

Objectives: Lung cancer (LC) is the largest single cause of death from cancer worldwide, and the lack of effective screening methods for early detection currently results in unsatisfactory curative treatments. We herein aimed to use breath analysis, a noninvasive and very simple method, to identify and validate biomarkers in breath for the screening of lung cancer.

Materials and methods: We enrolled a total of 2308 participants from two centers for online breath analyses using proton transfer reaction time-of-flight mass spectrometry (PTR-TOF-MS). The derivation cohort included 1007 patients with primary LC and 1036 healthy controls, and the external validation cohort included 158 LC patients and 107 healthy controls. We used eXtreme Gradient Boosting (XGBoost) to create a panel of predictive features and derived a prediction model to identify LC. The optimal number of features was determined by the greatest area under the receiver-operating characteristic (ROC) curve (AUC).

Results: Six features were defined as a breath-biomarkers panel for the detection of LC. In the training dataset, the model had an AUC of 0.963 (95% CI, 0.941–0.982), and a sensitivity of 87.1% and specificity of 93.5% at a positivity threshold of 0.5. Our model was tested on the independent validation dataset and achieved an AUC of 0.771 (0.718–0.823), and sensitivity of 67.7% and specificity of 73.0%.

Conclusion: Our results suggested that breath analysis may serve as a valid method in screening lung cancer in a borderline population prior to hospital

visits. Although our breath-biomarker panel is noninvasive, quick, and simple to use, it will require further calibration and validation in a prospective study within a primary care setting.

KEYWORDS

breath analysis, lung cancer, machine learning, PTR-TOF-MS, screening

Introduction

Lung cancer (LC) is the largest single cause of death from cancer worldwide (1), and the five-year net survival is in the range of 10–20% for most countries (2). However, early-stage LC is curable, with an overall five-year survival rate of 80% (3). There is therefore an urgency to the development of efficient approaches in the early detection of LC.

Low-dose computed tomography (CT) scanning for the population at high risk is commonly used in LC screening. To reduce cancer mortality, the United States (U.S.) Preventive Services Task Force recommended expanding LC screening to younger individuals and low-intensity smokers (4). However, high false-positive rates, over-diagnosis, limits to applicable coverage, and cumulative radiation exposure remain primary concerns with this type of screening modality (5).

Breath analysis (BA) provides an attractive option (6–8), because the growth of cancer cells is strongly linked to key metabolic pathways that produce detectable amounts of volatile organic compounds (VOCs) in exhaled breath (9). Previous studies have shown that BA can differentiate between LC patients and healthy controls, with an overall accuracy of 69.4% to 100% (10). However, the lack of reproducibility for breath biomarkers among different studies restricts the further implementation of these biomarkers in clinical practice (11). This lack of replicability is primarily because most breath biomarkers were recognized from small pilot studies (the largest study had a sample size of 193), and they lacked independent validation to evaluate their test accuracy (10, 12). Owing to the heterogeneity and variety of physiologic and clinical backgrounds of patients, this deficiency in large-scale samples hinders the development, validation, and implementation of appropriate biomarkers.

Gas chromatography in combination with mass spectrometry and electronic noses are widely used for the investigation of breath biomarkers in LC (10–12). Breath VOCs comprise a very complex matrix that contains a large variety of VOCs at trace amounts (ppbv to pptv) (13, 14). The major flaw of electronic noses in screening the reliable biomarkers of LC is the inferior provision of quantitative results with respect to unknown substances (15, 16). Mass-spectrometric techniques are particularly well suited for biomarker investigations because they offer the possibility of detecting a large variety of compounds of interest with high sensitivity and high accuracy (17–19). However, the commonly encountered issues during conventional gas chromatography in combination with mass spectrometry-based breath-profiling analysis in a large-scale study are the complicated sample-preparation procedure and time-consuming test processes. Direct mass spectrometry—such as selected ion flow tube mass spectrometry and proton transfer reaction-mass spectrometry—are sufficiently sensitive and rapid to allow real-time breath analysis (20, 21). Proton transfer reaction time-of-flight mass spectrometry (PTR-TOF-MS) combines time-of-flight mass spectrometry with a proton transfer flow-drift tube reactor, and provides a high mass-resolving power that enables the separation of isobaric molecules; this allows the measurement of a complete mass spectrum within a fraction of a second (22, 23). Compared with offline sampling such as sample collection into bags or onto traps, online sampling is beneficial in reducing artifacts of sample degradation during collection, storage, and handling or the introduction of impurities. To address the challenges inherent to a large-scale breath study, we herein employed a real-time, sensitive, and reliable analytical instrument, the PTR-TOF-MS, in combination with buffered end-tidal (BET) online sampling (24).

Machine learning-based prediction models have shown promising and even superior predictive performance compared with conventional statistical techniques (25), and the advantages of machine learning in large-scale data processing and its non-linear fitting capability make it particularly useful in resolving medical complications. Therefore, we herein incorporated machine-learning algorithms into the pipeline of LC screening of an individual based on breath-component analysis.

Recent efforts have been undertaken to identify and internally validate LC biomarkers using a relatively small

Abbreviations: AUC, area under the receiver-operating characteristic curve; BA, Breath analysis; BET, buffered end-tidal; EPV, events per variable; IQR, inter-quartile range; LC, lung cancer; PTR-TOF MS, proton transfer reaction-time of flight-mass spectrometry; XGBoost, eXtreme Gradient Boosting; VOCs, volatile organic compounds.

dataset (139 patients with lung cancer and 289 healthy adults), and the results suggested that breath testing may constitute a reliable approach for the detection of LC (26). The goal of the current study, then, was to define and externally validate breath testing for LC screening using breath data from a large number of samples from multiple centers. We therefore exploited a breath test that combined PTR-TOF-MS and a machine-learning algorithm to identify and validate the clinical applicability of our novel biomarkers.

Materials and methods

This study was conducted and reported in accordance with TRIPOD (27), the guideline for clinical-prediction model studies; and STARD-2015 (28), the reporting guideline for diagnostic test studies. Both checklists were completed and are provided in e-Tables 1 and 2 in the Supplement.

Study design and data collection

The dataset comprising biomarker discovery and model development was collected prospectively using a case-control design. Consecutive patients suspected to have LC were prepared for surgery or bronchoscopy in the Pulmonary Oncology Department of the Cancer Institute and Hospital, Tianjin Medical University, and were recruited between February of 2019 and January of 2020. Healthy subjects were enrolled after undergoing health checkups at the Cancer Institute and Hospital, Tianjin Medical University, from April 2019 to May 2019.

The validation dataset was also prospectively collected in a case-control design. Suspected LC patients who were prepared for surgery or bronchoscopy in the Department of Pulmonary Oncological Surgery were recruited from Tianjin Medical University General Hospital between October 2020 and June 2021, and healthy subjects (controls) were recruited from hospital staff of the General Hospital of Tianjin Medical University in November of 2020 and March and December of 2021.

The exclusion criteria for LC patients were those under 18 years of age; patients who showed a history of cancer or a synchronous cancer; or had undergone chemotherapy (with anticancer drugs), immunotherapy, hormonal therapy, or radiotherapy. The exclusion criteria for healthy controls were those under 18 years of age, undergoing pregnancy, individuals with a self-reported history of pulmonary disease, and those manifesting pulmonary nodules confirmed by CT images.

Information regarding a history of lung disease, medication use, fasting, and tobacco smoking was obtained through self-reporting. A history of lung disease was designated as an affirmative response to the question "Have you ever had lung disease?"; use of medications was defined as taking any type of drug (including sprays, pills, capsules, and decoctions) in the previous half-month; an empty stomach was characterized as an affirmative answer to the question "Have you eaten breakfast already?"; and smoking status was delineated as never smoking, being an ex-smoker, of currently smoking. Smoking denoted at least one cigarette every day, which continued to or averaged over six or more months; and an ex-smoker quit smoking four or more months prior to sampling. We determined the amount of smoking by counting the number of cigarettes smoked per day.

Calculation of sample size

In concert with the recommendations of TRIPOD and PROBAST regarding sample-size calculation, we determined the sample size needed for developing and validating the respective models. The sample size for model development was ascertained with the method recently proposed by Riley et al. (29), as well as using 10 events per variable (EPV) as a rule-of-thumb. We set Cox-Snell's adjusted R (2) to 0.1 and the desired shrinkage equal to 0.9 as recommended. Since machine-learning models may require additional data relative to fitting a statistical model, we added a conservative factor of 10%. Based on our calculations, the desired sample size for model development was 1868, with a conservative adjustment to 2055 (i.e., 1868*1.1). Cases and controls were collected at a ratio of approximately 1:1, and with 22 candidate variables our EPV was 47 (i.e., 2055*0.5/22), which was far larger than that required using rule-of-thumb.

We computed the sample size for model validation according to a requirement of at least 100 patients in both groups, with and without the outcome of interest (i.e., primary LC).

Outcome and reference standards

We obtained samples of lung-tissue lesions from LC patients by bronchoscopy or surgery for pathologic examination, and clinical status (including stage and type of LC) was confirmed by pathologic diagnosis within one month after sampling. The disease status of healthy controls was determined by physical examination; i.e., individuals younger than 45 years of age underwent lung X-rays while individuals older than 45 underwent either lung X-rays or lung CT scans.

PTR–TOF-MS analysis

PTR-TOF-MS (PTR-TOF-MS 1000, Ionicon Analytik GmbH, Innsbruck, Austria) offers quantitative analysis of the entire mass range (1–10,000 amu) within split-seconds and with an ultra-low detection limit (LoD<10 pptv) and high resolution (>2000 m/△m). The BET-sampling system (Ionicon Analytik GmbH, Innsbruck, Austria) also affords the two distinct

advantages of collecting the end-tidal fraction of exhaled breath gas and maintaining a normal breathing pattern for test subject after one exhalation. This system allows the measurement of endogenous compounds originating from the alveolar blood-gas exchange, and reduces the risk of hyperventilation.

Our procedure was as follows. The test subject exhaled directly into the buffer tube of the BET-sampling system equipped with a disposable and sterile mouthpiece (Polypropylene; Art. Nr. 31-30-0022, Germany) and the procedure was repeated three times. The buffer tube was maintained at 80°C by a heating system so as to eliminate the effect of condensation of humidified breath gas, and the collected gas was introduced into the ionization section by the inlet line of the instrument. The ionized molecules were then separated by their mass-to-charge ratio (m/z) and subsequently detected. The pressure and temperature in the drift tube were 2.3 mbar and 70° C, respectively, with an electric drift field of 600 V. A total of 318 features (m/z) were thus extracted from the acquired spectrum of each exhaled breath sample.

Data analysis

Candidate predictors

Raw PTR-MS spectra were acquired using the data acquisition software IoniTOF30. Data were preprocessed to extract all features that were organic compounds and expiratory concentrations that were higher than the respective inspiratory concentrations. Twenty-two features of endogenous VOCs that were ultimately determined for all test subjects in the discovery dataset were chosen as candidate features.

Feature selection

Before feature selection we first standardized our dataset with an estimated mean and variance from the training set (standardization of external validation set was also based on mean and variance from the training set), and to further reduce the candidate-feature set, we calculated the Spearman correlation between each pair of features. We then randomly removed one of the features from a pair with a correlation greater than 0.99, and this resulted in a final candidate set of 14 features. Finally, we ran an eXtreme Gradient Boosting (XGBoost) classifier (30) and ranked the remaining 14 features using the inherent feature importance from the classifier.

Model selection

The XGBoost models were iteratively trained with the feature subset ranked at the top, starting with the most important feature and with one feature added each time. At completion, we respectively compared 14 models with 1 to 14 features; and model performance was evaluated *via* a 10-fold cross validation. The area under the receiver-operating characteristic (ROC) curve (AUC) averaged over 10 validation results was used as our criterion for model selection. To achieve a balance between model performance and simplicity, we set the performance-reduction tolerance at 1%, indicating that the minimal model performance requirement was 99% of the highest AUC among the 14 models. We then chose the final model that met both the minimal performance requirement and that possessed the fewest features.

Statistical analysis of model performance

Continuous variables are expressed as median and inter-quartile ranges (IQRs), and categorical variables are expressed as counts and percentages. The discrimination of the predictive model was assessed using ROC curves and AUCs, while calibration was assessed with the calibration curve.

We also calculated diagnostic performance measures—including sensitivity, specificity, precision, recall, and accuracy—based upon confusion matrix with a pre-specified positivity threshold of 0.5.

The implementations of our feature engineering process, predictive model development, and validation were based on Python Scikit-learn 0.22.1 (31).

Ethics approval

The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committees of the Cancer Institute and Hospital, Tianjin Medical University; and Tianjin Medical University General Hospital. The present trial was registered with the Institutional Review Board of the Chinese Clinical Trial Registry (registration number: chiCTR1900023659), and all methods were conducted in accordance with relevant guidelines and regulations. Informed consent was obtained from all participants.

Results

Description of the derivation and validation datasets

The flow chart for patient recruitment is shown in Figure 1. For model derivation, we recruited a total of 2043 participants, including 1007 patients with primary LC and 1036 healthy controls from the Cancer Institute and Hospital, Tianjin Medical University. Mean age of the 1007 patients with primary LC (559 males, 55.51%) was 61 years (age range, 21–81 years), and the most-common smoking status of the patients was non-smoker—accounting for 45.08%. The principal tumor cell types were adenocarcinoma (62.36%), squamous cell carcinoma (15.89%), and small-cell carcinoma (9.04%). At

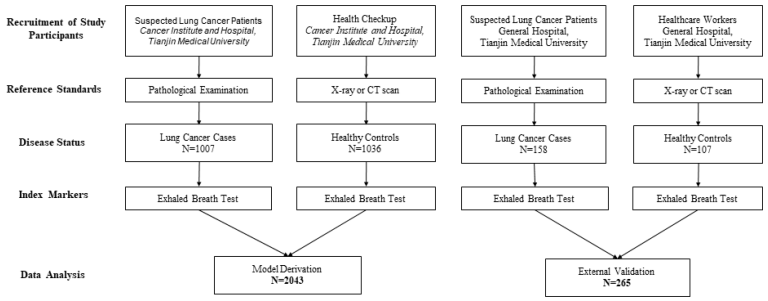


FIGURE 1
Flow chart for patient recruitment in model development and validation cohorts.

the time of LC diagnosis, we noted 273 patients with stage I disease (27.11%), 121 with stage II (12.02%), 128 patients with stage IIIIV (12.71%), and 170 patients with stage IV (16.88%). Of the enrolled patients, 387 (38.43%) reported that they were fasting at the time of breath sampling. The mean age of the 1036 healthy controls (536 males, 51.74%) was 45 years (age range, 22–90 years), 776 of the subjects were non-smokers (74.91%), and 857 (82.72%) were fasting at the time of breath sampling.

The independent-validation cohort comprised 265 subjects (including 158 patients with primary LC and 107 healthy controls) who came from the General Hospital of Tianjin Medical University. Mean age of the 158 patients with primary LC (63 males, 39.87%) was 63 years (age range, 33–78 years), and the most-common smoking status was smoker—accounting for 49.37%. At the time of LC diagnosis, we noted 133 patients with adenocarcinoma (84.18%), 15 with squamous cell carcinoma (9.49%), and four with small-cell lung cancer (2.53%). Of the enrolled patients, 17 (10.76%) reported that they were fasting at the time of breath sampling. The mean age of the 107 healthy controls (40 males, 37.38%) was 30 years (age range, 19–74 years), 94 of the subjects were non-smokers (87.85%), and nine (8.41%) were fasting at the time of breath sampling (the baseline characteristics of these individuals are shown in Table 1).

Development and validation of the prediction model

Feature selection and importance ranking

Candidate features were first selected by their pairwise correlations, and 14 features were retained for subsequent data analysis. The features selected and their distributions are depicted in Figure 2 (ranked by their importance), where green represents LC patients and red represents healthy subjects.

The model achieving the greatest AUC included the top 12 features, and it yielded an AUC (averaged across 10-fold cross-validation) of 0.970. Thus, the minimal performance requirement was 0.961 (i.e., 99%*0.970). The model with the fewest features that met this requirement was selected as the final model, and it included the top six features, with an AUC of 0.963. (Figure 3) The features included in the ultimate model were 'm77.0597 ([$C_3H_8O_2$] H+)', 'm95.0491 ([C_6H_6O] H+)', 'm33.0335 ([CH_4O] H+)', 'm59.0491 ([C_3H_6O] H+)', 'm137.0709 ([$C_7H_8N_2O$] H+)', and 'm68.0495 ([C_4H_5N] H+)'.

The final model was internally validated with 10-fold validation and externally validated with the independent-validation dataset, and the AUCs were 0.963 and 0.771 for internal and external validations, respectively (Figure 4). The calibration curves showed acceptable alignment for both the derivation and validation datasets (Figure 5).

With a predefined positivity threshold of 0.5, the sensitivity and specificity of the final model were 87.1% and 93.5% for the derivation data, respectively; and 67.7% and 73.0% for the validation data, respectively. The confusion matrix (Table 2) and diagnostic-accuracy measures for both the derivation and validation datasets are provided in Table 3.

Sensitivity and subgroup analyses on internal and external validation

In the sensitivity analyses, we evaluated the performance of the developed model in distinguish different histological subtypes from healthy controls, and also evaluated the model performance in cancer staging. In the subgroup analyses, we evaluated the model performance in gender groups (male and female), age groups (<45 and 45 years old), fasting groups, and smoking groups (evaluating the model performance in ever-smoking and non-smoking groups separately). The model performance is consistent with the overall results in the sensitivity and subgroup analysis. All these results are provided in e-Tables 3, 4 and 5 in the Supplementary.

10.3389/fonc.2022.975563

TABLE 1 Baseline characteristics of the individuals included in the study.

	Derivation		Validation	
	Lung cancer (n = 1007)	Healthy control (n = 1036)	Lung cancer (n = 158)	Healthy control (n = 107)
Male (%)	559 (55.51%)	536 (51.74%)	63 (39.87%)	40 (37.38%)
Age (IQR) [Range]	61 (54, 66) [21-81]	45 (35, 58) [22-90]	63 (58, 69) [33-78]	30 (24, 43)[19-74]
Smoking (%)				
Smokers	382 (37.94%)	209 (20.17%)	78 (49.37%)	12 (11.22%)
Ex-smokers	171 (16.98%)	51 (4.92%)	24 (15.19%)	1 (0.93%)
Non-smokers	454 (45.08%)	776 (74.91%)	56 (35.44%)	94 (87.85%)
BMI(IQR)	24.03 (22.04, 26.30)	24.06 (21.97, 26.30)	23.96 (21.64, 25.92)	22.48 (20.28, 25.45)
Fasting (%)	387 (38.43%)	857 (82.72%)	17 (10.76%)	9 (8.41%)
Adenocarcinoma (%)	628 (62.36%)	NA	133 (84.18%)	NA
Squamous cell carcinoma (%)	160 (15.89%)	NA	15 (9.49%)	NA
Small-cell lung cancer (%)	91 (9.04%)	NA	4 (2.53%)	NA
Missing	128 (12.71%)		6 (3.80%)	
Stage (%)				
0	31 (3.08%)	NA	–	NA
I	273 (27.11%)	NA	–	NA
II	121 (12.02%)	NA	–	NA
III	128 (12.71%)	NA	–	NA
IV	170 (16.88%)	NA	–	NA
Missing	284 (28.20%)			

NA, Not applicable for healthy control group.

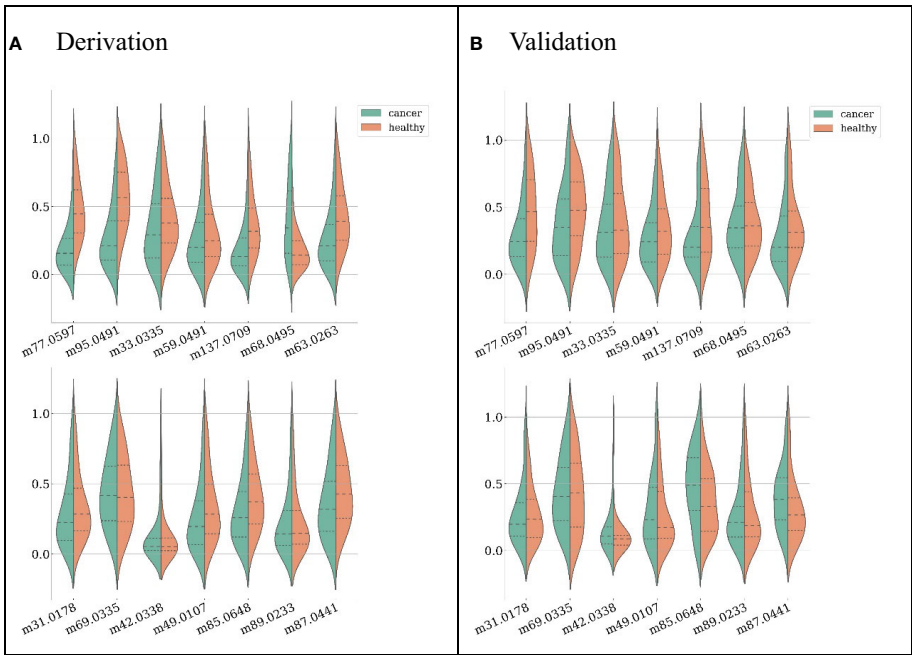


FIGURE 2
Feature distributions on the derivation dataset **(A)** and validation dataset **(B)**, ranked by their importance (the first feature from left on the first row is the most important). For each feature, both distributions from LC patients (green) and healthy subjects (red) are shown.

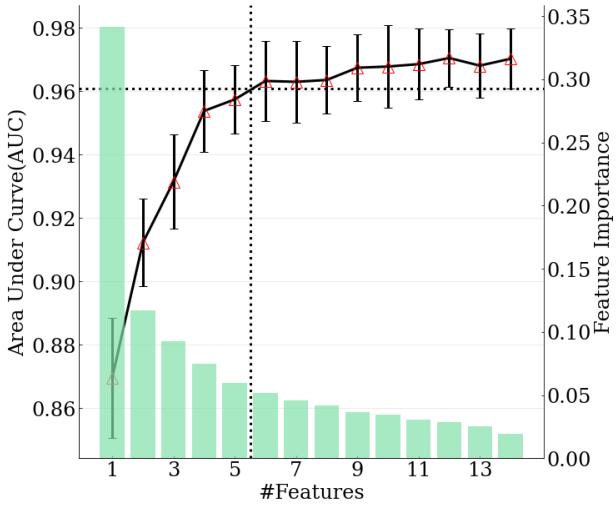


FIGURE 3
Relation between number of features selected in the model and model performance. Green bars correspond to feature importance. Black solid line corresponds to AUC calculated with top 1-14 features. Black dotted lines demonstrate the number of features selected when achieving 99% of maximum AUC.

Discussion

In the nearly 50 years since Linus Pauling first demonstrated the presence of VOCs in human breath, investigators have published over 50 reports showing a strong biologic rationale for using breath biomarkers in the detection of LC. Nevertheless, prior to reaching the clinical setting, this promising approach still faces the challenges of the inconsistent biomarkers exhibited in previous studies: these include limited study cohorts, single study sites, and a lack of validation.

In the current large-scale, multi-center biomarker study, efforts were made to define more reliable breath biomarkers. We first recruited a large cohort of 2043 subjects and analyzed their breaths to develop a predictive panel using machine learning so as to reduce the influence of patient-related individual differences. We thus designed a real-time, sensitive,

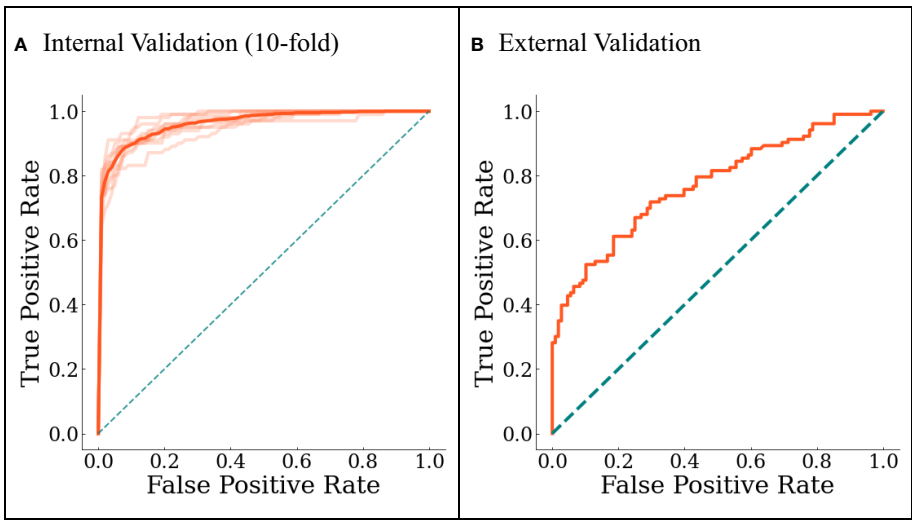


FIGURE 4
The ROC curves for **(A)** internal 10-fold cross-validation and **(B)** external validation. In **(A)**, darker line represents the averaged results.

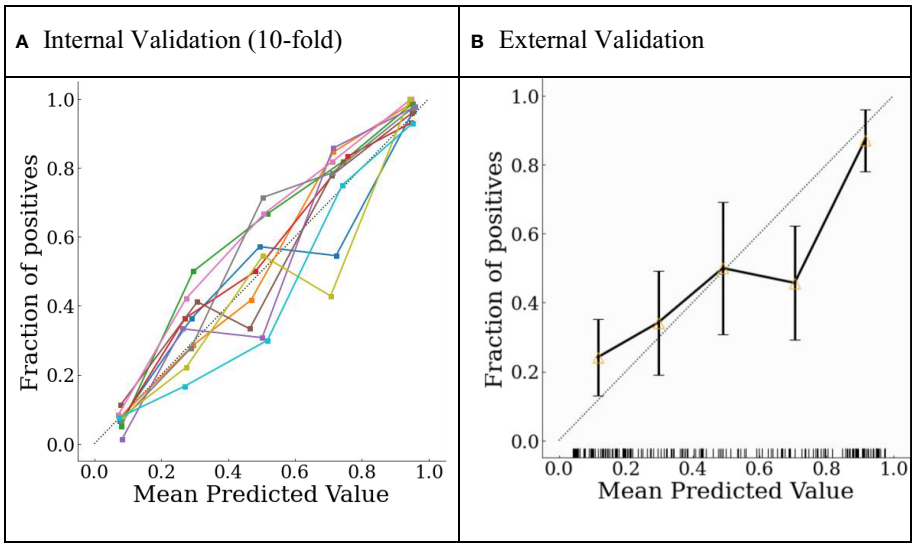


FIGURE 5
Probability calibration plots for **(A)** internal 10-fold cross-validation and **(B)** external validation.

TABLE 2 Confusion matrix of the derivation and validation datasets.

Derivation data		**Lung cancer diagnosed by the current gold standard**		**Total**
		Present	**Absent**	
Model prediction	Positive	True positive = 877	False positive = 67	944
	Negative	False negative= 130	True negative = 969	1099
Total		1007	1036	2043
Validation data		Lung cancer diagnosed by the current gold standard		Total
		Present	Absent	
Model prediction	Positive	True positive = 107	False positive = 29	136
	Negative	False negative= 51	True negative = 78	129
Total		158	107	265

and reliable instrument coupled with BET sampling for the online collection of alveolar air to reduce the influence of sampling and environmental confounders. Through the exploitation of the machine-learning algorithm XGboost, a panel of six features was defined that revealed an AUC of 0.963, a sensitivity of 87.1%, and a specificity of 93.5%. Second, our panel was validated from a dataset measured at a different hospital, and which achieved an AUC of 0.771. These data of large-scale breath testing and machine learning exhibited the potential to overcome the methodologic challenges of breath tests in the detection of LC, and showed that our metabolic breath panel generated a strong potential for application as a screening tool in clinical practice for the detection of LC.

TABLE 3 Model performance of diagnostic accuracy in the derivation and validation datasets.

	Training (95% CI)	**Validation (95% CI)**
AUC	0.963 (0.941–0.982)	0.771 (0.718–0.823)
Accuracy	0.904 (0.888–0.925)	0.704 (0.654–0.753)
Sensitivity/Recall	0.871 (0.822–0.926)	0.677 (0.598–0.750)
Specificity	0.935 (0.884–0.967)	0.730 (0.660–0.798)
PPV/Precision	0.930 (0.883–0.961)	0.706 (0.631–0.779)
F-score	0.899 (0.880–0.924)	0.690 (0.625–0.750)

The PTR-TOF-MS is one of the most powerful techniques for online monitoring of trace VOCs, it can detect mass spectrum peaks with m/z less than 500 and simultaneously achieved accurate concentration of these features, while the peak intensity as a substitute indicator for concentration was used in most mass spectrum techniques. In addition, it has a low detection limit of 10 ppt and a wide detection linear range of 5 orders of magnitude. These characteristics make PTR-TOF-MS hold potentially great value for model development of cancer

identification. In this study, we developed the model based on the features extracted from PTR-TOFMS data. The identified VOCs based on m/z was 1,3-Propanediol, phenol, methanol, acetone, m-aminobenzamide and butene nitrile according to the library established based on PTR-TOF-MS. Alcohols and ketones are most commonly detected compounds as biomarkers as lung cancer (32). The formation of some alcohols has been repeatedly reported in the literature to be associated with the growth and metastasis of cancer, suggesting the existence of significance of alcohols in indicating lung cancer (33). Acetone can be produced from the spontaneous decarboxylation of acetoacetate, and it has been used as a biomarker for activation of ketone metabolism, which suggesting that metabolism of ketone bodies might be important for lung cancer cells. It has been confirmed that when there is cancer cell activity in the body, abnormal cell proliferation triggers a stress response that causes increased secretion of adrenocorticotropic hormones (monohydroxy phenolics) in the body (34), suggesting that phenolic metabolites may have an indicator role for lung cancer. In addition, m-aminobenzamide and butene nitrile have not been reported in the literature.

There were, however, some limitations to our study. First, we employed a population-based, case-control design for recruiting participants, individuals with pulmonary nodules confirmed by CT images were excluded from healthy controls, which may lead to overestimate of the predictor–outcome association as well as the model performance. Although a phase-3 analysis (such as model development) was executed, our study can only be viewed as a phase-2 study for biomarker exploration according to the definition provided by Pepe et al. (35). We thus plan a follow-up phase-3 study with nested case-controls within a population cohort to confirm the performance of the proposed markers, and to validate the model in a real-world setting. Second, only 304 cases (30.19% of all LC cases) were diagnosed with early-stage LC in the derivation dataset (stage 0 + stage I), which can differ from the screening setting. Third, the two centers involved in this study were from the same city, which may have limited the robustness of the panel in general clinical practice. Finally, whether the breath panel established in this study excluded the interference of other lung diseases in otherwise healthy individuals requires further verification in the future. Alternatively, a new expiratory database that includes other lung diseases could be implemented. Since the breath panel was selected based upon machine learning, we also propose that our analysis will emerge as more robust when additional participating centers and individuals are recruited to the study.

In summary, we identified a breath-biomarker panel consisting of six features that was defined and validated as an effective tool for the detection of LC in a multi-center phase-2 study. The biomarker panel was applied to discriminate patients with LC from a healthy population (without LC), and whose screening performance was externally validated. This breath panel showed a robust potential for LC screening in clinical practice. However, additional prospective data are needed within a cohort-study design in a primary care setting where the prevalence of LC would be far lower, so as to confirm the validity of our findings and to establish the optimal predictive model.

Data availability statement

These data are not publicly available due to concerns for subject privacy. Requests to access the data can be directed to the corresponding author.

Ethics statement

The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committees of the Cancer Institute and Hospital, Tianjin Medical University and Tianjin Medical University General Hospital. The present trial was registered with the Institutional Review Board of the Chinese Clinical Trial Registry (registration number: chiCTR1900023659), and all methods were conducted in accordance with relevant guidelines and regulations. Informed consent was obtained from all participants. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

MS took responsibility for (i.e., is the guarantor of) the content of the manuscript, including the data and analysis. JL performed the experiments and assisted with manuscript writing. YZ and QY performed computational data analyses. QC provided samples for model derivation. ZP and JC provided samples for the validation cohort. MS conceived and designed the study, established collaborations, wrote the manuscript, and allocated funding for this study. JW provided critical technical assistance and consultation, and reviewed the manuscript. YL established collaborations and provided scientific direction. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Chinese Academy of Medical Sciences Initiative for Innovative Medicine (2018-I2M-AI-012).

Acknowledgments

We wish to thank all of the participants included in this study.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.975563/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. Allemani C, Matsuda T, Carlo VD, Harewood R, Matz M, Nikšić M, et al. Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37513025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet* (2018) 391(10125):1023–75. doi: 10.1016/S0140-6736(17)33326-3

3. Chu GCW, Lazare K, Sullivan F. Serum and blood based biomarkers for lung cancer screening: a systematic review. *BMC Cancer* (2018) 18(1):181. doi: 10.1186/s12885-018-4024-3

4. Melzer AC, Wilt TJ. Expanded access to lung cancer screening–implementing wisely to optimize health. *JAMA Netw Open* (2021) 4(3):e210275. doi: 10.1001/jamanetworkopen.2021.0275

5. Wang X, Zhi X, Yang Z, Tian H, Li Y, Li M, et al. A novel serum based biomarker panel has complementary ability to preclude presence of early lung cancer for low dose CT (LDCT). *Oncotarget* (2017) 8(28):45345–55. doi: 10.18632/oncotarget.17477

6. Chen T, Liu T, Li T, Zhao H, Chen Q. Exhaled breath analysis in disease detection. *Clin Chim Acta* (2021) 515:61–72. doi: 10.1016/j.cca.2020.12.036

7. Gaugg MT, Nussbaumer-Ochsner Y, Bregy L, Engler A, Stebler N, Gaisl T, et al. Real-time breath analysis reveals specific metabolic signatures of COPD exacerbations. *Chest* (2019) 156(2):269–76. doi: 10.1016/j.chest.2018.12.023

8. Yoon JW, Lee JH. Toward breath analysis on a chip for disease diagnosis using semiconductor-based chemiresistors: recent progress and future perspectives. *Lab Chip* (2017) 17(21):3537–57. doi: 10.1039/C7LC00810D

9. Pavlova NN, Thompson CB. The emerging hallmarks of cancer metabolism. *Cell Metab* (2016) 23(1):27–47. doi: 10.1016/j.cmet.2015.12.006

10. Antoniou SX, Gaude E, Ruparel M, Schee MPVD, Janes SM, Rintoul RC. The potential of breath analysis to improve outcome for patients with lung cancer. *J Breath R* (2019) 13(3):034002. doi: 10.1088/1752-7163/ab0bee

11. Saalberg Y, Wolff M. VOC breath biomarkers in lung cancer. *Clin Chim Acta* (2016) 459:5–9. doi: 10.1016/j.cca.2016.05.013

12. Hanna GB, Boshier PR, Markar SR, Romano A. Accuracy and methodologic challenges of volatile organic compound–based exhaled breath tests for cancer diagnosis: a systematic review and meta-analysis. *JAMA Oncol* (2019) 5(1):e182815. doi: 10.1001/jamaoncol.2018.2815

13. Liang Q, Chan YC, Changala PB, Nesbitt DJ, Ye J, Toscano J. Ultrasensitive multispecies spectroscopic breath analysis for real-time health monitoring and diagnostics. *PNAS* (2021) 118(40):e2105063118. doi: 10.1073/pnas.2105063118

14. Buszewski B, Grzywinski D, Ligor T, Stacewicz T, Bielecki Z, Wojtas J. Detection of volatile organic compounds as biomarkers in breath analysis by different analytical techniques. *Bioanalysis* (2013) 5(18):2287–306. doi: 10.4155/bio.13.183

15. Arnaldo D, Giorgio P, Marco S, Martinelli E, Roscioni C, Galluccio G, et al. An investigation on electronic nose diagnosis of lung cancer. *Lung Cancer* (2010) 68(2):170–6. doi: 10.1016/j.lungcan.2009.11.003

16. Li W, Jia Z, Xie D, Chen K, Cui J, Liu H. Recognizing lung cancer using a homemade e-nose: a comprehensive study. *Comput Biol Med* (2020) 120:103706. doi: 10.1016/j.compbiomed.2020.103706

17. Li J, Peng Y, Liu Y, Li W, Jin Y, Tang Z, et al. Investigation of potential breath biomarkers for the early diagnosis of breast cancer using gas chromatography-mass spectrometry. *Clin Chim Acta* (2014) 436:59–67. doi: 10.1016/j.cca.2014.04.030

18. Rudnicka J, Kowalkowski T, Buszewski B. Searching for selected VOCs in human breath samples as potential markers of lung cancer. *Lung Cancer* (2019) 135:123–9. doi: 10.1016/j.lungcan.2019.02.012

19. Zhou J, Huang ZA, Kumar U, Chen DDY. Review of recent developments in determining volatile organic compounds in exhaled breath as biomarkers for lung cancer diagnosis. *Anal Chim Acta* (2017) 996:1–9. doi: 10.1016/j.aca.2017.09.021

20. Smith D, Spaněl P, Herbig J, Beauchamp J. Mass spectrometry for real-time quantitative breath analysis. *J Breath R* (2014) 8(2):0271. doi: 10.1088/1752-7155/8/2/027101

21. Wehinger A, Schmid A, Mechtcheriakov S, Ledochowski M, Grabmer C, Gastlfh GA, et al. Lung cancer detection by proton transfer reaction mass-spectrometric analysis of human breath gas. *Int J Mass Spectrom* (2007) 265(1):49–59. doi: 10.1016/j.ijms.2007.05.012

22. Herbig J, Müller M, Schallhart S, Titzmann T, Graus M, Hansel A. On-line breath analysis with PTR-TOF. *J Breath R* (2009) 3(2):027004. doi: 10.1088/1752-7155/3/2/027004

23. Morisco F, Aprea E, Lembo V, Fogliano V, Vitaglione P, Mazzone G, et al. Rapid "breath-print" of liver cirrhosis by proton transfer reaction time-of-flight mass spectrometry. A pilot study. *PloS One* (2013) 8(4):e59658. doi: 10.1371/journal.pone.0059658

24. Herbig J, Titzmann T, Beauchamp J, Kohl I, Hansel A. Buffered end-tidal (BET) sampling-a novel method for real-time breath-gas analysis. *J Breath R* (2008) 2(3):037008. doi: 10.1088/1752-7155/2/3/037008

25. Navarro CLA, Damen JAA, Takada T, Nijman SWJ, Dhiman P, Ma J, et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. *BMJ* (2021) 375:n2281. doi: 10.1136/bmj.n2281

26. Meng S, Li Q, Zhou Z, Li H, Liu X, Pan S, et al. Assessment of an exhaled breath test using high-pressure photon ionization time-of-Flight mass spectrometry to detect lung cancer. *JAMA Netw Open* (2021) 4(3):e213486. doi: 10.1001/jamanetworkopen.2021.3486

27. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* (2015) 350:g7594. doi: 10.1161/CIRCULATIONAHA.114.014508

28. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* (2015) 351:h5527. doi: 10.1136/bmj.h5527

29. Riley RD, Ensor J, Snell KIE, Harrell FEJr., Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ* (2020) 368:m441. doi: 10.1136/bmj.m441

30. Niculescu-Mizil A, Caruana R. Predicting good probabilities with supervised learning, in: *Proceedings of the 22nd international conference on Machine Learning (ICML '05)*. New York, NY, United States: Association for Computing Machinery (2005). pp. 625–32.

31. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. *J Mach Learn Res* (2011) 12:2825–30. doi: 10.5555/1953048.2078195

32. Schallschmidt K, Becker R, Jung C, Bremser W, Walles T, Neudecker J, et al. Comparison of volatile organic compounds from lung cancer patients and healthy controls–challenges and limitations of an observational study. *J Breath Res* (2016) 10:046007. doi: 10.1088/1752-7155/10/4/046007

33. Ma K, Zhang L. Overview: lipid metabolism in the tumor microenvironment. *Adv Exp Med Biol* (2021) 1316:41–7. doi: 10.1007/978-981-33-6785-2_3

34. Duettmann W, Koidl C, Troppan K, Seeber K, Buzina W, Wölfler A, et al. Serum and urine galactomannan testing for screening in patients with hematological malignancies. *Med Mycol* (2014) 52(6):647–52. doi: 10.1093/mmy/myu019

35. Pepe MS, Feng Z, Potter JD, Thompson ML, Thornquist M, Winget M, et al. Phases of biomarker development for early detection of cancer. *J Natl Cancer Inst* (2001) 93(14):1054–61. doi: 10.1093/jnci/93.14.1054

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 29 September 2022
DOI 10.3389/fonc.2022.1023776

Check for updates

OPEN ACCESS

EDITED BY
Hongda Liu,
Nanjing Medical University, China

REVIEWED BY
Jianbo Wang,
Shandong University, China
Li Yu,
Zhuhai People's Hospital, China

*CORRESPONDENCE
Cui Tian
smootheve2019@163.com

SPECIALTY SECTION
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

RECEIVED 20 August 2022
ACCEPTED 13 September 2022
PUBLISHED 29 September 2022

CITATION
Li Z, Liang N, Wang N, Jia Y and Tian C (2022) WDR5 is a prognostic biomarker of brain metastasis from non-small cell lung cancer. *Front. Oncol.* 12:1023776. doi: 10.3389/fonc.2022.1023776

COPYRIGHT
© 2022 Li, Liang, Wang, Jia and Tian. 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.

WDR5 is a prognostic biomarker of brain metastasis from non-small cell lung cancer

Zheng Li[1], Nan Liang[1], Na Wang[2], Yan Jia[3] and Cui Tian[4]*

[1]Department of Neurosurgery, The Second Affiliated Hospital of Shandong First Medical University, Tai'an, China, [2]Department of Anesthesiology, The First People's Hospital of Tai'an, Tai'an, China, [3]Department of Intensive Care Unit, Shandong Provincial Tai'shan Hospital, Tai'an, China, [4]Department of Intensive Care Unit, The Second Affiliated Hospital of Shandong First Medical University, Tai'an, China

Background: Lung cancer (LC) is the most frequent caner type and causes the most cancer-related death. Brain metastases (BM) are the deadliest complications of lung cancer, and the prognostic biomarkers of BM are urgently needed.

Materials and methods: In our study, we established an inception cohort including 122 patients with asynchronous BM from NSCLC, and further selected 70 patients who received surgical resection, which compromised the validation cohort. With immunohistochemistry, we investigated the expression of WDR5 in the cohort. By chi-square method, the correlations between WDR5 and clinicopathological factors were analyzed. The prognostic indicators were analyzed with the univariate analysis, and independent prognostic factors were identified by multivariate analysis with Cox-regression model.

Results: WDR5 is frequently expressed in the cytoplasm of BM from NSCLC. Patients with low or high expression of WDR5 account for 60% and 40% respectively. High expression of WDR5 indicates poor prognosis of BM from NSCLC (P=0.001). In addition to WDR5, KPS is also a prognostic factor of BM, and high KPS predicts favorable prognosis (P=0.006). WDR5 is an independent prognostic biomarker for poor prognosis of BM from NSCLC, with the cancer-related odds as 2.48.

Conclusions: High expression of WDR5 can predict the poor prognosis of BM, and WDR5 is an independent prognostic biomarker of BM from NSCLC. Patients with WDR5 overexpression are more high-risk to suffer BM-related death and should receive more intense post-operational supervision.

KEYWORDS

WDR5, lung cancer, brain metastasis, prognosis, biomarker, patient cohort

Introduction

Lung cancer (LC) is the malignancy with the highest morbidity and it results in most cancer-related death worldwide, accounting for 18.4% of all cancer-related deaths (1). Approximately 2.1 million patients were diagnosed as lung cancer in 2018 (1). The public health burden of LC is much more challenging in developing countries such as China, because they usually have more severe air pollution and higher LC frequencies (2). Histologically, LC can be divided into non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC) (3). NSCLC accounts for approximately 85% of all LCs, which can be further categorized into adenocarcinoma (LAD), squamous cell (SQCC), and large cell carcinoma (LCC) (4, 5).

Brain metastases (BM) are the most common and also the most devastating complications for the patients with LC, taking up approximately 30% of patients with advanced-stage LC (6). Every year, about 150,000 cancer patients develop BM in the United States (7), and LC is the main cause of secondary BM (8, 9). The median survival time of BM from NSCLC ranges from 3 to 14.8 months (10). The effect of chemotherapy is very limited because of blood brain barrier, and surgical resection is one of the treatment options, especially to those who have the opportunity of surgical resection or have significant symptoms of intracranial compression (11). BM from NSCLC is highly heterogeneous and requires more specific molecular classification (12). However, the molecular pattern of BM from NSCLC are neglected, and more prognostic biomarkers of BM from NSCLC should be investigated for the precision classification and treatment.

WD repeat domain 5 (WDR5) is an essential component of histone methyltransferase complex SET1/MLL, which catalyzes the lysine 4 methylation of histone 3 (H3K4me) (13). H3K4me is one of the most important epigenetic modifications and its main function is to enhance the transcription and expression of the substrates, requiring the participation of WDR5 (14). Interestingly, WDR5 is reported to interact with both unmodified and methylated H3K4 *in vitro* (15). The ectopic expression of WDR5 and negative correlations between WDR5 and prognosis are reported in several cancer types including prostate cancer, bladder cancer, cholangiocarcinoma, breast cancer and colorectal cancer (16–20). In lung cancer, WDR5 is reported to promote proliferation of lung adenocarcinoma by inducing SOX9 expression (21). In LC cell line A549 cells, WDR5 positively regulates P53 stability and arrests A549 cell cycle in G1 phase (22), which is conflicting with most studies indicating WDR5 as a tumor-promotor. However, the expression and clinical indications of WDR5 in BM from NSCLC are still unknown.

In our study, we established a cohort consisting of 70 patients with BM from NSCLC, and investigated the expression of WDR5 in the cohort. We analyzed the clinicopathological correlation of WDR5, and estimated WDR5 prognostic significance by univariate and multivariate analyses.

Materials and methods

Patient cohort and ethics

The inception cohort of our study consisted of the patients who underwent surgical resection of NSCLC, and were clinically diagnosed as asynchronous BM from NSCLC in the Second Hospital Affiliated to Shandong First Medical University and The First People's Hospital of Tai'an from 2016 to 2020. The validation cohort was selected from the inception cohort following the criteria: (i) available formalin-fixed tumor tissues and follow-ups, (ii) no pre-operational adjuvant therapy including chemotherapy or radiotherapy, (iii) solitary and resectable BM lesion (iv) received gross total resection. Patients who died during the perioperative period were excluded from the study.

The study was approved by the Ethics Committee of the Second Affiliated Hospital of Shandong First Medical University and The First People's Hospital of Tai'an. The written informed consent was obtained from each patient.

Immunohistochemistry

Immunohistochemistry (IHC) was performed with streptavidin peroxidase complex method as previous described (23). The specimens were deparaffinized and rehydrated with xylene and graded alcohol first, followed by the antigen retrieval in boiled citrate buffer (pH = 6.0) for 30 minutes. After that, 3% hydrogen peroxide was applied to block the activity of endogenous peroxidase. 5% bovine serum albumin was administrated for diminishing the unspecific antigen binding. Primary antibody of WDR5 (R&D Systems, AF5810) was used to incubate the tissues at dilution of 1:50 overnight at 4°C, and the corresponding secondary antibody labelled with streptavidin-biotin-peroxidase reagent (Beyotime, Beijing, China) was applied to incubate the specimens for 30 minutes at room temperature. At last, the slides were incubated in the 3,3'-diaminobenzidine solution for visualization and tissues were counter-stained with haematoxylin.

Evaluation of IHC

The semi-quantification method of IHC results was applied in our study. Two senior pathologists who were unaware of the clinical data estimated the IHC results and cases without consensus received a third pathologist for final determination. The IHC results were semi-quantified by IHC scores according to previous studies (24, 25), which were the multiplication product of score for staining intensity and positive cell percentage. The scores for staining intensity were identified as: score 0 for negative staining; score 1 for weak staining; score 2 for moderate staining and score 3 for strong staining. The scores of positively stained cell percentage were delineated as: score 1

for <25% positive cells; score 2 for 25%-50% positive cells; score 3 for 50%-75% positive cells; score 4 for 75%-100% positive cells. The final score ranged from 0 to 12, and the cut-off was identified by receiver operating characteristic (ROC) curve as previous studies (22, 26). Our cohort was separated to $WDR5^{low}$ and $WDR5^{high}$ subsets based on the cut-off of IHC score, which was defined as 3.0 in our study.

Statistical analysis

All the statistical analyses were analyzed by SPSS 22.0 software (SPSS, Chicago, IL, USA). The correlations between WDR5 expression and clinicopathological factors were assessed by the chi-square test. The overall survival (OS) rates were calculated by Kaplan-Meier method and the statistical differences between subsets were assessed by the log-rank test. In multivariate analysis, cox-regression hazard model was applied to identify the independent prognostic factors. *P* value less than 0.05 was regarded as statistically significant.

Results

Basic information of the inception and validation cohort

The inception cohort consisted of 122 patients who were clinically diagnosed as asynchronous BM from NSCLC. There were 68 male and 54 female patients, with an average age as 61.4 years old. The validation cohort was selected from the inception cohort if patients received a surgical resection of BM. There were 70 patients in the validation cohort, comprised of 41 male patients and 29 female patients (Table 1). The average age of validation cohort was 62.7 years old, and the average survival time was 6.0 months. The basic line of inception and validation cohort had no significant difference in our study.

WDR5 expression in brain metastasis of NSCLC

The expression and intracellular localization of WDR5 in BM of NSCLC were detected with IHC. WDR5 was mainly expressed in the nucleus of BMs (Figure 1). As described in *Materials and Methods* in detail, we categorized our cohort into subsets with low or high expression of WDR5. In our study, patients with low or high expression of WDR5 accounted for 60% and 40% respectively.

Clinical relevance of WDR5 in BM from NSCLC

The correlations between WDR5 expression and clinicopathological factors were analyzed with the chi-square test (Table 2). The clinicopathological factors included the gender and age of patients, the tumor size, histological type,

TABLE 1 Basic information of the inception and validation cohort.

Factors	Inception		Validation	
	number	percentage	number	percentage
Age				
≤60	45	36.89%	29	41.43%
>60	77	63.11%	41	58.57%
Gender				
Male	68	55.74%	41	58.57%
Female	54	44.26%	29	41.43%
Tumor size				
<3cm	44	36.07%	27	38.57%
≥3cm	78	63.93%	43	61.43%
Histological type				
Squamous carcinoma	67	54.92%	37	52.86%
Adenocarcinoma	55	45.08%	33	47.14%
KPS				
<80	50	40.98%	22	31.43%
≥80	72	59.02%	48	68.57%
WDR5				
Low	–	–	42	60.00%
High	-	-	28	40.00%

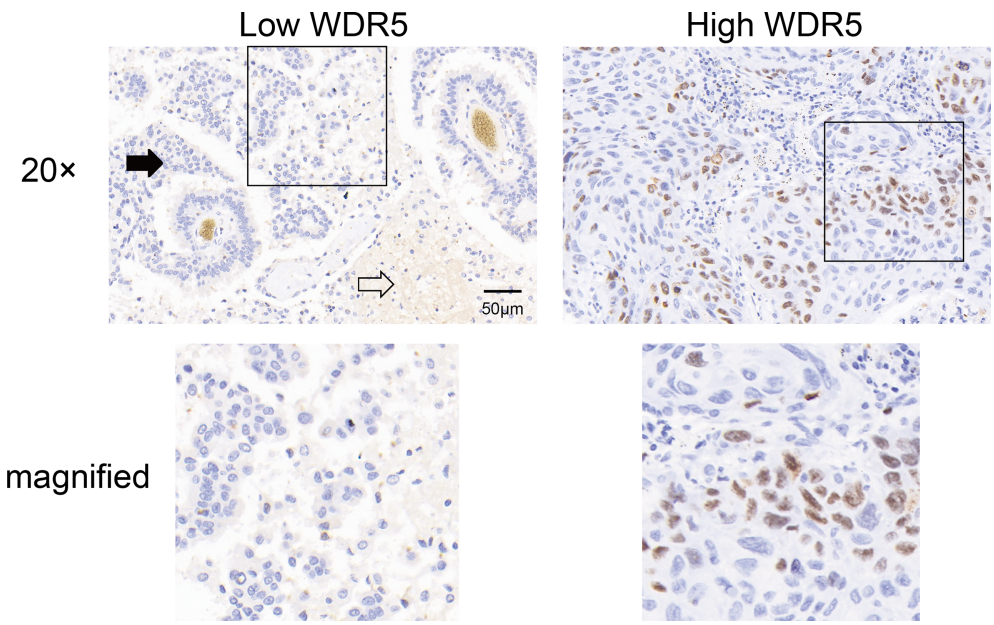


FIGURE 1
The expression of WDR5 in BM from NSCLC. The expression of WDR5 in BM from NSCLC was detected with IHC, dividing the cohort into high and low WDR5 expression. Black arrow indicates BM, and white arrow indicates brain tissue.

KPS score. In our study, no remarkable associations between WDR5 expression and above clinicopathological factors were observed.

WDR5 predicted poor prognosis of BM from NSCLC

Using log-rank method, we analyzed the prognostic significance of WDR5 and other clinicopathological factors (Table 3). High expression of WDR5 substantially correlated with poor prognosis of BM from NSCLC, suggesting that WDR5 was a prognostic biomarker (Figure 2A). In addition, high KPS was a favorable prognostic biomarker of BM from NSCLC (P=0.006) (Figure 2B). Interestingly, BMs from adenocarcinoma seemed to have a poorer prognosis compared with those from squamous carcinoma (Figure 2C). The 1-year OS rates of BM from squamous carcinoma and adenocarcinoma were 24.7% and 4.1%, respectively (Table 3). Other clinicopathological factors such as age, gender and tumor size, had no significant association with OS rates (Figures 2D-F).

By Cox-regression model, we further identified independent prognostic factors of BM from NSCLC (Table 3). The factors with P value less than 0.1 in univariate analysis were enrolled to the multivariate analysis, including WDR5 expression, KPS status and histological type of BM. All these factors were confirmed as the independently prognostic factors of BM. The odds of patients with high WDR5 of BM were 2.48 times of those with low WDR5 expression (P=0.001). More intriguingly, patients with BM from adenocarcinoma were more vulnerable to BM-related deaths than those with BM from squamous carcinoma.

TABLE 2 The correlation between WDR5 and the clinicopathological factors.

Factors	WDR5		P*
	Low	High	
Age			
≤60	17	12	0.843
>60	25	16	
Gender			
Male	25	16	0.843
Female	17	12	
Tumor size			
<3cm	18	9	0.367
≥3cm	24	19	
Histological type			
Squamous carcinoma	20	17	0.282
Adenocarcinoma	22	17	
KPS			
<80	11	11	0.248
≥80	31	17	

*Represents that data were analyzed with chi-square test.

Discussion

More than half of BMs originate from lung cancer (27), and up to 50% of all LC suffer BM in different course of the LC (28). Although many progresses have been made in BM treatment such as whole brain irradiation, the prognosis of BM from NSCLC is still extremely dismal (29). Clinical management of BM from NSCLC is affected by many aspects such as the performance status and the overall health of the patient, meanwhile, the standard treatment of BM is still in exploration (6). Just like the primary tumor, BM is also very heterogeneous, but unfortunately, the molecular characters of BM from NSCLC are not well defined. Therefore, how to select the high-risk patients and apply the individual treatment is very important. Here we identified WDR5 as an independent prognostic biomarker of NSCLC-associated BM, indicating that patients with high WDR5 expression are much more high-risk to suffer BM and should receive more intense post-operational supervision.

Understanding the molecular mechanisms and screening potential biomarkers of NSCLC-associated BM may help discover more novel therapeutic drug targets to improve the prognosis of BM. Unfortunately, there is no effective method to predict the high-risk patients who are susceptible to BM. Several studies attempted to select high-risk patients of BM. For example, historical data from 1973 to 2011 using the SEER database showed that 9% NSCLC patients without previously non-metastasis would suffer BM during the disease course, whereas this number was as high as 18% to patients who suffer small cell lung cancer patients without previous metastasis (30). In general, age of ≤ 60 years, non-squamous cell carcinoma and the presence of clinical bulky mediastinal lymph nodes (> 2 cm) are reported to be correlated with a high BM rate in locally advanced NSCLC (31–34). However, these results have not got total consensus, and most studies are focused on the clinicopathological factors but not the molecular features. Consistent with previous studies (35, 36), we demonstrated that low KPS and adenocarcinoma are associated with more unfavorable prognoses in our study. Although the sample size of our cohort of BM from NSCLC is not small (n=70), more multi-center and perspective cohort studies should be conducted to identify more effective biomarkers and possible drug targets of BM.

The oncogenic role of WDR5 has not been well elucidated. Most studies focused on the function of WDR5 as a component of histone methyltransferase complex. Emerging evidence gradually revealed the precise regulatory mechanism of WDR5 on H3K4me. For example, the phosphorylation of WDR5 is also shown to modulate the WDR5 function and thus regulate H3K4me of substrates (37). Moreover, more evidence showed that WDR5 can promote gene transcription in addition to modulating H3K4me, such as binding with c-Myc and thus regulating transcription (38). Here we identified WDR5 as an independent biomarker indicating the unfavorable prognosis of BM from NSCLC, but underlying molecular mechanism was not investigated here. Moreover, whether WDR5 overexpression results in easier BM is also an interesting and worthy topic to investigate because there is still no effective biomarker to predict BM risk to date. Several WDR5 inhibitors are in clinical trial for

TABLE 3 Prognostic significance in univariate and multivariate analyses.

Factors	1-year OS	P^*	HR	95%CI	$P^{\#}$
Age					
≤60	18.2	0.542	–	–	–
>60	12.3		–	–	
Gender					
Male	9.6	0.630	–	–	–
Female	18.1		-	-	
Tumor size					
<3cm	21.7	0.459	–	–	–
≥3cm	9.1		–	–	
Histological type					
Squamous carcinoma	24.7	0.063	1		
Adenocarcinoma	4.1		1.91	1.09-3.33	0.024
KPS					
<80	4.5	0.006	1		
≥80	21.8		2.23	1.26-3.94	0.006
WDR5					
Low	21.9	0.001	1		
High	4.6		2.48	1.43-4.31	0.001

*Represents that data were analyzed with log-rank test; #represents that data were analyzed with Cox-regression hazard model.

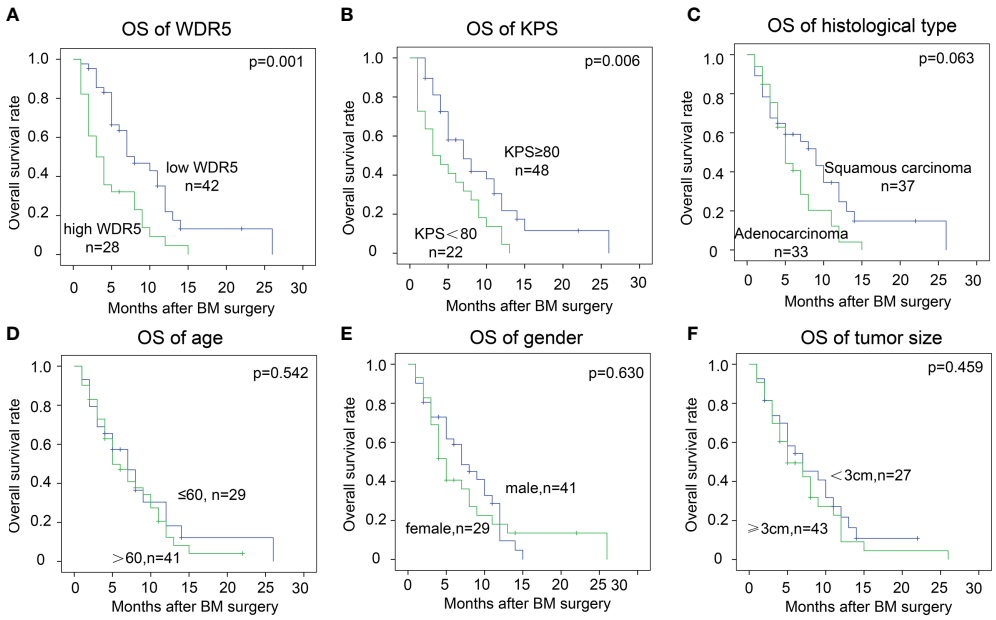


FIGURE 2
Prognostic significance of WDR5 and clinicopathological factors. The cohort was divided into subsets according to WDR5 expression and clinicopathological factors. High WDR5 **(A)** and low KPS **(B)** were significantly associated with low OS rate. Patients' age **(C)**, gender **(D)**, tumor size **(E)** and histological type **(F)** had no statistically significant correlation with OS of BM.

the treatment of acute myeloid leukemia. Our results also indicate that WDR5 may be a potential drug target of BM from NSCLC, and targeting WDR5 may be a possible treatment option.

In conclusion, we established a cohort of patients with asynchronous BM from NSCLC, and investigated the expression of WDR5 for the first time. Interestingly, high expression of WDR5 can predict the poor prognosis of BM, and WDR5 is an independent prognostic biomarker of BM from NSCLC. Our results indicate that patients with WDR5 overexpression are more high-risk to suffer BM-related death, and should receive stricter post-operational supervision. Our study provides more evidence for the molecular classification of BM from NSCLC, and may help guide the individual treatment of BM.

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 study was approved by the Ethics Committee of the Second Affiliated Hospital of Shandong First Medical University and The First People's Hospital of Tai'an. The written informed consent was obtained from each patient. The patients/participants provided their written informed consent to participate in this study.

Author contributions

ZL, NL, NW and YJ collected the specimen and established the cohort. ZL and NL performed the experiments and analyzed the data. CT designed the study and wrote the paper. 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.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: Cancer J Clin* (2018) 68:394–424. doi: 10.3322/caac.21492

2. Cao M, Chen W. Epidemiology of lung cancer in China. *Thorac Cancer* (2019) 10:3–7. doi: 10.1111/1759-7714.12916

3. Hirsch FR, Scagliotti GV, Mulshine JL, Kwon R, Curran WJ Jr, Wu YL, et al. Lung cancer: current therapies and new targeted treatments. *Lancet* (2017) 389:299–311. doi: 10.1016/S0140-6736(16)30958-8

4. Succony L, Rassl DM, Barker AP, McCaughan FM, Rintoul RC. Adenocarcinoma spectrum lesions of the lung: Detection, pathology and treatment strategies. *Cancer Treat Rev* (2021) 99:102237. doi: 10.1016/j.ctrv.2021.102237

5. The L. Lung cancer: some progress, but still a lot more to do. *Lancet* (2019) 394:1880. doi: 10.1016/S0140-6736(19)32795-3

6. Johung KL, Yeh N, Desai NB, Williams TM, Lautenschlaeger T, Arvold ND, et al. Extended survival and prognostic factors for patients with ALK-rearranged non-Small-Cell lung cancer and brain metastasis. *J Clin Oncol Off J Am Soc Clin Oncol* (2016) 34:123–9. doi: 10.1200/JCO.2015.62.0138

7. Bovi JA, White J. Radiation therapy in the prevention of brain metastases. *Curr Oncol Rep* (2012) 14:55–62. doi: 10.1007/s11912-011-0208-6

8. Subramanian A, Harris A, Piggott K, Shieff C, Bradford R. Metastasis to and from the central nervous system–the 'relatively protected site'. *Lancet Oncol* (2002) 3:498–507. doi: 10.1016/s1470-2045(02)00819-7

9. Nathoo N, Chahlavi A, Barnett GH, Toms SA. Pathobiology of brain metastases. *J Clin Pathol* (2005) 58:237–42. doi: 10.1136/jcp.2003.013623

10. Sperduto PW, Kased N, Roberge D, Xu Z, Shanley R, Luo X, et al. Summary report on the graded prognostic assessment: an accurate and facile diagnosis-specific tool to estimate survival for patients with brain metastases. *J Clin Oncol Off J Am Soc Clin Oncol* (2012) 30:419–25. doi: 10.1200/JCO.2011.38.0527

11. Lippitz B, Lindquist C, Paddick I, Peterson D, O'Neill K, Beaney R, et al. Stereotactic radiosurgery in the treatment of brain metastases: the current evidence. *Cancer Treat Rev* (2014) 40:48–59. doi: 10.1016/j.ctrv.2013.05.002

12. Achrol AS, Rennert RC, Anders C, Soffietti R, Ahluwalia MS, Nayak L, et al. Brain metastases. *Nat Rev Dis Primers* (2019) 5:5. doi: 10.1038/s41572-018-0055-y

13. Guarnaccia AD, Tansey WP. Moonlighting with WDR5: A cellular multitasker. *J Clin Med* (2018) 7(2):21. doi: 10.3390/jcm7020021

14. Calo E, Wysocka J. Modification of enhancer chromatin: what, how, and why? *Mol Cell* (2013) 49:825–37. doi: 10.1016/j.molcel.2013.01.038

15. Dou Y, Milne TA, Tackett AJ, Smith ER, Fukuda A, Wysocka J, et al. Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell* (2005) 121:873–85. doi: 10.1016/j.cell.2005.04.031

16. Dai X, Guo W, Zhan C, Liu X, Bai Z, Yang Y, et al. WDR5 expression is prognostic of breast cancer outcome. *PloS One* (2015) 10:e0124964. doi: 10.1371/journal.pone.0124964

17. Chen X, Xie W, Gu P, Cai Q, Wang B, Xie Y, et al. Upregulated WDR5 promotes proliferation, self-renewal and chemoresistance in bladder cancer. *Via Mediating H3K4 Trimethylation Sci Rep* (2015) 5:8293. doi: 10.1038/srep08293

18. Carugo A, Genovese G, Seth S, Nezi L, Rose JL, Bossi D. *Inn vivo* functional platform targeting patient-derived xenografts identifies WDR5-myc association as a critical determinant of pancreatic cancer. *Cell Rep* (2016) 16:133–47. doi: 10.1016/j.celrep.2016.05.063

19. Tan X, Chen S, Wu J, Lin J, Pan C, Ying X, et al. PI3K/AKT-mediated upregulation of WDR5 promotes colorectal cancer metastasis by directly targeting ZNF407. *Cell Death Dis* (2017) 8:e2686. doi: 10.1038/cddis.2017.111

20. Chen T, Li K, Liu Z, Liu J, Wang Y, Sun R, et al. WDR5 facilitates EMT and metastasis of CCA by increasing HIF-1alpha accumulation in myc-dependent and independent pathways. *Mol Ther J Am Soc Gene Ther* (2021) 29:2134–50. doi: 10.1016/j.ymthe.2021.02.017

21. Ji A, Qian L, Tian Z, Cui J. WDR5 promotes the proliferation of lung adenocarcinoma by inducing SOX9 expression. *Biomarkers Med* (2021) 15:1599–609. doi: 10.2217/bmm-2021-0184

22. Xie Q, Li Z, Chen J. WDR5 positively regulates p53 stability by inhibiting p53 ubiquitination. *Biochem Biophys Res Commun* (2017) 487:333–8. doi: 10.1016/j.bbrc.2017.04.060

23. Du K, Wu X, Ji X, Liang N, Li Z. Early growth response 1 promoted the invasion of glioblastoma multiforme by elevating HMGB1. *J Neurosurgical Sci* (2020). doi: 10.23736/S0390-5616.20.05107-3

24. Liu Z, Liu J, Chen T, Wang Y, Shi A, Li K, et al. Wnt-TCF7-SOX9 axis promotes cholangiocarcinoma proliferation and pemigatinib resistance in a FGF7-FGFR2 autocrine pathway. *Oncogene* (2022) 41:2885–96. doi: 10.1038/s41388-022-02313-x

25. Liu J, Ren G, Li K, Liu Z, Wang Y, Chen T, et al. The Smad4-MYO18A-PP1A complex regulates beta-catenin phosphorylation and pemigatinib resistance by inhibiting PAK1 in cholangiocarcinoma. *Cell Death Differentiation* (2022) 29:818–31. doi: 10.1038/s41418-021-00897-7

26. Qiu B, Chen T, Sun R, Liu Z, Zhang X, Li Z, et al. Sprouty4 correlates with favorable prognosis in perihilar cholangiocarcinoma by blocking the FGFR-ERK signaling pathway and arresting the cell cycle. *EBioMedicine* (2019) 50:166–77. doi: 10.1016/j.ebiom.2019.11.021

27. Budczies J, von Winterfeld M, Klauschen F, Bockmayr M, Lennerz JK, Denkert C, et al. The landscape of metastatic progression patterns across major human cancers. *Oncotarget* (2015) 6:570–83. doi: 10.18632/oncotarget.2677

28. Yousefi M, Bahrami T, Salmaninejad A, Nosrati R, Ghaffari P, Ghaffari SH, et al. Lung cancer-associated brain metastasis: Molecular mechanisms and therapeutic options. *Cell Oncol (Dordr)* (2017) 40:419–41. doi: 10.1007/s13402-017-0345-5

29. Gavrilovic IT, Posner JB. Brain metastases: epidemiology and pathophysiology. *J Neuro-Oncol* (2005) 75:5–14. doi: 10.1007/s11060-004-8093-6

30. Goncalves PH, Peterson SL, Vigneau FD, Shore RD, Quarshie WO, Islam K, et al. Risk of brain metastases in patients with nonmetastatic lung cancer: Analysis of the metropolitan Detroit surveillance, epidemiology, and end results (SEER) data. *Cancer* (2016) 122:1921–7. doi: 10.1002/cncr.30000

31. Ji Z, Bi N, Wang J, Hui Z, Xiao Z, Feng Q, et al. Risk factors for brain metastases in locally advanced non-small cell lung cancer with definitive chest radiation. *Int J Radiat Oncol Biol Phys* (2014) 89:330–7. doi: 10.1016/j.ijrobp.2014.02.025

32. Ceresoli GL, Reni M, Chiesa G, Carretta A, Schipani S, Passoni P, et al. Brain metastases in locally advanced nonsmall cell lung carcinoma after multimodality treatment: risk factors analysis. *Cancer* (2002) 95:605–12. doi: 10.1002/cncr.10687

33. Mujoomdar A, Austin JH, Malhotra R, Powell CA, Pearson GD, Shiau MC, et al. Clinical predictors of metastatic disease to the brain from non-small cell lung carcinoma: primary tumor size, cell type, and lymph node metastases. *Radiology* (2007) 242:882–8. doi: 10.1148/radiol.2423051707

34. Na II, Lee TH, Choe DH, Cheon GJ, Kim CH, Koh JS, et al. A diagnostic model to detect silent brain metastases in patients with non-small cell lung cancer. *Eur J Cancer* (2008) 44:2411–7. doi: 10.1016/j.ejca.2008.07.003

35. Gaspar L, Scott C, Rotman M, Asbell S, Phillips T, Wasserman T, et al. Recursive partitioning analysis (RPA) of prognostic factors in three radiation therapy oncology group (RTOG) brain metastases trials. *Int J Radiat Oncol Biol Phys* (1997) 37:745–51. doi: 10.1016/s0360-3016(96)00619-0

36. Hanibuchi M, Kim SJ, Fidler IJ, Nishioka Y. The molecular biology of lung cancer brain metastasis: an overview of current comprehensions and future perspectives. *J Med Invest JMI* (2014) 61:241–53. doi: 10.2152/jmi.61.241

37. Kikani CK, Wu X, Paul L, Sabic H, Shen Z, Shakya A, et al. Pask integrates hormonal signaling with histone modification *via* Wdr5 phosphorylation to drive myogenesis. *eLife* (2016) 5:e17985. doi: 10.7554/eLife.17985

38. Thomas LR, Foshage AM, Weissmiller AM, Tansey WP. The MYC-WDR5 nexus and cancer. *Cancer Res* (2015) 75:4012–5. doi: 10.1158/0008-5472.CAN-15-1216

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 10 October 2022
DOI 10.3389/fonc.2022.1005962

Check for updates

OPEN ACCESS

EDITED BY
Qingqing Zhu,
The First Affiliated Hospital of Soochow University, China

REVIEWED BY
Depeng Jiang,
Second Affiliated Hospital of Chongqing Medical University, China
Song Xu,
Tianjin Medical University General Hospital, China
Lai Baochang,
Xi'an Jiaotong University Health Science Center, China

*CORRESPONDENCE
Kai Yuan
yuankai1978@163.com

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

SPECIALTY SECTION
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

RECEIVED 28 July 2022
ACCEPTED 20 September 2022
PUBLISHED 10 October 2022

CITATION
Chen P, Quan Z, Song X, Gao Z and Yuan K (2022) *MDFI* is a novel biomarker for poor prognosis in LUAD. *Front. Oncol.* 12:1005962. doi: 10.3389/fonc.2022.1005962

COPYRIGHT
© 2022 Chen, Quan, Song, Gao and Yuan. 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.

MDFI is a novel biomarker for poor prognosis in LUAD

Pengyu Chen[1,2†], Zhen Quan[1,2†], Xueyu Song[1,2], Zhaojia Gao[1,3] and Kai Yuan[1,3*]

[1]Division of Thoracic Surgery, The Affiliated Changzhou Second People's Hospital of Nanjing Medical University, Changzhou, China, [2]School of Medicine, Dalian Medical University, Dalian, China, [3]Heart and Lung Disease Laboratory, The Affiliated Changzhou Second People's Hospital of Nanjing Medical University, Changzhou, China

Background: Approximately 80% of lung cancers are non-small cell lung cancers (NSCLC). Lung adenocarcinoma (LUAD) is the main subtype of NSCLC. The incidence and mortality of lung cancer are also increasing yearly. Myogenic differentiation family inhibitor (*MDFI*) as a transcription factor, its role in lung cancer has not yet been clarified.

Methods: LUAD data were downloaded from The Cancer Genome Atlas (TCGA) database and Gene Expression Omnibus (GEO), analyzed and plotted using the R language. Associations between Clinical information and *MDFI* expression were assessed using logistic regression analyses to explore the effects of *MDFI* on LUAD. Two sets of tissue microarrays (TMAs) further confirmed the overexpression of *MDFI* in LUAD and its impact on prognosis. In addition, we examined the correlation between *MDFI* and immune infiltration. To investigate the effect of *MDFI* on the biological behavior of LUAD tumor cells by GSEA and GO/KEGG analysis. The survival status and somatic mutational characteristics of patients according to *MDFI* levels were depicted and analyzed.

Results: Expression of high *MDFI* in LUAD tissues *via* analyzing TCGA dataset (P <0.001). Kaplan-Meier survival analysis indicated a poor prognosis for those patients with LUAD who had upregulated *MDFI* expression levels (P <0.001). This was also verified by two groups of TMAs (P=0.024). Using logistic statistics analysis, *MDFI* was identified as an independent predictive factor and was associated with poor prognosis in LUAD (P <0.001, P =0.021). Assessment of clinical characteristics, tumor mutation burden (TMB), and tumor microenvironment (TME) between high- and low-expression score groups showed lower TMB, richer immune cell infiltration, and better prognosis in the low-risk group.

Conclusion: This study showed that *MDFI* was overexpressed in LUAD and was significantly associated with poor prognosis, indicating that *MDFI* may be used as a potential novel biomarker for the diagnosis and prognosis of LUAD. *MDFI* is associated with immune infiltration of LUAD and it is reasonable to speculate that it plays an important role in tumor proliferation and spread. In view of the

significant differences in *MDFI* expression between different biological activities, LUAD patients with *MDFI* overexpression may obtain more precise treatment strategies in the clinic.

KEYWORDS

MDFI, LUAD, TCGA, high expression, prognosis

Introduction

Lung cancer is one of the most common malignant tumors in the world, and is the malignant tumor with the highest mortality rate in the world. NSCLC accounts for approximately 80% of all lung cancers. Its cancer cells grow and divide slowly and spread and metastasize relatively late. About 75% of patients are in the middle and advanced stages, and the 5-year survival rate is very low (1). LUAD is the most common subtype of NSCLC (2), accounting for 50% of all lung cancer diagnoses, and its frequency is increasing (3, 4). Among all NSCLC subtypes, LUAD is the most heterogeneous and aggressive, and has a very high tumor mutational burden associated with *EGFR*, *KRAS*, *BRAF*, *ERBB2*, *TP53*, *ALK*, *STK11*, and TTE1 mutations (5, 6). In the past 50 years, China has reported a significant increase in the incidence and mortality of lung cancer. The incidence and mortality of male lung cancer rank first among all malignant tumors in China, and the incidence of female lung cancer ranks second and the mortality rate ranks second (7). However, 5 years survival for patients with stage I NSCLC is roughly 80%, and patients with stage II to stage III disease have a 5 years survival of 13–60% (8). Although the addition of adjuvant chemotherapy in patients with a specific stage can increase survival by 5-10%, there is a significant amount of toxicity associated with it (9). The opportunity to improve survival is evident in early-stage disease and is driving research to integrate targeted therapies and ICIs (1). There is space for improvement in the treatment of LUAD and scientists delve into the identification of molecular markers associated with tumors and combine them with pathological classifications that affect personalized treatment of patients. In order to highly accurately predict patient survival and/or response to individualized treatment, new biomarkers were identified in those with LUAD.

By excavating the TCGA database, we found that *MDFI* may be a novel lung cancer–related candidate target. *MDFI* is a Protein Coding gene. Diseases associated with *MDFI* include Erythema Infectiosum and Viral Exanthem (10). This protein is a transcription factor that negatively regulates other myogenic family proteins (11). *MDFI* is overexpressed in breast, colorectal, and liver cancers and may promote tumorigenesis in certain tissues (12–14). However, the role of *MDFI* in lung cancer has not yet been reported. We attempted to elaborate the prognostic value of *MDFI* in LUAD by exploring TCGA database in this study.

Methods and materials

Patient data acquisition

We searched the GEO database for high-throughput sequencing or microarray data on LUAD and selected 4 LUAD transcriptome datasets with different sample sizes: GSE43458, GSE62948, GSE116959, GSE139032 (15–18). All of these datasets including lung tumors and para-tumor tissues. We searched the TCGA database (19) and obtained patient data for the LUAD cohort based on legitimate research objectives (20). The TCGA-LUAD cohort contained a total of 599 participants, including 59 normal patients and 539 patients with lung adenocarcinoma tumors. This includes mRNA data, clinicopathological data, and somatic mutation data. Some patients with missing data were excluded, and 535 oncology patients were enrolled in the study (21).

Tissue microarray and analysis of immunohistochemical results

To further evaluate the expression of *MDFI* in NSCLC, we obtained a tissue microarray (TMA) from Superbiotek (Shanghai, China). It containing 60 pairs of NSCLC specimens and para-tumor tissues. In addition, we constructed a TMA containing 140 NSCLC tissues [included LUSC (n = 80), LUAD (n = 51), adenosquamous carcinoma (n = 5), Bronchioloalveolar carcinoma (n = 3), and sarcomatoid carcinoma (n = 1)] and 10 normal lung tissues. The reasonable tumor stage of these patients was determined based on the World Health Organization criteria and the International TNM classification (22). The patient did not receive radiotherapy or chemotherapy, and did not have other tumors within 5 years before surgery (Supplementary Table 1). Two groups of TMAs were

deparaffinized with a conventional protocol and rehydrated according to a standard protocol for immunohistological (IHC) examination. Primary antibodies *MDFI* goat antibody (1:500, ProSci, PSI-42-165) were used (23, 24).

Then analyzing the staining results by Image Pro Plus 6 software (IPP6), the staining area (Area) and integrated optical density (IOD) can be obtained. The mean optical density (MOD) can be obtained by taking the ratio of the two, and the MOD value is positively correlated with the staining intensity of the tissue. By analyzing the MOD of *MDFI* in NSCLC, para tumor and normal tissues can be further studied. We mapped the immunostaining heat map of tumor tissue and normal lung tissue adjacent to the tumor, and labeled the corresponding expression scores. In addition, we also compared the expression levels of *MDFI* in the tumor and normal groups in 4 GEO validation sets.

Gene expression and survival time

High-volume data downloaded from the TCGA was managed using the R programming language (25). The results of unpaired and paired samples were analyzed by independent and paired sample t-test, respectively (26). We have selected several groups of characteristics with more obvious differences for comparative display. Boxplots plots, using Age, Pathologic stage, T stage, Smoker and so on as the variable, were generated to calculate differential expression of *MDFI*. Differences in global gene expression levels between the normal tissues and tumor tissues of patients with LUAD were analyzed with an R package and $P < 0.05$. Kaplan-Meier analysis was used to evaluate the prognostic value of *MDFI* in LUAD patients (25).

Immune cell and immune function analysis

The gene set analysis of variance (GSVA) procedure of the 'GSVA' and 'GSEA Base' package of R software was used to calculate gene signature enrichment scores (GSVA scores) for immune cells and immune function in each sample, which is called relative immune cell abundance [10.1038/ng765.] (27, 28). Heat maps of immune cells and immune function within disease groups and normal groups have been established using the "pheat map" software package. We then performed a correlative analysis of immune cells and immune function separately using the corresponding package, which was calculated using the Pearson correlation method (29). The scores of immune cells and immune function were also compared in the *MDFI* high expression group and low expression group (30).

Functional enrichment analysis

First, we divided the TCGA-LUAD cohort patients into two groups with high and low *MDFI* expression levels according to the *MDFI* expression. Differential analysis between groups was performed using the 'limma' package of R software (31), and differential genes were screened according to the criteria of adj. P <0.05, |log FC| >0.5. To explore the functions of the differentially expressed genes in *MDFI*, the screened genes were enriched for Gene Ontology (32) and Kyoto Encyclopedia of Genes and Genomes (33) pathway terms, respectively (34). We used the "cluster Profiler" package in R software to carry out the analysis and plotted the enriched pathway results separately (35).

Somatic mutation analysis

We downloaded mutation data from TCGA for the LUAD patient cohort. To investigate the relationship between *MDFI* and mutations, we first divided the mutation profile of patients into high *MDFI* mutation group and low *MDFI* mutation group depending on the expression of *MDFI*. We assessed the mutations in the two groups separately using the R software "Maftools", and plotted the waterfall of mutations (36). We analyzed differences in mutations in different *MDFI* expression groups, as well as the differences in the number and location of mutations in the same mutated gene due to different *MDFI* expression (37).

Predicted drugs

Finally, we used protein-drug interaction data from the DSigDB database (38) to identify potential drugs that could benefit lung adenocarcinoma. We adopted the following tests: t-test or wilcoxon test for differences between groups depending on the data, and pearson or spearman method for correlation analysis. Survival analysis was performed using Kaplan-Meier and Cox regression analysis, and the results were evaluated using LogRank test, respectively (25).

Statistical analysis

R software (version 4.1.1; The R Foundation for Statistical Computing) was used for all statistics of this article. All statistical tests were double-sided, and P values less than 0.05 were evaluated as significant (28).

Result

MDFI expression in TCGA dataset

Significantly increased levels of *MDFI* expression in LUAD and LUSC compared to normal lung tissue by pan-cancer analysis using TGCA data (P <0.001) (Figure 1A). Likewise, high expression of *MDFI* in LUAD and low expression in normal lung tissue was verified in four GEO validation sets (Figures 1B–E). However, in terms of Kaplan-Meier curve, LUAD with high *MDFI* expression has a significantly worse prognosis (P < 0.001), while the expression of *MDFI* was not significant with the prognosis of LUSC (P = 0.125). Therefore, the study focused on the role of *MDFI* in LUAD (Figures 1F–K).

Characteristics of patients

A total of 535 LUAD patients with the required clinical features were acquired from TCGA data portal in July 2022 (Table 1). Among them, 249 were male (46.5%) and 286 were

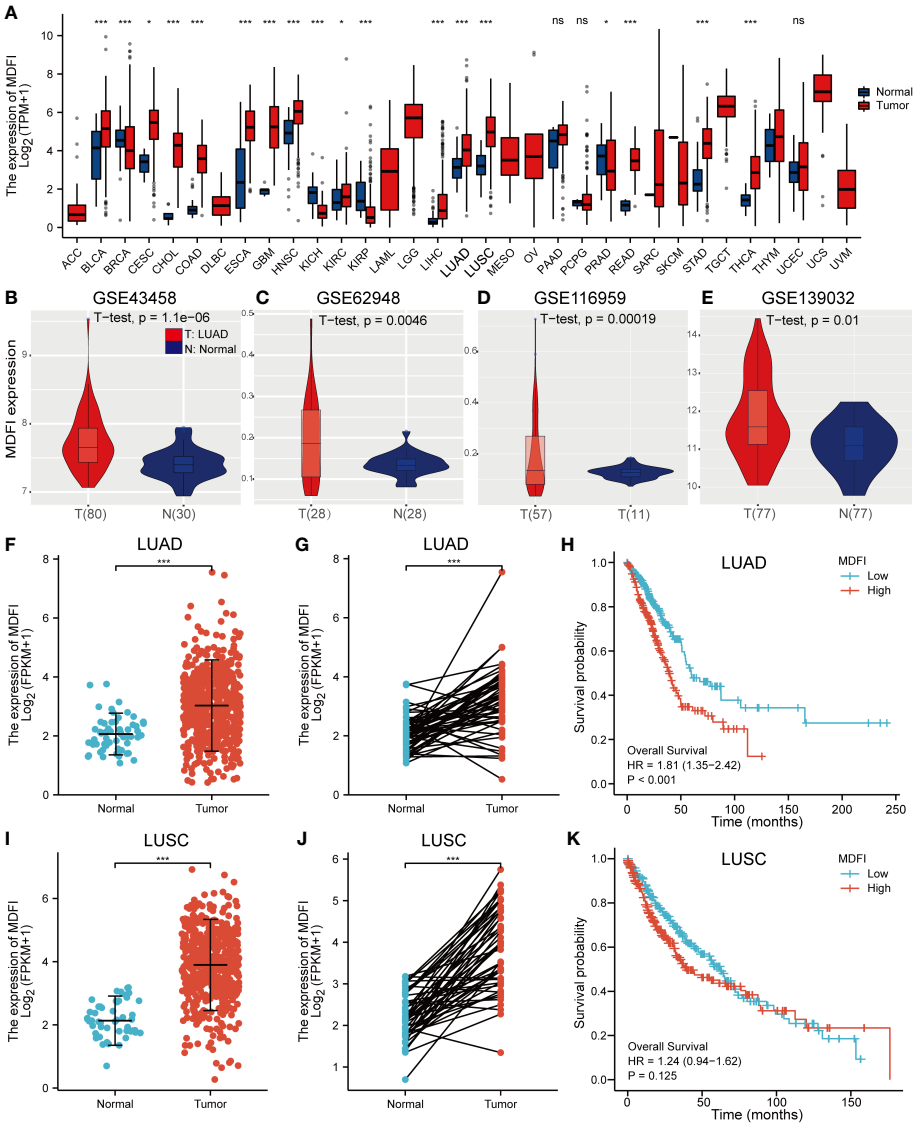


FIGURE 1
Analysis of MDFI expression in tumors, **(A)** Pan-cancer analysis of MDFI, **(B–E)** Expression analysis of MDFI in 4 sets of GEO datasets, **(F)** Expression analysis of MDFI in unpaired samples of LUAD, **(G)** MDFI in the LUAD expression analysis in paired sample, **(H)** Prognostic analysis of the survival of MDFI in LUAD, **(I–K)** Expression analysis of MDFI in LUSC. **(A, F–K)** Data from TCGA database. *p < 0.05, ***p < 0.001. ns, no significance.

TABLE 1 Baseline Table of 535 NSCLC Patients in TCGA Database.

Characteristic	levels	Overall
n		535
Gender, n (%)	Female	286 (53.5%)
	Male	249 (46.5%)
Age, n (%)	<=65	255 (49.4%)
	>65	261 (50.6%)
Race, n (%)	Asian	7 (1.5%)
	Black or African American	55 (11.8%)
	White	406 (86.8%)
Smoker, n (%)	No	75 (14.4%)
	Yes	446 (85.6%)
Pathologic stage, n (%)	Stage I	294 (55.8%)
	Stage II	123 (23.3%)
	Stage III	84 (15.9%)
	Stage IV	26 (4.9%)
T stage, n (%)	T1	175 (32.9%)
	T2	289 (54.3%)
	T3	49 (9.2%)
	T4	19 (3.6%)
N stage, n (%)	N0	348 (67.1%)
	N1	95 (18.3%)
	N2	74 (14.3%)
	N3	2 (0.4%)
M stage, n (%)	M0	361 (93.5%)
	M1	25 (6.5%)
Primary therapy outcome, n (%)	PD	71 (15.9%)
	SD	37 (8.3%)
	PR	6 (1.3%)
	CR	332 (74.4%)
Anatomic neoplasm subdivision, n (%)	Left	205 (39.4%)
	Right	315 (60.6%)
Anatomic neoplasm subdivision2, n (%)	Central Lung	62 (32.8%)
	Peripheral Lung	127 (67.2%)
OS event, n (%)	Alive	343 (64.1%)
	Dead	192 (35.9%)
DSS event, n (%)	Alive	379 (76%)
	Dead	120 (24%)
PFI event, n (%)	Alive	309 (57.8%)
	Dead	226 (42.2%)
Age, median (IQR)		66 (59, 72)
number_pack_years_smoked, median (IQR)		37 (20, 50)

PFI, Progression Free Interval; OS, Overall Survival; DSS, Disease-Specific Survival; PD, progressive disease; SD, stable disease; PR, partial response; CR, complete Response.

female (53.5%). The median age of all patients was 66 years. In terms of LUAD pathologic stage, 294 patients were stage I (55.8%), 123 patients were stage II (23.3%), 84 patients were stage III (15.9%), and 26 patients were stage IV (4.9%). Subjects included 406 White (86.8%) and 62 non-White (13.2%). In terms of primary treatment outcome, PD + SD were 108 (24.2%), PR+CR were 338 (75.8%).

Immunohistochemical analysis

MDFI-specific antibody staining was performed on the two groups of TMAs constructed in advance, we could see various staining situations and plot 60 pairs of tumor and para-tumor immunostaining heatmap (Figures 2A–G). Meanwhile, we selected one column each of invasive breast cancer and clear

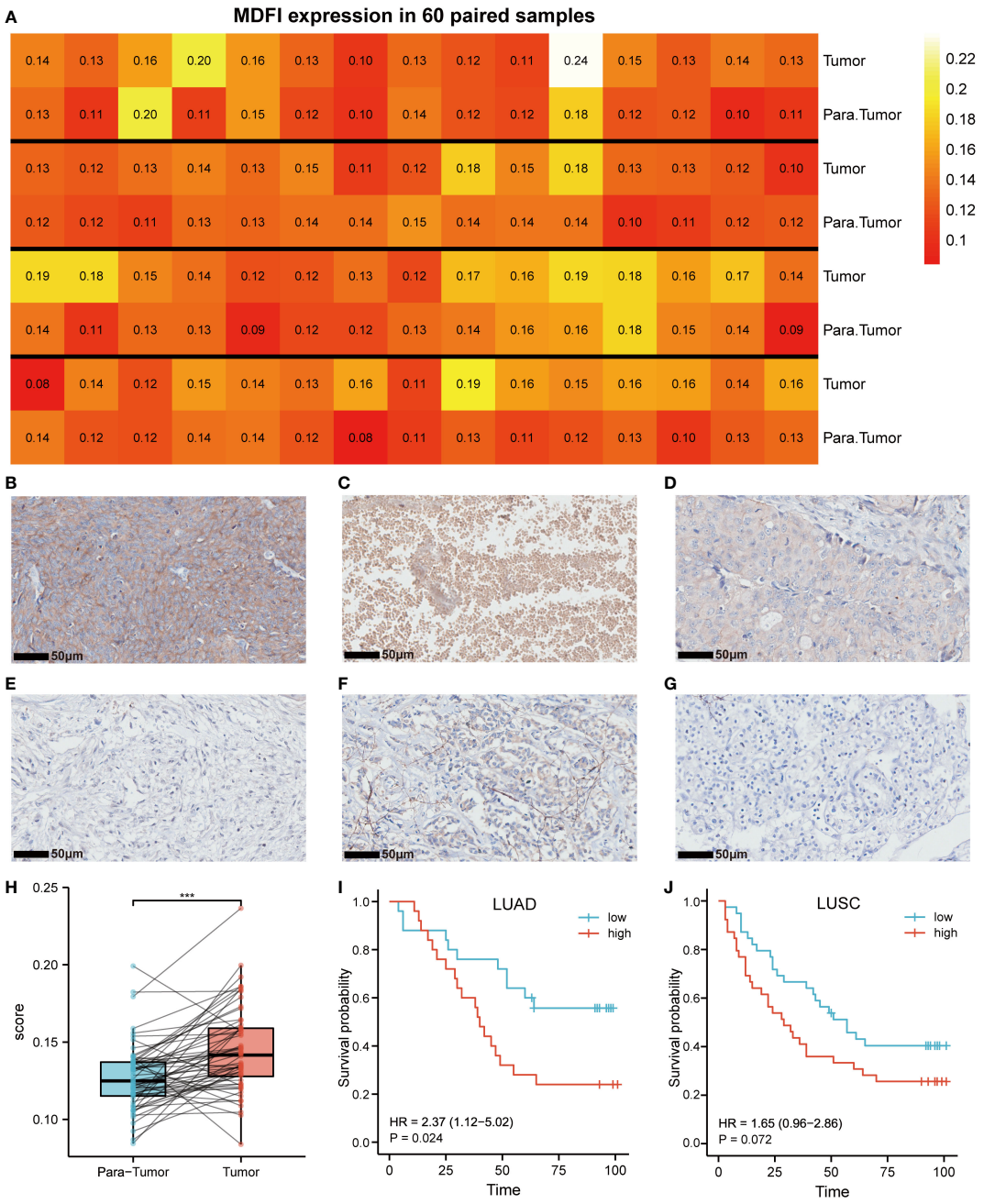


FIGURE 2
Expression of MDFI in tumor tissues. **(A)** Heat map of MDFI expression in 60 pairs of samples, **(B)** High expression of MDFI in lung cancer tissues, **(C)** MDFI is moderately expressed in lung cancer tissues, **(D)** Low expression of MDFI in lung cancer tissues, **(E)** Unstained normal tissue, **(F)** MDFI staining in invasive breast cancer tissue, **(G)** MDFI staining in clear cell renal cell carcinoma tissue, **(H)** Box plot of MDFI expression in 60 pairs of paired samples, **(I)** Prognastic analysis of MDFI in LUAD specimens, **(J)** prognostic analysis of MDFI in LUSC specimens, ***p < 0.001.

cell renal cell carcinoma as the reference sample for the same staining procedure (Figures 2F–G). The staining results were processed by IPP6 software to obtain MOD values. By MOD analysis of 60 pairs of tumor and para-tumor, the results suggested that the expression of *MDFI* in tumors was indeed higher than that in normal tissues ($P < 0.001$) (Figure 2H). Further, the K-M curve suggested that high-expressing *MDFI* had a worse prognosis in LUAD (n = 51, $P = 0.024$), whereas the difference was not significant in LUSC (n = 80, $P = 0.072$) (Figures 2I–J). Our experimental results similarly corroborate

previous results predicted by the TCGA database. High *MDFI* expression can be used as an independent prognostic marker in LUAD.

Correlation between *MDFI* expression and clinical features

In LUAD patients, the relationship between *MDFI* and clinical characteristics is summarized in Table 2. Logistic regression analysis showed that *MDFI* gene expression is a categorical dependent variable associated with poor prognostic clinical features. High expression of *MDFI* was significantly associated with T stage (P =0.013), N stage (P =0.004), Pathologic Stage (P <0.001) and Primary therapy outcome (P =0.025) (Table 2). High expression of *MDFI* was significantly correlated with clinical stage (P <0.001), T stage (P <0.001), PFI event (P <0.05), OS event (P <0.001) and DSS event (P <0.01) (Figures 3A–H).

High expression of *MDFI* is an independent risk factor for overall survival

High *MDFI* expression was associated with poor prognosis, as shown in Kaplan-Meier survival analysis (Figures 3I–P). Analysis of different clinical characteristics showed that high *MDFI* expression was significantly associated with poor prognosis in LUAD patients with T1&T2 stage (p=0.004), N0&N1 stage (P =0.002), M0 stage (P =0.001), less than 65 years (P =0.005), Primary therapy outcome: CR (P =0.003), smoker (P =0.001), pathologic stage I&II (P =0.005) and female (P =0.003) (Figures 3B–I). Univariate Cox analysis demonstrated that high *MDFI* expression was significantly correlated with poor overall survival (P <0.001). Multivariate Cox analysis confirmed MDIF gene expression was an independent risk factor for overall survival in patients with LUAD (P =0.021) (Table 3).

Diagnostic value of *MDFI* expression in LUAD

We performed ROC curve analysis of *MDFI* gene expression data to evaluate the diagnostic value of this gene. The AUC area was 0.771 (CI =0.724-0.818). These results indicate that *MDFI* expression has certain value in the diagnosis of LUAD. A nomogram was constructed to predict the 1-, 3-, and 5-year survival probability of patients in combination with the expression level of *MDFI*, TNM stage, and pathological stage (Figures 3Q, R).

Immune infiltration by *MDFI*

To investigate the relationship between *MDFI*-associated immune cells and immune function, we screened the immune cell marker genes associated with the *MDFI* high expression group at P <0.05 (Figure 4A). We found that high *MDFI* expression was associated with most immune cells, with significant marker genes for Effector memory CD8 T cells and Monocyte, such as *TRIB2* and *MARCKSL1*.

Immuno-infiltration analysis was then performed in the *MDFI* high and low expression groups, in which aDCs (P < 0.001), iDCs (P < 0.001), Mast cells (P < 0.001), Neutrophils (P < 0.001), T helper cells (P < 0.01) and TIL (P < 0.05) showed differential expression (Figure 4B). In terms of immune function, HLA (P < 0.001), Para-inflammation

TABLE 2 Logistic analysis of the correlation between *MDFI* expression and clinical characteristics.

Characteristics	Total (N)	Odds Ratio (OR)	*P* value
T stage (T2&T3&T4 vs. T1)	532	0.630 (0.436-0.906)	0.013
N stage (N1&N2&N3 vs. N0)	519	0.582 (0.401-0.843)	0.004
M stage (M1 vs. M0)	386	0.685 (0.291-1.550)	0.370
Pathologic stage (Stage II&Stage III&Stage IV vs. Stage I)	527	0.540 (0.381-0.764)	<0.001
Gender (Male vs. Female)	535	1.262 (0.898-1.775)	0.180
Age (>65 vs. <=65)	516	1.205 (0.853-1.704)	0.289
Residual tumor (R2&R1 vs. R0)	372	0.331 (0.092-0.956)	0.057
Anatomic neoplasm subdivision (Right vs. Left)	520	1.209 (0.851-1.721)	0.290
number_pack_years_smoked (>=40 vs. <40)	369	0.762 (0.506-1.147)	0.193
Smoker (Yes vs. No)	521	0.807 (0.492-1.317)	0.393
Primary therapy outcome (PR&CR vs. PD&SD)	446	1.653 (1.069-2.573)	0.025
Anatomic neoplasm subdivision2 (Peripheral Lung vs. Central Lung)	189	1.156 (0.629-2.131)	0.640
Race (White vs. Asian&Black or African American)	468	1.088 (0.637-1.863)	0.757

PD, progressive disease; SD, stable disease; PR, partial response; CR, complete response.

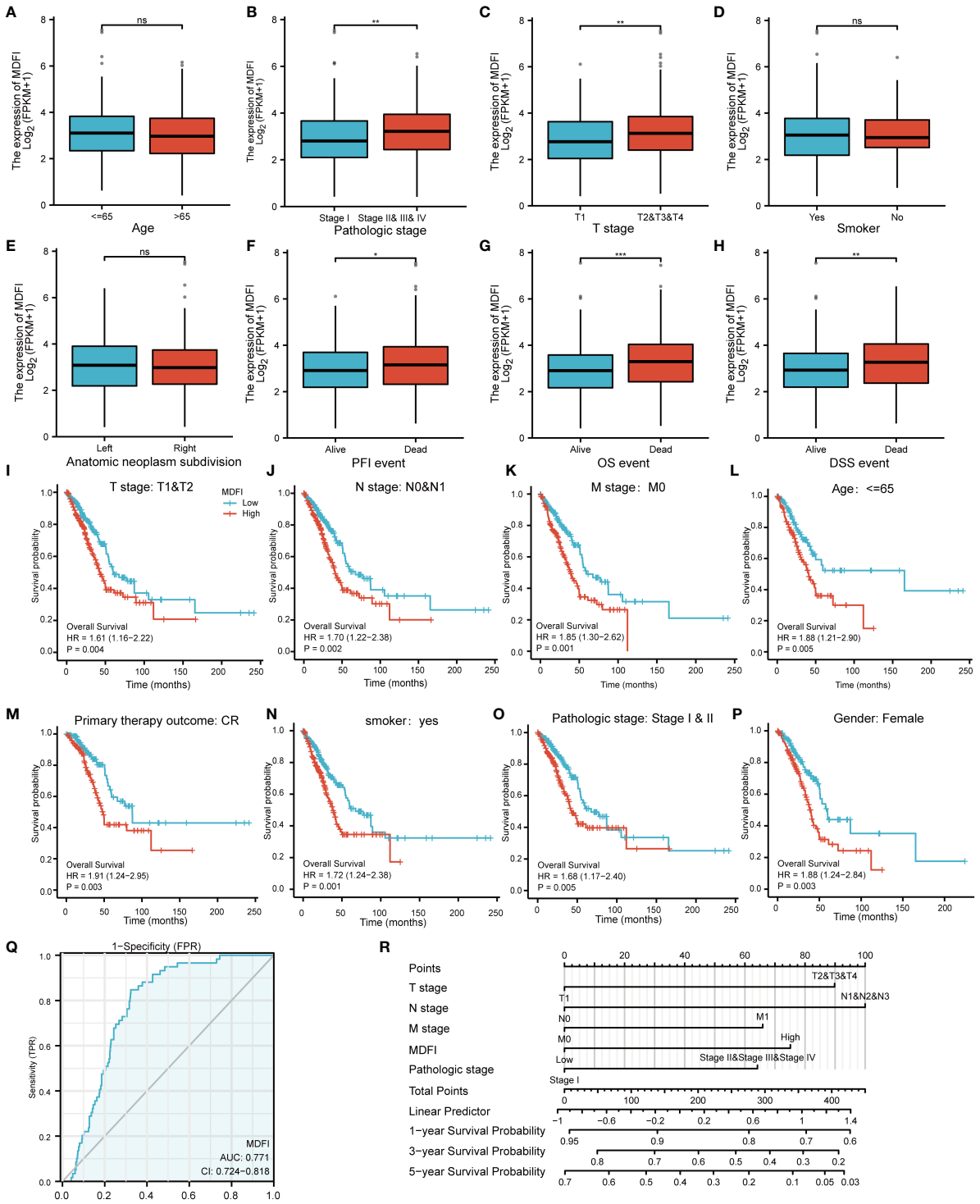


FIGURE 3
Association of MDFI expression with different LUAD characteristics and prognosis, date from TCGA, **(A–H)** Expression of MDFI in relation to different characteristics of LUAD, **(I–P)** Association between high MDFI expression in prognosis of LUAD with different characteristics, **(Q)** MDFI's ROC curve analysis at LUAD, **(R)** Construction of a nomogram to assess the prognostic value of MDFI in LUAD. *p < 0.05, **p < 0.01, ***p < 0.001, ns, p > 0.05. PFI, Progression Free Interval; OS, Overall Survival; DSS, Diseases-Specific Survival; CR, complete response.

TABLE 3 Univariate and multivariate Cox regression analyses of clinical characteristics associated with overall survival.

Characteristics	Total(N)	Univariate analysis		Multivariate analysis	
		Hazard ratio (95% CI)	*P* value	Hazard ratio (95% CI)	*P* value
T stage (T3&T4 vs. T1&T2)	523	2.317 (1.591-3.375)	**<0.001**	1.589 (0.878-2.875)	0.126
N stage (N2&N3 vs. N0&N1)	510	2.321 (1.631-3.303)	**<0.001**	1.934 (0.800-4.671)	0.143
M stage (M1 vs. M0)	377	2.136 (1.248-3.653)	**0.006**	1.593 (0.597-4.248)	0.352
Pathologic stage (Stage III&Stage IV vs. Stage I&Stage II)	518	2.664 (1.960-3.621)	**<0.001**	1.111 (0.440-2.804)	0.824
Gender (Male vs. Female)	526	1.070 (0.803-1.426)	0.642		
Age (>65 vs. <=65)	516	1.223 (0.916-1.635)	0.172		
Race (Asian&Black or African American vs. White)	468	0.678 (0.415-1.109)	0.121		
Smoker (Yes vs. No)	512	0.894 (0.592-1.348)	0.591		
Anatomic neoplasm subdivision (Right vs. Left)	512	1.037 (0.770-1.397)	0.810		
Anatomic neoplasm subdivision2 (Peripheral Lung vs. Central Lung)	182	0.913 (0.570-1.463)	0.706		
Primary therapy outcome (PR&CR vs. PD&SD)	439	0.377 (0.268-0.530)	**<0.001**	0.345 (0.227-0.525)	**<0.001**
MDFI (High vs. Low)	526	1.892 (1.408-2.542)	**<0.001**	1.606 (1.075-2.398)	**0.021**

PD, progressive disease; SD, stable disease; PR, partial response;, CR, complete response; P, <0.05.
P values in bold in Table 3 represent P <0.05.

($P < 0.01$) and Type II IFN Reponse ($P < 0.05$) were significant (Figure 4C).

Enrichment analysis of *MDFI*-related genes

To investigate the potential role of *MDFI* in LUAD, GO analysis and KEGG enrichment were used to analyze the function of *MDFI* differential genes. Among them, GO: BP/CC/MF showed 827 *MDFI* high expression group-related genes mainly acting in cornification, keratinization, cornified extracellular matrix and envelope structural constituent (Figures 5A, B). GSEA enrichment results were consistent with the direction of the above analysis, with up-regulated genes showing significant performance in regulation of actin cytoskeleton (P <0.001), focal adhesion (P <0.001) and extracellular matrix (ECM) receptor interaction (P =0.001) (Figures 5C–E).

MDFI-related gene co-expression analysis

We performed co-expression analysis of genes closely related to *MDFI* expression to characterize genes associated with *MDFI* expression in the LUAD cohort (Figure 6A). Further detailed analysis of the association of cellular Matrix Links and intermediate filament bioactivity-related genes with their *MDFI* revealed several prominent genes, such as *TNS4*, *ITGB4*, *PL AUR*, *HMGA1*, and *FSCN1*. Most of these genes are positively correlated with *MDFI* expression, and these genes have an impact on both LUAD and prognosis (Figures 6B–H). Next, the interaction networks of these co-expressed genes were visually analyzed to embody the association of these genes and biological functions. The network of these gene interactions suggests that the significantly associated biological functions are Cell differentiation, cytoskeleton, regulation of cell differentiation and positive regulation of cellular process (Figure 6I).

Relationship between somatic mutations and *MDFI* expression in LUAD

To investigate the critical role of *MDFI* in tumor progression and tumor cell dissemination and metastasis, we explored whether the distribution of mutations in the LUAD cohort was influenced by *MDFI* gene expression. We collected mutation profiles from the LUAD cohort in the TCGA database and plotted waterfalls according to MDIF expression. Interestingly, the top five mutated genes in the two groups of differential genes were the same *TP53* (57%; 42%), *TTN* (46%; 40%), *MUC16* (42%; 39%), *CSMD3* (42%; 38%) and *RYR2* (39%; 29%) (Supplementary Figures A, B).

Mutated genes were more frequent in the high-expression group, including *MTCL1*, *HYDIN*, *DCHS1*, *FRMD3*, *TP53*, *SEMA3D*, *BTAF1*, *ENPP2*, *BNIP5*, *NPTX2*, *TANC1* and *DPYS*, whereas the low-expression group included *ZNF268*, *SIGLEC10* and *OLFM4* (Figure 7B).

Then we performed a visual analysis of mutations in the pathways involved in both sets of genes, including the number of genes mutated in the pathway and the number of samples mutated (Figures 8A, B). It can be seen that in terms of

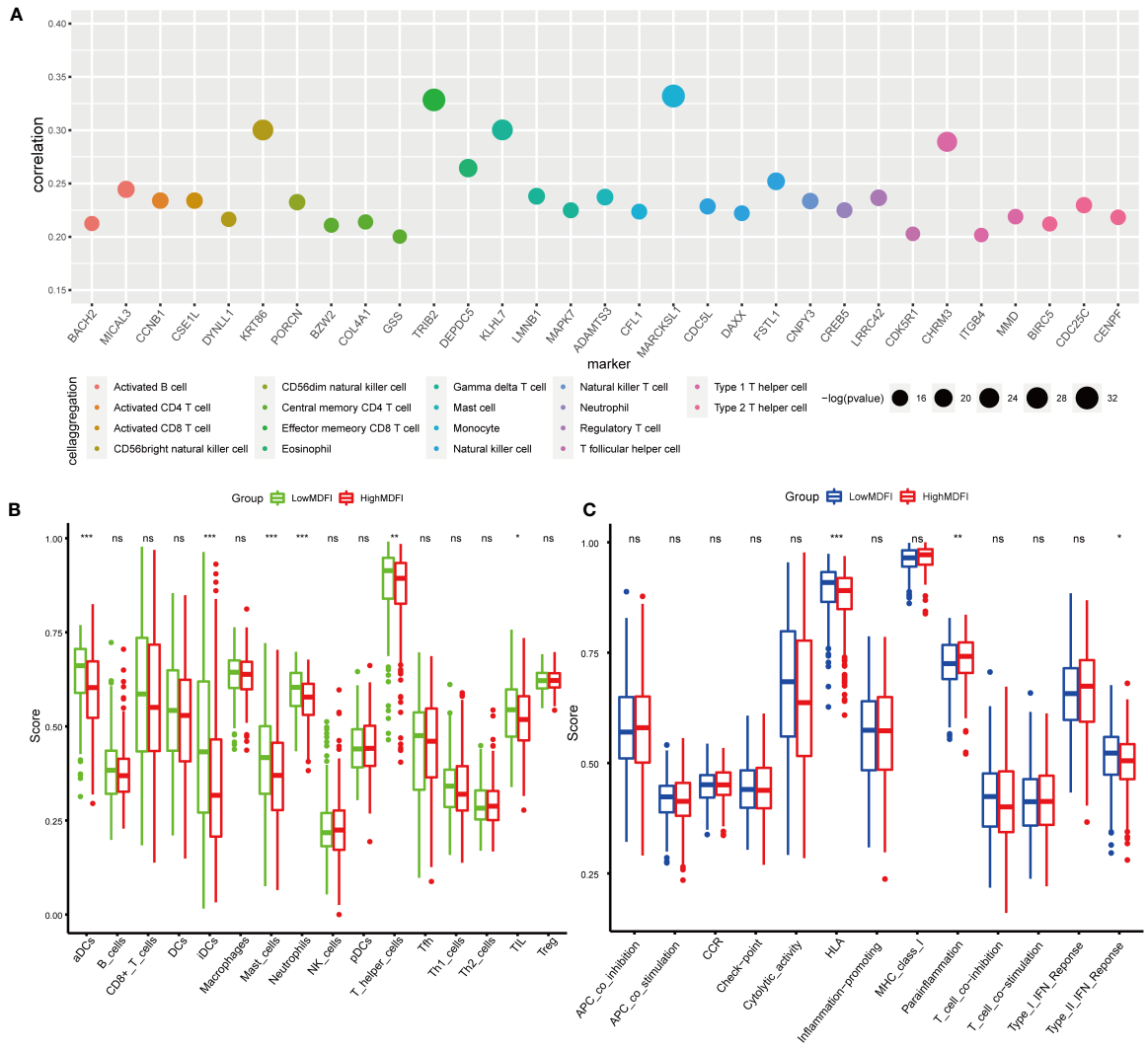


FIGURE 4
Immuno-infiltration analysis of MDFI in LUAD, data from TCGA, **(A)** Maker genes of immune cells associated with MDFI, **(B)** Relationship between the expression profile of MDFI and immune cells, **(C)** Relationship between the expression profile of MDFI and immune function. *p < 0.05, **p < 0.01, ***p < 0.001, ns, p > 0.05.

Pathway, the mutation in the *MDFI* low group was still lower than that in the *MDFI* high, and then the RTK-RAS pathway and WNT pathway of the first two pathways in the *MDFI* high group were separately plotted as waterfalls (Figures 8C, D).

Predicted drugs

Bazedoxifene (adj. *P* =0.027), Pentadecafluorooctanoic acid (adj. *P* =0.0313), Hexachlorobiphenyl (adj. *P* =0.0313), Pentachlorphenyl (adj. *P* =0.0313), Arbutin (adj. *P* =0.0313), Kojic acid (adj. *P* =0.0313) and Nickel sulfate (adj. *P* =0.049371) were screened for potential drugs that may be beneficial in lung adenocarcinoma using protein-drug interaction data from the DSigDB database, and may be effective in LUAD patients with *MDFI* overexpression (Table 4).

Discussion

MDFI (Myogenic differentiation Family Inhibitor) is a Protein Coding gene (39). This protein is a transcription factor that negatively regulates other myogenic family proteins (11). It is highly expressed in a variety of cancer tissues, including Liver hepatocellular carcinoma, Breast invasive carcinoma, Colon adenocarcinoma/Rectum adenocarcinoma, etc (12–14). There

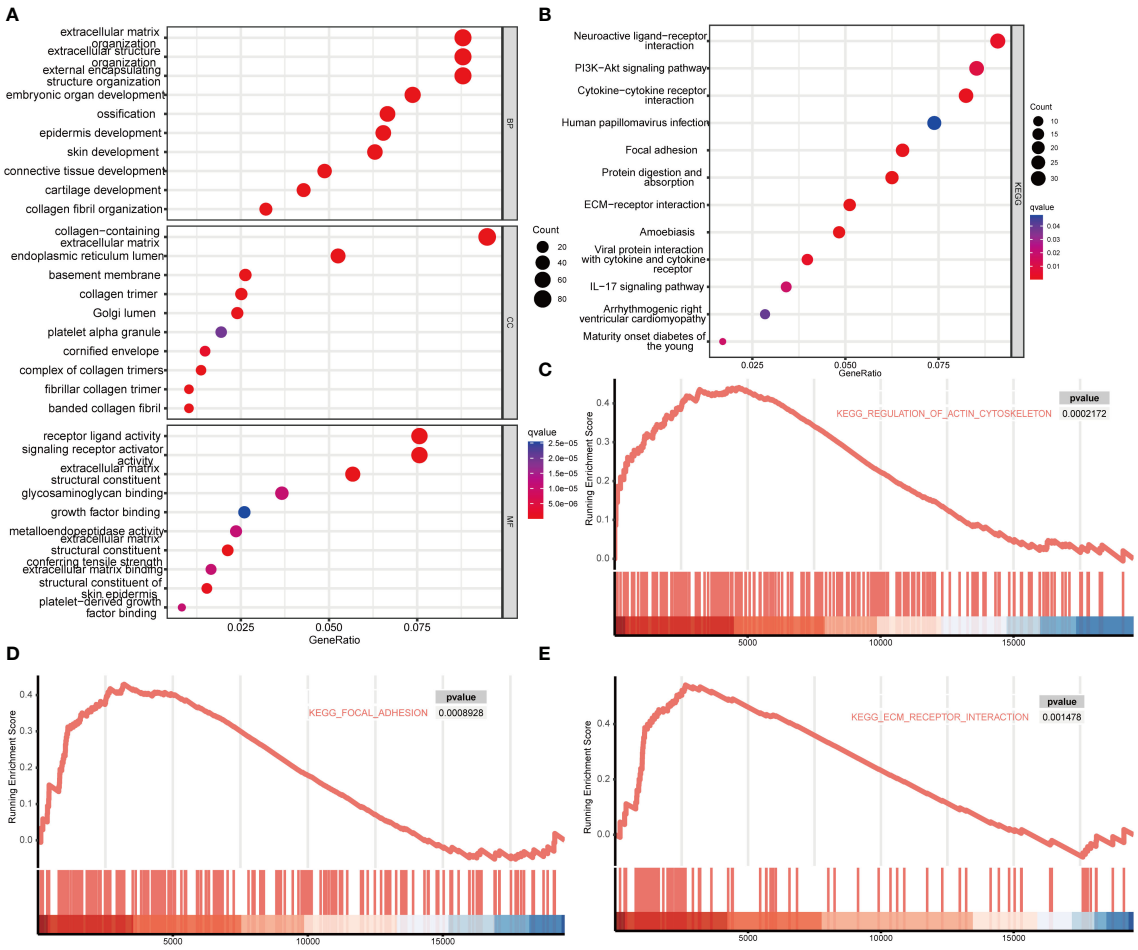


FIGURE 5
Enrichment analysis of MDFI in LUAD, data from TCGA. **(A)** GO. BP/CC/MF enrichment analysis, **(B)** KEGG enrichment analysis, **(C–E)** GSEA enrichment analysis.

is currently limited literature on the potential prognostic impact of *MDFI* in NSCLC. Therefore, we conducted a study on the potential role of *MDFI* in NSCLC, analyzing *MDFI* expression in NSCLC cohorts for the first time. In the context of TCGA data, we retrospectively analyzed 535 LUAD patients. The experimental results highlight that in LUAD patients, *MDFI* expression was significantly higher in tumors than in para-tumor. Similarly, it was validated in 4 GEO datasets and external immunohistochemistry experiments. Further, the expression levels of *MDFI* also differed in clinical stage, T stage, PFI event and OS event. We found that both RNA-seq data and TMA results showed that the expression level of *MDFI* was related to clinical prognosis. LUAD patients with high *MDFI* expression have a worse prognosis. Moreover, the K-M curves of female patients aged < 65 years with earlier tumor stage also supported the above results. Similarly, multivariate analysis showed that *MDFI* can be used as an independent prognostic factor in LUAD and is associated with poor prognosis in LUAD. Moreover, the AUC in our plotted ROC curve indicates that *MDFI* can be used as an indicator to predict LUAD.

The analysis of CIBERSORT showed that the expression level of *MDFI* was positively correlated with the infiltration level of most immune cells in LUAD, including CD4 T cells, Activated CD8 T cells, Effector memory CD8 T cells and Monocyte. The marker genes of these immune cells play an important role in tumor immune response, immune escape, proliferation, migration and invasion, promoting the progression of this LUAD. *TRIB2* and *MARCKSL1* are the most prominent marker genes in these immune cells. Critical role of *TRIB2* in cancer and drug resistance to therapy, *TRIB2* interacts with *MAPKK*, *AKT* and *NFkB* proteins and participates in cell survival, proliferation and immune responses (40, 41). Ectopic or intrinsic high expression of *TRIB2* induces drug resistance by promoting phosphorylation of *AKT* through its *COP1* structural domain. significantly increased expression of *TRIB2* in tumor tissue correlates with increased phosphorylation of *AKT*,

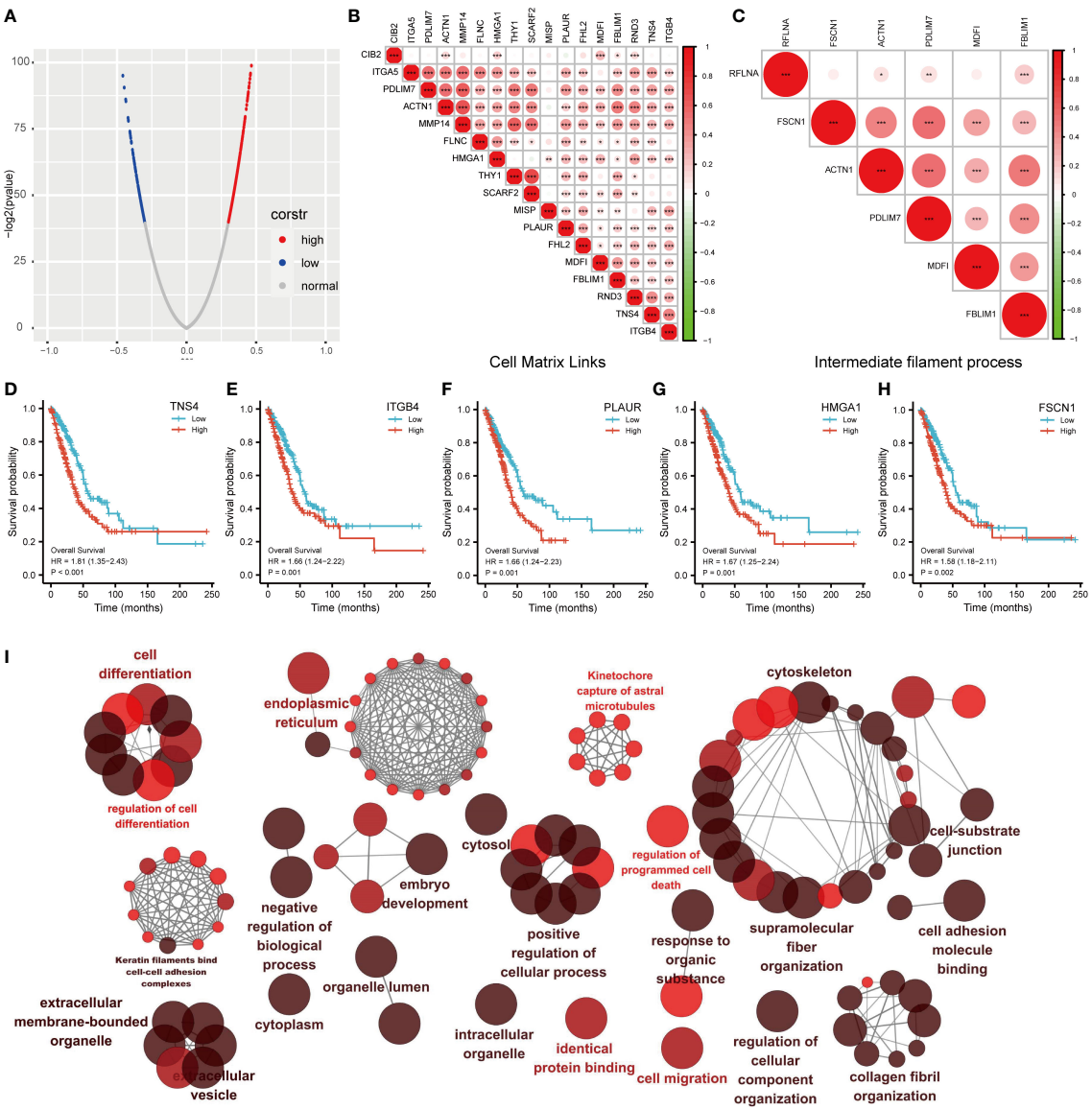


FIGURE 6
Correlation analysis of MDFI in LUAD **(A)** MDFI-related genes in LUAD, **(B)** Heat map of MDFI-related genes in Cell Matrix Links, **(C)** Heat map of MDFI-related genes in intermediate filament process, **(D–H)** Prognostic analysis of some MDFI-related genes in LUAD, **(I)** Visualization of the interaction network of genes strongly associated with MDFI by Cytoscape (ClueGo module). Node size indicates the mapped gene number; the node color schedule represents the p value, the darker the color, the smaller the p-value. **(A–H)** Data from TCGA. *p < 0.05, **p <0.01, ***p < 0.001.

FOXO3a, *MDM2* and impaired treatment response. This ultimately led to extremely poor clinical outcomes (42). When *MARCKSL1* phosphorylation is inhibited, actin mobility is increased, filamentous sodium formation is impaired, and laminar lipid formation is enhanced, as is cell migration, and we speculate that the same process may be going on in LUAD, thereby promoting tumor cell proliferation, migration, and invasion (43). *MARCKSL1* promoted the progression of lung adenocarcinoma by regulating epithelial–mesenchymal transition (EMT) (44). These results suggest that *MDFI* may play an important role in immune escape in the LUAD microenvironment (45). Moreover, it can be seen that the HLA presentation pathway is more pronounced in the low *MDFI* expression group, while tumor cells are more likely to evade immune detection in the absence of the HLA presentation pathway, thereby promoting tumor progression (46).

Keratin, the major intermediate filament protein of epithelial cells, and the cytoskeleton play multiple key roles in the cell, from cell migration to organelle dynamics (47, 48). They not only have a positive biological role in tumor progression and

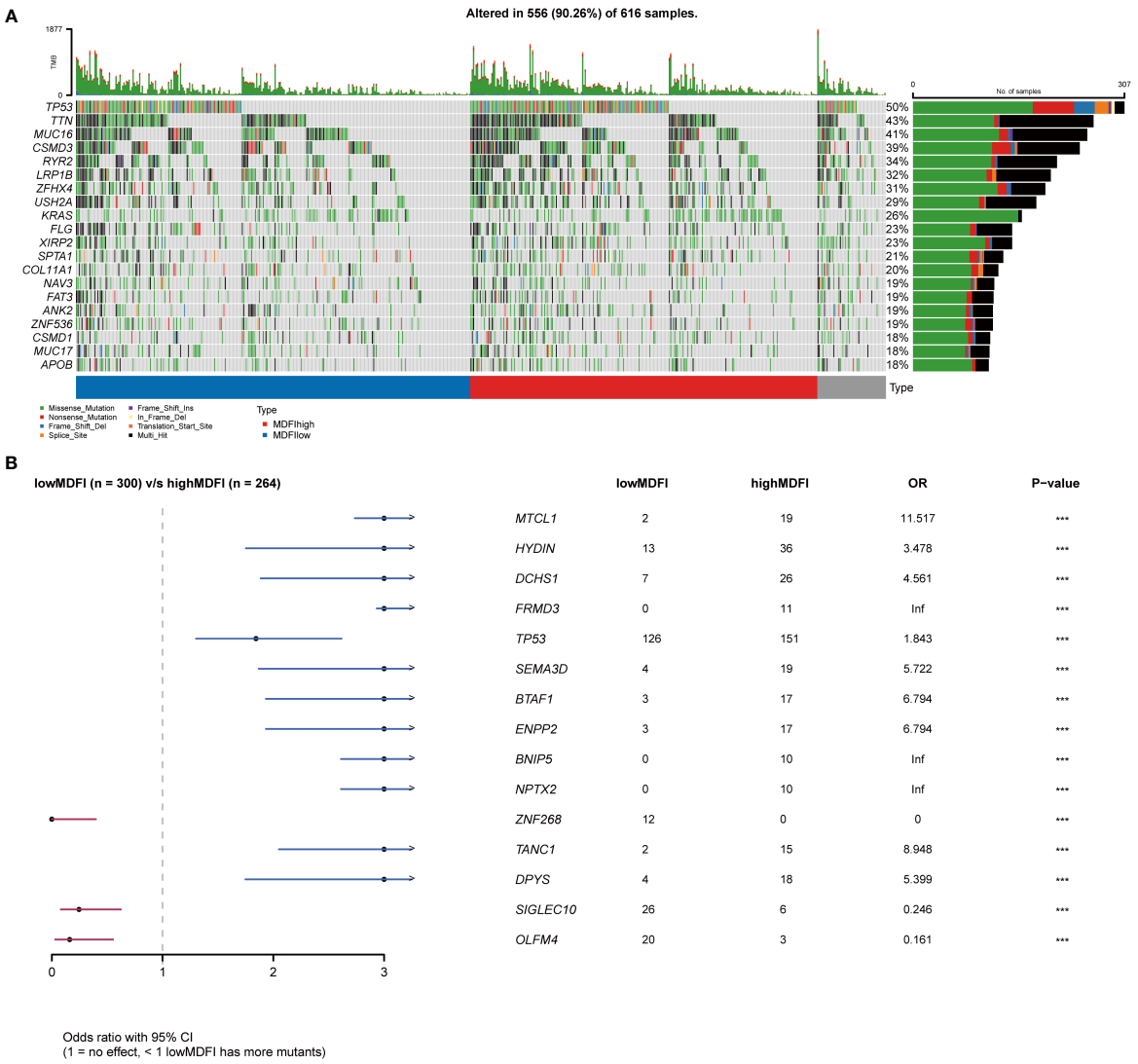


FIGURE 7
Relationship between MDFI and LUAD somatic mutation, data from TCGA. **(A)** Somatic mutation in MDFI-low and MDFI-high expression groups, **(B)** Comparison of mutation in MDFI high expression group and low expression group. ***p < 0.001.

tumor cell dissemination and metastasis, but also often maintain their specific expression pattern during malignant transformation of cancer, so they are also widely used as tumor markers in cancer diagnosis. Both GO/KEGG enrichment and GSEA analysis indicate that *MDFI* has a prominent performance in the above biological processes. Keratins act as epithelial cell markers, which makes their role in cancer progression, diagnosis and treatment an important focus of research. Among them, keratin 1(K1) can act as a cell surface receptor in cancer, and KEGG enrichment analysis also shows that cytokine receptor action is more significant (49, 50); while keratin 17 (K17) plays a role in DNA damage response and tumor initiation. Moreover, keratin is an essential element of the cytoskeleton in normal and malignant epithelial cells (49, 51). Cancers often maintain their specific keratin expression pattern during malignant transformation, and therefore keratin is widely used as a tumor marker in cancer diagnosis. Keratin plays an active biological role in tumor cell dissemination and metastasis (52).

We speculate that detection and treatment for *MDFI* may allow earlier diagnosis of tumors and limit further tumor growth. Several genes that were highly correlated and positively correlated in the subsequent co-expression analysis of *MDFI* showed the same differential performance for the prognosis of LUAD, suggesting that low expression is better for the prognosis. It further strengthens the important role and predictive value of

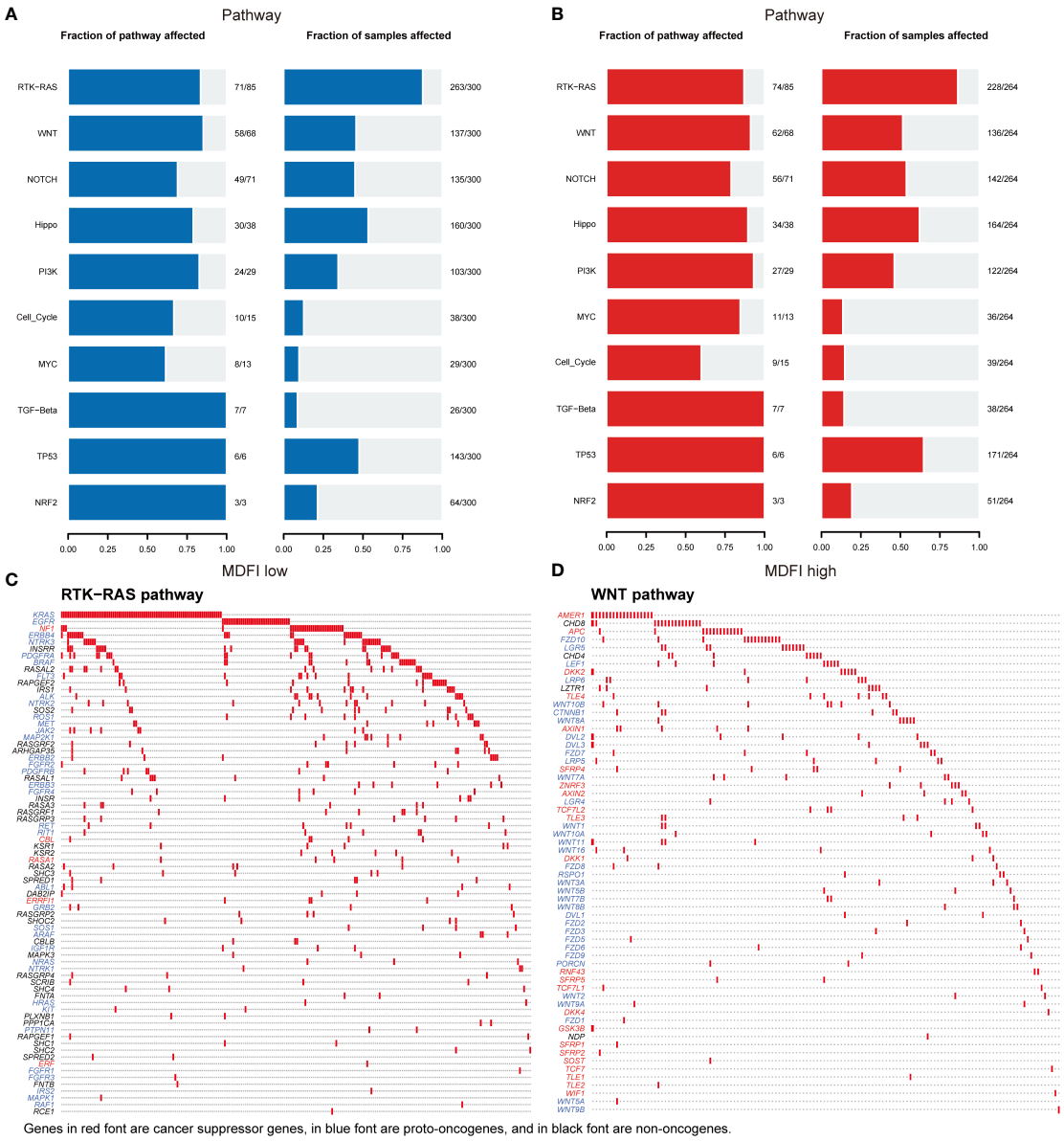


FIGURE 8
MDFI in LUAD involves mutations in the pathway data from TCGA **(A)** Pathway mutations of MDFI-low in LUAD, **(B)** Pathway mutations of MDFI-high in LUAD, **(C)** Mutations involved in RTK-RAS pathway, **(D)** Mutations involved in the WNT pathway.

MDFI in the prognosis of LUAD. The network of these gene interactions suggests Cell differentiation, cytoskeleton, regulation of cell differentiation and positive regulation of cellular process that are considered to be important process processes of tumor progression (53). In the analysis of somatic mutations in the high and low groups, the proportion of the top five mutated genes in the low expression group was lower than that in the high expression group, and it was speculated that the tumor cells with high expression of *MDFI* produced a large number of DNA replication errors in the proliferation and spread, indicating a worse prognosis (54). We mapped the RTK-RAS pathway and the WNT pathway of the first two pathways in the high *MDFI* group as waterfalls, respectively, LUAD-derived Wnts increase the proliferation/stemness potential of cancer cells, and LUAD cells use paracrine Wnt1 signaling to induce immune resistance, which could provide a new pathway therapeutic option for LUAD with high *MDFI* expression (55–57). Finally, we predicted that Bazedoxifene performed most prominently in *MDFI*-related LUAD. Bazedoxifene, or combined with oxaliplatin, significantly

TABLE 4 Predict potential drug.

Term	Overlap	*P*	adj. *P*	Combined Score
Bazedoxifene CTD 00004022	1/45	0.00225	0.027	121662.4978
Pentadecafluorooctanoic acid CTD 00001078	1/203	0.01015	0.0313	90874.00322
2,2',4,4',5,5'-Hexachlorobiphenyl CTD 00000731	1/208	0.0104	0.0313	90369.46872
3,3',4,4',5-Pentachlorobiphenyl CTD 00001077	1/237	0.01185	0.0313	87657.53732
arbutin CTD 00005438	1/282	0.0141	0.0313	84030.01469
kojic acid CTD 00000624	1/313	0.01565	0.0313	81844.61948
NICKEL SULFATE CTD 00001417	1/576	0.0288	0.049371	68904.42357

adj. P: adjusted P –value.

induced apoptosis, inhibited cell viability, colony formation, and cell migration in colon cancer cells, and we speculated that it may have the same effect in LUAD (58).

The above data and experimental results all suggest that *MDFI* is a closely related gene in LUAD, but our study still has some limitations. Our exploration of the role of *MDFI* in LUAD is based on data already available in GEO and TCGA databases, coupled with external data collected for validation and the specimens are old, and there are individual unstained conditions. And, no *in vivo* and *in vitro* experiments were performed to further verify the role of *MDFI* in immune escape and proliferation and spread of tumors, which also points the direction for future work.

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 Research Ethics Committee of Changzhou Second People's Hospital Affiliated to Nanjing Medical University. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

KY designed the research. XS and ZG performed the research and collected the data. ZQ analyzed the data. PC drafted the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was funded by Changzhou Sci & Tech Program (Grant number: CZ20220025), Changzhou High-Level Medical Talents Training Project (Grant number: 2022CZBJ069), "333 Project" of Jiangsu Province (Grant number: BRA2020157), 333 High-Level Talent Training Project (Grant number: 2022-2), "Six One Project," Research Projects of High-level Medical Personnel of Jiangsu Province (Grant number: LGY2019025), High-level Talent Selection and Training Project of the 16th Batch of "Six Talent Peak" in Jiangsu Province (Grant number: WSN-245), Medical Scientific Research Foundation of Jiangsu Commission of Health (Grant number: H2018083).

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.2022.1005962/full#supplementary-material

References

1. Thai AA, Solomon BJ, Sequist LV, Gainor JF, Heist RS. Lung cancer. *Lancet* (2021) 398(10299):535–54. doi: 10.1016/S0140-6736(21)00312-3

2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* (2018) 68(6):394–424. doi: 10.3322/caac.21492

3. Nagy-Mignotte H, Guillem P, Vesin A, Toffart AC, Colonna M, Bonneterre V, et al. Primary lung adenocarcinoma: Characteristics by smoking habit and sex. *Eur Respir J* (2011) 38(6):1412–9. doi: 10.1183/09031936.00191710

4. Barta JA, Powell CA, Wisnivesky JP. Global epidemiology of lung cancer. *Ann Glob Health* (2019) 85(1):1–16. doi: 10.5334/aogh.2419

5. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, et al. International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. *J Thorac Oncol* (2011) 6(2):244–85. doi: 10.1097/JTO.0b013e318206a221

6. Devarakonda S, Morgensztern D, Govindan R. Genomic alterations in lung adenocarcinoma. *Lancet Oncol* (2015) 16(7):e342–51. doi: 10.1016/S1470-2045(15)00077-7

7. Gao S, Li N, Wang S, Zhang F, Wei W, Li N, et al. Lung cancer in people's Republic of China. *J Thorac Oncol* (2020) 15(10):1567–76. doi: 10.1016/j.jtho.2020.04.028

8. Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WE, et al. The IASLC lung cancer staging project: Proposals for revision of the TNM stage groupings in the forthcoming (Eighth) edition of the TNM classification for lung cancer. *J Thorac Oncol* (2016) 11(1):39–51. doi: 10.1016/j.jtho.2015.09.009

9. Pignon JP, Tribodet H, Scagliotti GV, Douillard JY, Shepherd FA, Stephens RJ, et al. Lung adjuvant cisplatin evaluation: A pooled analysis by the LACE collaborative group. *J Clin Oncol* (2008) 26(21):3552–9. doi: 10.1200/JCO.2007.13.9030

10. European Virology 2000. 17-21 September 2000, Glasgow, scotland. Abstracts. *J Clin Virol* (2000) 18(1-3):1–330. doi: 10.1016/S1386-6532(00)00098-6.

11. Chen CM, Kraut N, Groudine M. Weintraub h. I-mf, a novel myogenic repressor, interacts with members of the MyoD family. *Cell* (1996) 86(5):731–41. doi: 10.1016/s0092-8674(00)80148-8

12. Lin ZY, Chuang WL. Genes responsible for the characteristics of primary cultured invasive phenotype hepatocellular carcinoma cells. *BioMed Pharmacother* (2012) 66(6):454–8. doi: 10.1016/j.biopha.2012.04.001

13. Soran A, Ozbas S, Ozcinar B, Isik A, Dogan L, Senol K, et al. Intervention for hepatic and pulmonary metastases in breast cancer patients: Prospective, multi-institutional registry study-IMET, protocol MF 14-02. *Ann Surg Oncol* (2022) 29(10): 6337–38. doi: 10.1245/s10434-022-12320-7

14. Sui Y, Li X, Oh S, Zhang B, Freeman WM, Shin S, et al. Opposite roles of the JMJD1A interaction partners MDFI and MDFIC in colorectal cancer. *Sci Rep* (2020) 10(1):8710. doi: 10.1038/s41598-020-65536-6

15. Kabbout M, Garcia MM, Fujimoto J, Liu DD, Woods D, Chow CW, et al. ETS2 mediated tumor suppressive function and MET oncogene inhibition in human non-small cell lung cancer. *Clin Cancer Res* (2013) 19(13):3383–95. doi: 10.1158/1078-0432.CCR-13-0341

16. Mansfield AS, Wang L, Cunningham JM, Jen J, Kolbert CP, Sun Z, et al. DNA Methylation and RNA expression profiles in lung adenocarcinomas of never-smokers. *Cancer Genet* (2015) 208(5):253–60. doi: 10.1016/j.cancergen.2014.12.002

17. Enfield KSS, Marshall EA, Anderson C, Ng KW, Rahmati S, Xu Z, et al. Epithelial tumor suppressor ELF3 is a lineage-specific amplified oncogene in lung adenocarcinoma. *Nat Commun* (2019) 10(1):5438. doi: 10.1038/s41467-019-13295-y

18. Moreno Leon L, Gautier M, Allan R, Ilie M, Nottet N, Pons N, et al. The nuclear hypoxia-regulated NLUCAT1 long non-coding RNA contributes to an aggressive phenotype in lung adenocarcinoma through regulation of oxidative stress. *Oncogene* (2019) 38(46):7146–65. doi: 10.1038/s41388-019-0935-y

19. The cancer genome atlas (2022). Available at: https://cancergenome.nih.gov/.

20. Hutter C, Zenklusen JC. The cancer genome atlas: Creating lasting value beyond its data. *Cell* (2018) 173(2):283–5. doi: 10.1016/j.cell.2018.03.042

21. Zhao J, Guo C, Ma Z, Liu H, Yang C, Li S. Identification of a novel gene expression signature associated with overall survival in patients with lung adenocarcinoma: A comprehensive analysis based on TCGA and GEO databases. *Lung Cancer* (2020) 149:90–6. doi: 10.1016/j.lungcan.2020.09.014

22. Wittekind C, Brierley JD, Gospodarowicz MK. *TNM classification of malignant Tumours 2017*. (2016) New Jersey: Wiley-Blackwell.

23. Liu J, Tian Y, Yi L, Gao Z, Lou M, Yuan K. High KIF11 expression is associated with poor outcome of NSCLC. *Tumori* (2022) 108(1):40–6. doi: 10.1177/0300891620988342

24. Waldvogel HJ, Curtis MA, Baer K, Rees MI, Faull RL. Immunohistochemical staining of post-mortem adult human brain sections. *Nat Protoc* (2006) 1(6):2719–32. doi: 10.1038/nprot.2006.354

25. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* (2018) 173(2):400–16 e11. doi: 10.1016/j.cell.2018.02.052

26. Wickham H. *ggplot2: Elegant graphics for data analysis*. (2016). New York: Springer-Verlag

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

28. Morgan M, Anders S, Lawrence M, Aboyoun P, Pages H, Gentleman R. ShortRead: A bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics*. (2009) 25(19):2607–8. doi: 10.1093/bioinformatics/btp450

29. Adler J, Parmryd I. Recent review on colocalization seem to misunderstand the Pearson correlation coefficient. *J Microsc* (2007) 227(Pt 1):83. doi: 10.1111/j.1365-2818.2007.01788.x

30. Liu XS, Zhou LM, Yuan LL, Gao Y, Kui XY, Liu XY, et al. NPM1 is a prognostic biomarker involved in immune infiltration of lung adenocarcinoma and associated with m6A modification and glycolysis. *Front Immunol* (2021) 12:724741. doi: 10.3389/fimmu.2021.724741

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

32. Gene Ontology (2022). Available at: http://www.geneontology.org. [Accessed July 2022]

33. Kyoto Encyclopedia Genes Genomes. (2022). Available at: http://www.genome.jp/kegg/pathway.html. [Accessed July 2022]

34. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* (2005) 102(43):15545–50. doi: 10.1073/pnas.0506580102

35. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* (2021) 2(3):100141. doi: 10.1016/j.xinn. 2021.100141

36. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res* (2018) 28(11):1747–56. doi: 10.1101/gr.239244.118

37. Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD, Cibulskis K, et al. Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* (2008) 455(7216):1069–75. doi: 10.1038/nature07423

38. DSigDB Database (2022). Available at: http://dsigdb.tanlab.org/DSigDBv1.0/. [Accessed July 2022]

39. Davis RL, Weintraub H, Lassar AB. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* (1987) 51(6):987–1000. doi: 10.1016/0092-8674(87)90585-X

40. Mayoral-Varo V, Jimenez L, Link W. The critical role of TRIB2 in cancer and therapy resistance. *Cancers (Basel)* (2021) 13(11):1–17. doi: 10.3390/cancers13112701

41. Yokoyama T, Nakamura T. Tribbles in disease: Signaling pathways important for cellular function and neoplastic transformation. *Cancer Sci* (2011) 102(6):1115–22. doi: 10.1111/j.1349-7006.2011.01914.x

42. Hill R, Madureira PA, Ferreira B, Baptista I, Machado S, Colaco L, et al. TRIB2 confers resistance to anti-cancer therapy by activating the serine/threonine protein kinase AKT. *Nat Commun* (2017) 8:14687. doi: 10.1038/ncomms14687

43. Bjorkblom B, Padzik A, Mohammad H, Westerlund N, Komulainen E, Hollos P, et al. C-jun n-terminal kinase phosphorylation of MARCKSL1 determines actin stability and migration in neurons and in cancer cells. *Mol Cell Biol* (2012) 32(17):3513–26. doi: 10.1128/MCB.00713-12

44. Liang W, Gao R, Yang M, Wang X, Cheng K, Shi X, et al. MARCKSL1 promotes the proliferation, migration and invasion of lung adenocarcinoma cells. *Oncol Lett* (2020) 19(3):2272–80. doi: 10.3892/ol.2020.11313

45. Zhang C, Zhang J, Xu FP, Wang YG, Xie Z, Su J, et al. Genomic landscape and immune microenvironment features of preinvasive and early invasive lung adenocarcinoma. *J Thorac Oncol* (2019) 14(11):1912–23. doi: 10.1016/j.jtho.2019.07.031

46. Hazini A, Fisher K, Seymour L. Deregulation of HLA-I in cancer and its central importance for immunotherapy. *J Immunother Cancer* (2021) 9(8):1–17. doi: 10.1136/jitc-2021-002899

47. Schiavon CR, Zhang T, Zhao B, Moore AS, Wales P, Andrade LR, et al. Actin chromobody imaging reveals sub-organellar actin dynamics. *Nat Methods* (2020) 17(9):917–21. doi: 10.1038/s41592-020-0926-5

48. Ehrlich F, Fischer H, Langbein L, Praetzel-Wunder S, Ebner B, Figlak K, et al. Differential evolution of the epidermal keratin cytoskeleton in terrestrial and aquatic mammals. *Mol Biol Evol* (2019) 36(2):328–40. doi: 10.1093/molbev/msy214

49. Nair RR, Hsu J, Jacob JT, Pineda CM, Hobbs RP, Coulombe PA. A role for keratin 17 during DNA damage response and tumor initiation. *Proc Natl Acad Sci USA* (2021) 118(13):1–10. doi: 10.1073/pnas.2020150118

50. Ogunnigbagbe O, Bunick CG, Kaur K. Keratin 1 as a cell-surface receptor in cancer. *Biochim Biophys Acta Rev Cancer* (2022) 1877(1):188664. doi: 10.1016/j.bbcan.2021.188664

51. Perone Y, Farrugia AJ, Rodriguez-Meira A, Gyorffy B, Ion C, Uggetti A, et al. SREBP1 drives keratin-80-dependent cytoskeletal changes and invasive behavior in endocrine-resistant ERalpha breast cancer. *Nat Commun* (2019) 10(1):2115. doi: 10.1038/s41467-019-09676-y

52. Werner S, Keller L, Pantel K. Epithelial keratins: Biology and implications as diagnostic markers for liquid biopsies. *Mol Aspects Med* (2020) 72:100817. doi: 10.1016/j.mam.2019.09.001

53. Otasek D, Morris JH, Boucas J, Pico AR, Demchak B. Cytoscape automation: Empowering workflow-based network analysis. *Genome Biol* (2019) 20(1):185. doi: 10.1186/s13059-019-1758-4

54. Lenormand C, Lipsker D. Somatic mutations in "Benign" disease. *N Engl J Med* (2021) 385(11):e34. doi: 10.1056/NEJMc2110545

55. Reck M, Carbone DP, Garassino M, Barlesi F. Targeting KRAS in non-small-cell lung cancer: Recent progress and new approaches. *Ann Oncol* (2021) 32 (9):1101–10. doi: 10.1016/j.annonc.2021.06.001

56. Skronska-Wasek W, Gosens R, Konigshoff M, Baarsma HA. WNT receptor signalling in lung physiology and pathology. *Pharmacol Ther* (2018) 187:150–66. doi: 10.1016/j.pharmthera.2018.02.009

57. Kerdidani D, Chouvardas P, Arjo AR, Giopanou I, Ntaliarda G, Guo YA, et al. Wnt1 silences chemokine genes in dendritic cells and induces adaptive immune resistance in lung adenocarcinoma. *Nat Commun* (2019) 10(1):1405. doi: 10.1038/s41467-019-09370-z

58. Wei J, Ma L, Lai YH, Zhang R, Li H, Li C, et al. Bazedoxifene as a novel GP130 inhibitor for colon cancer therapy. *J Exp Clin Cancer Res* (2019) 38(1):63. doi: 10.1186/s13046-019-1072-8

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 13 October 2022
DOI 10.3389/fonc.2022.1021084

Check for updates

OPEN ACCESS

EDITED BY
Hongda Liu,
Nanjing Medical University, China

REVIEWED BY
Yuqing Huang,
Beijing Haidian Hospital, Peking University, China
Shouliang Qi,
Northeastern University, China

*CORRESPONDENCE
Fan Yang
yangfan@pkuph.edu.cn

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

SPECIALTY SECTION
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

RECEIVED 17 August 2022
ACCEPTED 26 September 2022
PUBLISHED 13 October 2022

CITATION
Chen X, Xu H, Qi Q, Sun C, Jin J, Zhao H, Wang X, Weng W, Wang S, Sui X, Wang Z, Dai C, Peng M, Wang D, Hao Z, Huang Y, Wang X, Duan L, Zhu Y, Hong N and Yang F (2022) AI-based chest CT semantic segmentation algorithm enables semi-automated lung cancer surgery planning by recognizing anatomical variants of pulmonary vessels. *Front. Oncol.* 12:1021084. doi: 10.3389/fonc.2022.1021084

COPYRIGHT
© 2022 Chen, Xu, Qi, Sun, Jin, Zhao, Wang, Weng, Wang, Sui, Wang, Dai, Peng, Wang, Hao, Huang, Wang, Duan, Zhu, Hong and Yang. 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.

AI-based chest CT semantic segmentation algorithm enables semi-automated lung cancer surgery planning by recognizing anatomical variants of pulmonary vessels

Xiuyuan Chen[1,2†], Hao Xu[1,2†], Qingyi Qi[3], Chao Sun[3], Jian Jin[1,2], Heng Zhao[1,2], Xun Wang[1,2], Wenhan Weng[1,2], Shaodong Wang[1,2], Xizhao Sui[1,2], Zhenfan Wang[1,2], Chenyang Dai[4], Muyun Peng[5], Dawei Wang[6], Zenghao Hao[6], Yafen Huang[6], Xiang Wang[5], Liang Duan[4], Yuming Zhu[4], Nan Hong[3] and Fan Yang[1,2*]

[1]Department of Thoracic Surgery, Peking University People's Hospital, Beijing, China, [2]Thoracic Oncology Institute, Peking University People's Hospital, Beijing, China, [3]Department of Radiology, Peking University People's Hospital, Beijing, China, [4]Thoracic Surgery Department, Shanghai Pulmonary Hospital, Shanghai, China, [5]Thoracic Surgery Department, The Second Xiangya Hospital of Central South University, Changsha, China, [6]Institute of Advanced Research, Infervision Medical Technology Co., Ltd, Beijing, China

Background: The recognition of anatomical variants is essential in preoperative planning for lung cancer surgery. Although three-dimensional (3-D) reconstruction provided an intuitive demonstration of the anatomical structure, the recognition process remains fully manual. To render a semiautomated approach for surgery planning, we developed an artificial intelligence (AI)–based chest CT semantic segmentation algorithm that recognizes pulmonary vessels on lobular or segmental levels. Hereby, we present a retrospective validation of the algorithm comparing surgeons' performance.

Methods: The semantic segmentation algorithm to be validated was trained on non-contrast CT scans from a single center. A retrospective pilot study was performed. An independent validation dataset was constituted by an arbitrary selection from patients who underwent lobectomy or segmentectomy in three institutions during Apr. 2020 to Jun. 2021. The golden standard of anatomical variants of each enrolled case was obtained via expert surgeons' judgments based on chest CT, 3-D reconstruction, and surgical observation. The performance of the algorithm is compared against the performance of two junior thoracic surgery attendings based on chest CT.

Results: A total of 27 cases were included in this study. The overall case-wise accuracy of the AI model was 82.8% in pulmonary vessels compared to 78.8% and 77.0% for the two surgeons, respectively. Segmental artery accuracy was 79.7%, 73.6%, and 72.7%; lobular vein accuracy was 96.3%, 96.3%, and 92.6% by the AI model and two surgeons, respectively. No statistical significance was found. In subgroup analysis, the anatomic structure-wise analysis of the AI algorithm showed a significant difference in accuracies between different lobes ($p = 0.012$). Higher AI accuracy in the right-upper lobe (RUL) and left-lower lobe (LLL) arteries was shown. A trend of better performance in non-contrast CT was also detected. Most recognition errors by the algorithm were the misclassification of LA^{1+2} and LA^{3}. Radiological parameters did not exhibit a significant impact on the performance of both AI and surgeons.

Conclusion: The semantic segmentation algorithm achieves the recognition of the segmental pulmonary artery and the lobular pulmonary vein. The performance of the model approximates that of junior thoracic surgery attendings. Our work provides a novel semiautomated surgery planning approach that is potentially beneficial to lung cancer patients.

KEYWORDS

pulmonary vessel, artificial intelligence, semantic segmentation, surgery planning, lung cancer

Introduction

Lung cancer is one of the leading causes of cancer-related morbidity and mortality worldwide, with an estimated 2.2 million new cases and 1.8 million deaths (1). With the increased frequency of computed tomography (CT) screening, especially thin-section CT, the early detection rate of small-sized lung cancer and ground-glass opacity has dramatically increased (2). Anatomic lobectomy and segmentectomy are the main curative treatments for early-stage lung cancer, especially segmentectomy, which preserves more lung tissue (3, 4). However, pulmonary arteries and veins are highly variable; understanding the anatomical structure of each patient during preoperative surgical planning is crucial yet challenging. The misclassification of segmental or even lobular vessels can occur even for experienced surgeons, which can lead to bleeding, increased surgical resection, or other catastrophic consequences.

Traditionally, chest CT is the most common tool for preoperative planning that typically consists of three steps: 1. 3-D reconstruction; 2. variant recognition; and 3. intraoperative projection. In the first step, surgeons rely on their own spatial imagination or 3-D reconstruction software to perform a 3-D reconstruction of anatomical structures. Second, normal anatomy and anatomical variations require careful identification, which relies heavily on the experience of the surgeon. Third, the surgeon needs to project the reconstruction of the preoperative 3-D anatomy to the intraoperative anatomy, that is, the surgeon matches and identifies the anatomy seen during the operation according to the preoperative 3-D reconstruction. All three steps rely solely on human effort, which impairs the accuracy and efficiency of preoperative planning. The recent development of artificial intelligence (AI), however, has shown potential in optimizing this practice.

AI algorithms have been widely applied in every aspect of medicine recently (5–9). From the screening of the pulmonary nodule (10) to the diagnosis of skin cancer (11) and diabetic retinopathy (12) and even in the development of new treatment drugs (13–15), utilizing the AI algorithm has been revolutionizing. Among all applications, the pattern recognition of medical imaging is the most reliable, of which semantic segmentation excels due to its interpretability and robustness in highly specialized tasks.

According to literatures, semantic segmentation algorithms are capable of detecting red blood cells for sickle cell disease in microscopic images (16); deciding the tumor border in pathological images (17, 18); recognizing the infection area of coronavirus disease of 2019 (COVID-19) lesions on chest CTs (19); distinguishing the brachia plexus, fetal head, and lymph node from ultrasound images (20); segmenting the thalamus, caudate nucleus, and lenticular nucleus in brain MRI (21); and diagnosing gastrointestinal

cancer margins during endoscopy (22). Aiming to optimize the surgical planning process, we have previously developed a fully automated 3-D reconstruction algorithm (23) to classify and reconstruct the pulmonary artery and vein. The performance of the algorithm is the same as manual reconstruction. Although the granularity of these applications is coarse, the clinical significancy is solid.

In this study, we go one more step into the surgical planning process. We developed a fine-grained chest CT semantic segmentation algorithm that systematically identifies 18 segmental pulmonary arteries and 5 lobular pulmonary veins. We evaluated the independent performance of the algorithm using the CT data from 27 patients who had undergone lobectomy or segmentectomy at three medical institutions. This algorithm would facilitate the realization of a semiautomated surgical planning process, which is one of the backbones for the development of a fully automated thoracoscopic surgical system.

Methods

Patient enrollment

Patients who underwent lobectomy or segmentectomy at Peking University People's Hospital, Shanghai Pulmonary Hospital, and the Second Xiangya Hospital of Central South University between Apr. 2020 to Jun. 2021 were retrospectively reviewed. The inclusion criteria were as follows: (1) preoperative thin-section (<1.25 mm), either non-enhanced chest CT images or contrast-enhanced CT chest CT images available, (2) the time interval between CT examination and surgery of less than 1 month, and (3) the video of surgery and preoperative 3-D reconstruction of pulmonary vessels and bronchi available. Among patients who met the above criteria, 27 cases representing most of the common lobectomies and segmentectomies were arbitrarily selected from three participating centers.

Chest computed tomography acquisition

All enrolled patients underwent a chest CT examination at most a month prior to surgery. The whole lung scan from the thoracic entrance to the bottom of the lung was completed after one inhalation and holding breath using CT instruments from GE Healthcare (Chicago, IL, USA), Philips Healthcare (Amsterdam, Netherlands), Siemens Healthineers (Forchheim, Germany), and United Imaging (Shanghai, China). CT images were reconstructed by using different convolutional kernels with a layer thickness of less than 1.25 mm.

Deep learning algorithm for automated semantic segmentation of pulmonary vessels

Based on the automatic 3-D reconstruction system of pulmonary blood vessels and bronchi (InferOperate Thoracic Surgery) (23), a pulmonary blood vessel semantic segmentation system was developed using deep learning (DL) algorithms for the automatic segmentation and name of the segmental pulmonary arteries and lobular pulmonary veins, aiming to provide guidance for surgery and promote the application of anatomical lobectomy and segmentectomy. In brief, segmental arteries and lobular veins were manually and concisely segmented based on the automatically constructed 3-D blood vessels by senior thoracic surgeons and then used for training the model. The ResUNet was utilized as the backbone, and the label propagation algorithm was employed to reduce the misclassification of segmental arteries and lobular veins; a schematic roadmap is shown in Figure 1. Label propagation is usually utilized to classify a massive number of unlabeled examples in the presence of a few labeled examples (24, 25). In our scenario, the label of each pixel is propagated to adjacent pixels according to the similarity. During each step of propagation, each pixel adds the label values propagated by its surrounding pixels according to the

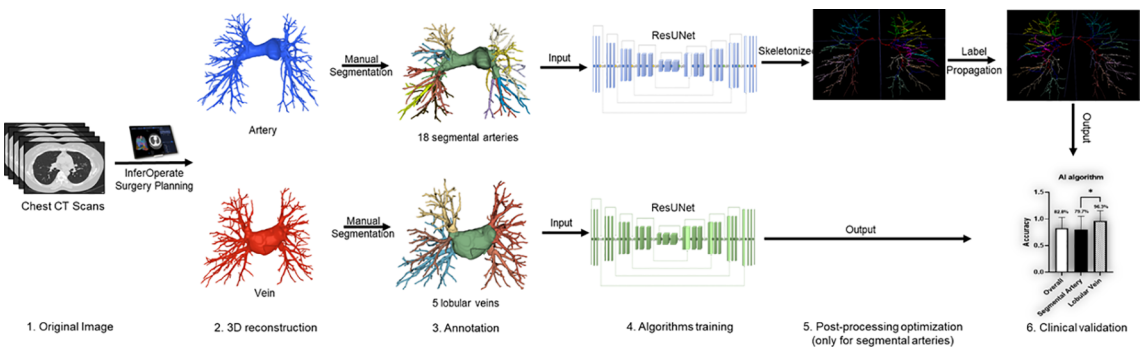


FIGURE 1
Workflow diagram of the deep learning (DL)-based pulmonary blood vessel segmentation system.

propagation probability and updates the probability distribution of their respective label types. By keeping the labels of known pixels unchanged at the initial value and then restarting a new round of propagation until each pixel of different categories is divided into a range, the color mixing of blood vessels could be well reduced. In this multicenter study, we validated the robustness and generalizability of the pulmonary blood vessel segmentation system by examining its segmentation performance on either plain CT scans or contrast-enhanced CT scans.

The golden standard for anatomical structures compositions of targeted lesions

To quantitatively evaluate the performance of the pulmonary blood vessel segmentation system, the golden standard for calculating accuracy was firstly established by three senior thoracic surgeons with CT images, surgery videos, and preoperative 3-D reconstructions as references. Taking the clinical needs into account, the anatomical structures related to target lesions and their relationships were analyzed. Two senior thoracic surgeons reviewed these cases back to back, and disagreements would be settled by a third senior thoracic surgeon. To simplify the establishment procedure, possible anatomical variants that represent the spatial pattern of relevant anatomical structures were enumerated in a table, which is inherited from our previous study with minor simplification (23) and made available to the surgeons (Supplementary Table 1).

Reader study

To comprehensively analyze the performance of the pulmonary blood vessel segmentation system, we invited two junior attendings in thoracic surgery at Peking University People's Hospital to participate in a reader study in comparison to the AI algorithm. The recognition by both the algorithm and junior attendings is based on chest CT. The recognized anatomical variant is selected or described based on Supplementary Table 1. By comparing with the golden standard, recognition accuracy was calculated to evaluate their performance and compare it with the DL-based system.

Evaluation index and statistical analysis

Accuracy was utilized as the evaluation index in the independent performance test and reader study, including anatomical structure-wise accuracy and case-wise accuracy. The former was defined as the correctly segmented/recognized targeted pulmonary vessel structures in the 3-D reconstruction divided by the total related structures, while the latter was achieved by averaging the anatomical structure-wise accuracy of each case. Patients' demographics and clinical features (age, sex, and smoking history), tumor characteristics (tumor location, tumor size, and histology), and surgery characteristics (blood loss and operation time) were analyzed. Continuous data with normality distribution and the homogeneity of variance were analyzed using one-way ANOVA; otherwise, the Mann–Whitney test or Kruskal–Wallis test was utilized. Meanwhile, the categorical variables were processed by chi-square or Fisher's exact tests as appropriate.

Results

Clinical characteristics

To validate the performance and practice of the DL-based semantic segmentation algorithm, we arbitrarily selected 27 patients who underwent lobectomy or segmentectomy from Apr. 2020 to Jun. 2021 in three medical centers, among which nine were enrolled from Peking University People's Hospital, nine were enrolled from Shanghai Pulmonary Hospital, and nine were enrolled from The Second Xiangya Hospital of Central South University. The clinical characteristics are delineated in Table 1. The median age was 59 with the interquartile range (IQR) from 53.5 to 64.5 years. A total of 14 patients (51.85%) were women, and 13 (48.15%) were men. Most of the enrolled patients were non-smokers (88.89%). Five of the enrolled cases were pathologically diagnosed with benign lesions and 22 with malignant ones (MIA or invasive adenocarcinoma). Regarding the location of lesions, 10 RUL, 4 right-middle lobe (RML), 5 right-lower lobe (RLL), 6 left-upper lobe (LUL), and 2 LLL were included. With respect to surgical procedures, 10 received segmentectomy and 17 underwent lobectomy. No significant differences between baseline data were found across the three institutions except for blood loss (30 ml vs. 50 ml vs. 50 ml, $p = 0.006$). All enrolled cases were successfully processed by the semantic segmentation algorithm without systemic failures. The outputs of the algorithm are demonstrated in Supplementary Figures 1–6. The median inference time was 100 s.

Performance of the semiautomated surgical planning algorithm

The independent performance of the DL-based semantic segmentation algorithm in classifying segmental arteries and lobular veins was first validated. The overall case-wise accuracy

10.3389/fonc.2022.1021084

TABLE 1 Clinical characteristics of the validation dataset.

Variable	Total	Center A	Center B	Center C	p-value
Number of cases, n	27	9	9	9	
Age, median (IQR), years	59 [40, 75]	62 (49–63)	58 (55–63)	57 (53–67)	0.981
Sex, n(%)					0.145
Female	14 (51.9)	3 (33.3)	4 (44.4)	7 (77.8)	
Male	13 (48.1)	6 (66.7)	5 (55.6)	2 (22.2)	
Smoking history, n(%)					
Yes	3 (11.1)	1 (11.1)	1 (11.1)	1 (11.1)	1
No	24 (88.9)	8	8	8	
Histology, n(%)					0.267
Benign lesion	5 (18.5)	0 (0)	2 (22.2)	3 (33.3)	
MIA*	7 (25.9)	4 (44.4)	1 (11.1)	2 (22.2)	
Invasive adenocarcinoma	15 (55.6)	5 (55.6)	6 (66.7)	4 (44.4)	
Tumor location**					0.569
RUL	10 (37.0)	4 (44.4)	1 (11.1)	5 (55.6)	
RML	4 (14.8)	1 (11.1)	2 (22.2)	1 (11.1)	
RLL	5 (18.5)	1 (11.1)	3 (33.3)	1 (11.1)	
LUL	6 (22.2)	2 (22.2)	3 (33.3)	1 (11.1)	
LLL	2 (7.4)	1 (11.1)	0 (0)	1 (11.1)	
Surgery type, n(%)					
Segmentectomy	10 (37)				
Lobectomy	17 (63)				
Blood loss, median (IQR), ml	50 [20, 50]	30 [20, 50]	50 [40, 50]	50 [50, 50]	0.006
Operation time, median (IQR), min	110 [60, 195]	120 [110, 146]	100 [90, 120]	110 [100, 120]	0.355
CT instrument manufacturer					<0.05
GE	6 (22.2)	1 (11.1)	5 (55.6)	0 (0)	
Siemens	15 (55.6)	6 (66.7)	0 (0)	9 (100)	
Philips	4 (14.8)	0 (0)	4 (44.4)	0 (0)	
UIH	2 (7.4)	2 (22.2)	0 (0)	0 (0)	
Chest CT slice thickness, mm					0.072
0.625	3 (11.1)	0 (0)	3 (33.3)	0 (0)	
0.8	1 (3.7)	0 (0)	0 (0)	1 (11.1)	
1	20 (74.1)	8 (88.9)	4 (44.4)	8 (88.9)	
1.25	3 (11.1)	1 (11.1)	2 (22.2)	0 (0)	
Imaging convolutional kernels					<0.05
stnd	3 (11.1)	1 (11.1)	2 (22.2)	0 (0)	
Lung	3 (11.1)	0 (0)	3 (33.3)	0 (0)	
B70f	7 (25.9)	0 (0)	0 (0)	7 (77.8)	
Br40d\3	5 (18.5)	5 (55.6)	0 (0)	0 (0)	
Br60f\3	1 (3.7)	0 (0)	0 (0)	1 (11.1)	
BI64d\3	1 (3.7)	1 (11.1)	0 (0)	0 (0)	
BI57d\2	1 (3.7)	0 (0)	0 (0)	1 (11.1)	
Lung iDose(3)	1 (3.7)	0 (0)	1 (11.1)	0 (0)	
Lung iDose(4)	3 (11.1)	0 (0)	3 (33.3)	0 (0)	
Soft B	2 (7.4)	2 (22.2)	0 (0)	0 (0)	

* Microinvasive adenocarcinoma.

** RUL, right-upper lobe; RML, right-middle lobe; RLL, right-lower lobe; LUL, left- upper lobe; LLL, left-lower lobe.

of the algorithm was 82.8% (Figure 2A). In the segmental artery recognition task, the accuracy was 79.7% (Figure 2A). Higher accuracy was seen in non-contrast CT (83.8% vs. 65.3%, p = 0.094) and segmentectomies (91.7% vs. 72.7%, p = 0.053) when compared to contrast CT scans and lobectomies, respectively (Figure 2B). Arteries in the LLL showed the highest recognition accuracy at 100.0%, followed by the RUL (90.0%), the RML (87.5%), the RLL (70.5%), and the LUL (58.3%) (Figure 2C). In the lobular vein recognition task, the accuracy was 96.3% (Figure 2A). In subgroup analyses, the accuracy reached 95.2% and 100.0% on non-contrast CT and contrast CT scans, respectively; similar accuracies (94.1% vs. 100.0%) were observed for lobectomy and segmentectomy cases (Figure 2D). Of note, the RUL, RLL, LUL, and LLL reached 100.0% accuracy while in the RML, the accuracy was lower (75.0%) (Figure 2E).

Performance of junior thoracic surgery attendings

Two junior attendings participated in the reader study for recognizing pulmonary vascular anatomical structures and obtained the overall accuracy of 78.8% and 77.0%, respectively (Figure 3A). In the segmental artery recognition task, the accuracy was 73.6% and 72.7%, respectively. Accuracies slightly favored non-contrast over contrast CT (76.0% vs. 65.3%, p = 0.355; 76.0% vs. 61.1%, p = 0.245) and favored segmentectomy over lobectomy (78.3% vs. 70.8%, p = 0.473; 83.3% vs. 66.4%, p = 0.074) (Figure 3B). No significant difference across lobes were shown (RUL 66.7%, RML 100.0%, RLL 65.7%, LUL 70.8%, LLL 83.3%, p = 0.247; RUL 76.7%, RML 75.0%, RLL 65.7%, LUL 66.7%, LLL 83.3%, p = 0.823) (Figure 3C). In the lobular vein recognition task, the accuracy was 96.3% and 92.6%, respectively (Figure 3A). No significant difference between non-contrast and contrast CT was shown (Figure 3D). The accuracy of the RLL was 80% in both surgeons and LUL 83.3% in surgeon B (Figure 3E). All other lobular vein recognitions reached 100.0% accuracy (Figure 3E).

Performance comparison between the artificial intelligence algorithm and surgeons' performance

The case-wise artery recognition accuracy showed no significant difference between the AI algorithm and two surgeons in three-way analysis (Figure 4A). Upon further pairwise analysis, surgeon A showed significantly lower accuracy in the RUL (p = 0.037) (Figure 4B). In non-contrast CT, contrast CT, lobectomy and segmentectomy subgroups, no significant differences were shown across the AI algorithm and two surgeons (p = 0.493, 0.825, 0.872, 0.396) (Figure 4C). Different CT manufacturers, slice thicknesses, or convolutional kernels showed no significant impact on the AI algorithm and two surgeons (Figure 5).

Error analysis of artificial intelligence and human recognition

Error analyses were performed by regrouping the results at the segmental structure level instead of the case level. All recognition errors by the algorithm are listed in Supplementary Figures 7–18. Most errors were observed in segmental arteries, while few errors were seen in lobular veins. The analysis of the AI recognition showed significant differences across different segmental arteries (p = 0.002) (Table 2), among which RA^{1}, RA^{2}, RA^{3}, RA^{4+5}, RA^{6}, LA^{6}, and LA^{8-10} showed higher accuracies. Failure in distinguishing LA^{1+2} and LA^{3} was the most frequently seen error in the algorithm; accuracy in right-lower basal segmental arteries was also below 50%. The error of two surgeons was analyzed on a structure-wise basis

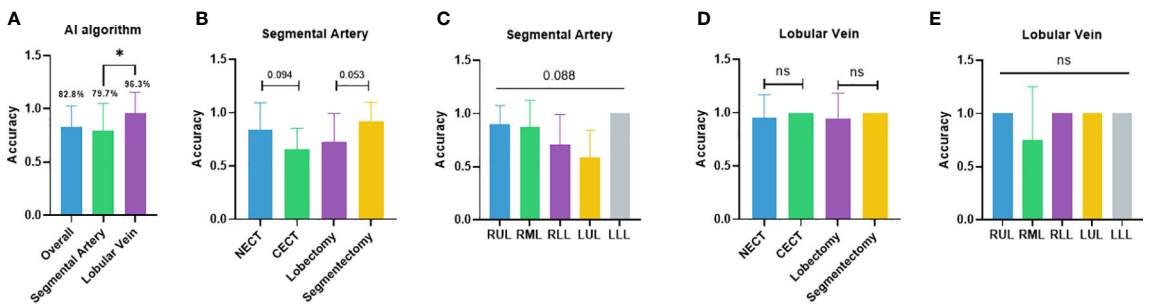


FIGURE 2
Performance of the semiautomated surgical planning algorithm. **(A)** Case-wise overall accuracy and accuracy in arteries and veins; **(B)** artery accuracy comparison between contrast and non-contrast CT, lobectomy and segmentectomy; **(C)** artery accuracy comparison between different lobes; **(D)** vein accuracy comparison between contrast and non-contrast CT, lobectomy, and segmentectomy; **(E)** vein accuracy comparison between different lobes. NECT, non-enhanced CT; CECT, contrast-enhanced CT; RUL, right-upper lobe; RML, right-middle lobe; RLL, right-lower lobe; LUL, left-upper lobe; LLL, left- lower lobe. *p<0.05; ns non-significant.

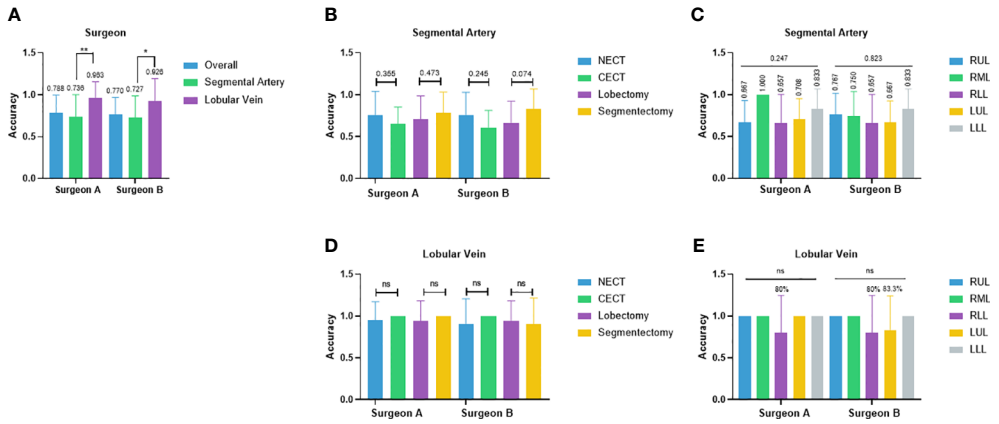


FIGURE 3
Performance of two junior thoracic surgery attendings. **(A)** Case-wise overall accuracy and accuracy in arteries and veins; **(B)** Artery accuracy comparison between contrast and non-contrast CT, lobectomy, and segmentectomy; **(C)** artery accuracy comparison between different lobes; **(D)** vein accuracy comparison between contrast and non-contrast CT, lobectomy, and segmentectomy; **(E)** vein accuracy comparison between different lobes. NECT, non-enhanced CT; CECT, contrast-enhanced CT; RUL, right-upper lobe; RML, right-middle lobe; RLL, right-lower lobe; LUL, left-upper lobe; LLL, left-lower lobe. *p<0.05, **p<0.01, ns non-significant.

(Table 2). For surgeon A, a significant difference across different segmental arteries was detected (p = 0.033), among which RA^{1} and RA^{4+5} showed higher accuracy and LA^{1+2} showed significant lower accuracy. The accuracies of RA^{2}, RA^{6}, LA^{4}, LA^{5}, and LA^{6} were also below 50%. For surgeon B, no significant differences across different segmental arteries were detected (p = 0.083). The accuracies of LA^{1+2}, LA^{4}, LA^{5}, and LA^{6} were less than 50%.

Discussion

In this study, we first reported a CT-based pulmonary vessel semantic segmentation algorithm for semiautomated surgical planning and comprehensively evaluated its clinical value by performing a validation study on multicenter datasets. Our results showed that the independent performance of the semantic segmentation algorithm approximated that of junior attendings in thoracic surgery. The overall recognition accuracy ranged from 70% to 80%. Subgroup analyses uncovered the possible influential factors on algorithm performance, including the usage of a contrast agent, radiation doses, the CT instrument manufacturer, and reconstruction convolutional kernels, and explored its potential applicable scenarios, such as lobectomy and segmentectomy.

A comprehensive understanding of the blood vessel structure, especially the artery structure, is the most crucial task in preoperative planning to minimize the chance of intraoperative bleeding and misresection (26–28). Given the fact that arteries are more hazardous than veins in surgeries

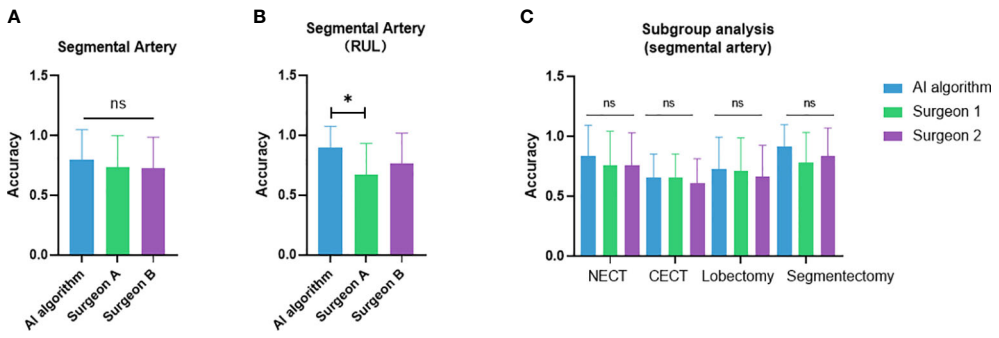


FIGURE 4
Performance comparison between the artificial intelligence algorithm and two surgeons. **(A)** Three-way analysis between three groups; **(B)** Pairwise analysis showed significantly lower accuracy in the RUL of surgeon A compared to the AI algorithm; **(C)** subgroup analysis between non-contrast vs. contrast CT and lobectomy vs. segmentectomy. NECT, non-enhanced CT; CECT, contrast-enhanced CT. *p<0.05; ns, non-significant.

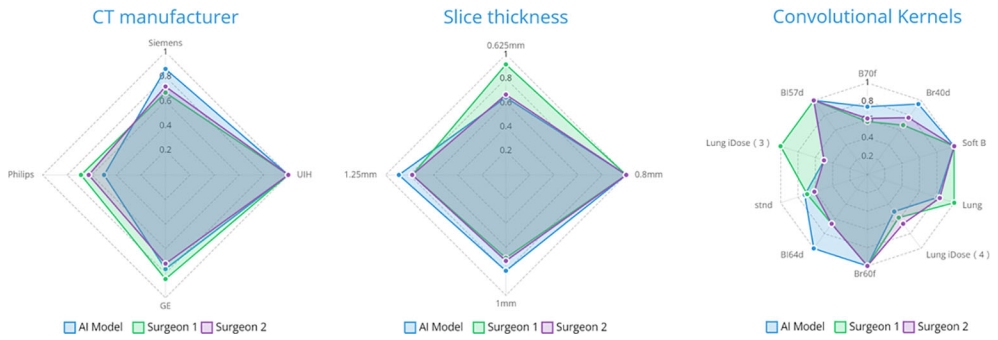


FIGURE 5
The impact of radiological parameters on accuracy on three parties.

and require delicate recognition for precise resection, in addition to the enormous number of vein variants on a segmental scale that cannot be fully represented due to the limited amount of training data, we thus split the surgical planning into two separate recognition tasks: segmental artery recognition and lobular vein recognition by developing a model that realizes the fine-grained recognition on arteries and relatively coarse-grained recognition on veins. Model performance was also evaluated by segmental artery recognition accuracy and lobular vein recognition accuracy on both case-wise and structure-wise.

In artery recognition, the algorithm showed similar accuracy compared to surgeons. It is worth notice that two parties exhibited complementary superiorities across lobes. The algorithm excels in the RUL, while surgeons excel in bilateral basal segments. Our result indicates a valid clinical application of this version of algorithm in assisting surgical planning in certain lobes, especially for junior thoracic surgeons. Vein recognition showed high accuracy in both parties; the algorithm may help in verifying the surgeon's recognition. Corroborating our result, our previous study[9] showed that the accuracy of the manual identification of the anatomical variant by thoracic surgeons using automated 3-D reconstruction is 85%, which is similar to our semiautomated approach. As for the time efficiency, compared to our previous study[9] that showed a median recognition time of 120 s by surgeons using 3-D reconstruction images, the human recognition of anatomical structure in this research is no longer needed, and the variant recognition time can be ignored, which is accomplished in less than 5 s.

When analyzing the possible influential factors for this semantic segmentation algorithm, we found the similar trend of lower accuracy in the contrast CT subgroup for both the AI algorithm and human performance. Contrast CTs are usually used in cases in which blood vessels were overlayed by the lesion. Such situation is more challenging for the algorithm than for surgeons. Since the model was also trained based on mostly non-contrast CT, the extra information provided by contrast CT may have not been fully utilized.

In terms of applicable scenarios, higher recognition accuracy was observed in segmentectomy compared with lobectomy for both

TABLE 2 Accuracy of the segmental artery in structure-wise error analysis.

Tumor location*	Segmental artery	AI	Surgeon A	Surgeon B
RUL	A1	1.00	0.86	1.00
	A2	0.80	0.30	0.50
	A3	0.75	0.50	0.50
RML	A4, A5	0.86	1.00	0.71
RLL	A6	0.75	0.33	1.00
	A7–10	0.20	0.60	0.60
LUL	A1+2	0.00	0.25	0.25
	A3	0.40	0.80	0.80
	A4, A5	0.60	0.40	0.20
LLL	A6	1.00	0.00	0.00
	A8–10	1.00	1.00	1.00
In group p-value		0.002	0.033	0.083

* RUL, right-upper lobe; RML, right-middle lobe; RLL, right-lower lobe; LUL, left-upper lobe; LLL, left-lower lobe.

the AI algorithm and the thoracic surgeon. Meanwhile, the error analysis indicated that the algorithm struggled in certain segmental arteries (LA^{1+2}, right basal segmental arteries, etc.) and surgeons struggled more in bilateral upper lobe arteries (RA^{2}, RA^{3}, and LA^{1+2}). Our preliminary results indicated a better performance of the AI algorithm on segmentectomy. Considering the possible bias introduced by a limited number of cases and unevenly distributed surgery types, a larger dataset needs to be employed to further confirm the observation. As for the inferior performance on LA^{1+2} and right basal segmental arteries, more similar training data are needed to enhance the model performance.

A different performance between the AI algorithm and surgeons was observed across lobes that may reflect their difference in mechanics. The algorithm performed well in RUL, while surgeons did not. On one hand, the RUL has relatively well-defined anatomical patterns that are easier for the algorithm to learn, while the 3-D structure is difficult for humans to imagine. On the other hand, the RLL, especially basal segmental arteries, have more anatomical variants that cannot be completely defined and represented during model training, while the imagination of a 3-D structure is relatively easy for humans; thus, surgeons perform better than the algorithm.

To our knowledge, this is the first study that validated the performance of a semiautomated surgical planning algorithm based on the semantic segmentation of CT imaging. Recently, a number of attempts of using semantic segmentation in medical imaging have been made. We have seen its application from the microscopic scale of segmenting red blood cells for sickle cell disease (16) to segmenting the COVID-19 infection area from chest CT (19) or lesions from endoscope images (22). However, the above-mentioned applications are mostly single-labeled tasks, using a large training set to achieve an organ or lesion recognition task. Our work is the first in the class semantic segmentation application that systemically presented the blood vessel anatomy of an organ. The complexity of this task is yet unparalleled.

In conclusion, the semiautomated surgical planning algorithm achieves similar accuracy in both segmental artery and lobular vein recognition compared to the junior attendings of thoracic surgery. As a continuation of our previous fully automated 3-D reconstruction study, the addition of a fine-grained semantic segmentation algorithm greatly enhanced their competency and practicality in aiding accurate preoperative planning and even made intraoperative intelligent interaction possible. With increased training data and refined labeling in the near future, the model will achieve higher accuracy and benefit both surgeons and patients in lobectomy and segmentectomy. Since the study is based on a small sample size, the promising results are to be confirmed in a large-scale validation study in the future.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by Peking University People's Hospital ethics committee (2022PHB011-002). Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

XC, HX, and DW: drafting manuscript, data labeling, result analyzing, model training; QQ, CS, CD, and MP: dicom data collection, data labeling; JJ, HZ, and ZW: data labeling; SW: label consultation; XuW and WW: reader study; XS, XiW, and LD: golden standard; ZH and YH: model training, label validation; YZ, NH, and FY: draft revision. All authors contributed to the article and approved the submitted version.

Funding

The work is funded by the Ministry of Science and Technology of the People's Republic of China, Grant/Award Number: 2020AAA0109600. This research is also supported by grants from Shanghai Hospital Development Center (SHDC12021111).

Acknowledgments

We would like to thank Tong Zou for his contribution to this work.

Conflict of interest

Authors DW, ZH, and YH were employed by Infervision Medical Technology Co., Ltd.

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

The reviewer YH declared a shared parent affiliation with the author(s) XC, HX, QQ, CS, JJ, HZ, XuW, WW, SW, XS, ZW, NH and FY to the handling editor at the time of review.

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.2022.1021084/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. Ganti AK, Klein AB, Cotarla I, Seal B, Chou E. Update of incidence, prevalence, survival, and initial treatment in patients with non-small cell lung cancer in the US. *JAMA Oncol* (2021) 7(12):1824–32. doi: 10.1001/jamaoncol.2021.4932

3. Saji H, Okada M, Tsuboi M, Nakajima R, Suzuki K, Aokage K, et al. Segmentectomy versus lobectomy in small-sized peripheral non-small-cell lung cancer (JCOG0802/WJOG4607L): a multicentre, open-label, phase 3, randomised, controlled, non-inferiority trial. *Lancet* (2022) 399(10335):1607–17. doi: 10.1016/S0140-6736(21)02333-3

4. Okada M, Koike T, Higashiyama M, Yamato Y, Kodama K, Tsubota N. Radical sublobar resection for small-sized non-small cell lung cancer: a multicenter study. *J Thorac Cardiovasc Surg* (2006) 132(4):769–75. doi: 10.1016/j.jtcvs.2006.02.063

5. Wang L, Song H, Wang M, Wang H, Ge R, Shen Y, et al. Utilization of ultrasonic image characteristics combined with endoscopic detection on the basis of artificial intelligence algorithm in diagnosis of early upper gastrointestinal cancer. *J Healthc Eng.* (2021) 2021:2773022. doi: 10.1155/2021/2773022

6. Prabhakar B, Singh RK, Yadav KS. Artificial intelligence (AI) impacting diagnosis of glaucoma and understanding the regulatory aspects of AI-based software as medical device. *Comput Med Imaging Graph.* (2021) 87:101818. doi: 10.1016/j.compmedimag.2020.101818

7. Piantadosi G, Sansone M, Fusco R, Sansone C. Multi-planar 3D breast segmentation in MRI *via* deep convolutional neural networks. *Artif Intell Med* (2020) 103:101781. doi: 10.1016/j.artmed.2019.101781

8. Mhiri I, Khalifa AB, Mahjoub MA, Rekik I. Brain graph super-resolution for boosting neurological disorder diagnosis using unsupervised multi-topology connectional brain template learning. *Med Imag. Anal* (2020) 65:101768. doi: 10.1016/j.media.2020.101768

9. Onishi Y, Teramoto A, Tsujimoto M, Tsukamoto T, Saito K, Toyama H, et al. Automated pulmonary nodule classification in computed tomography images using a deep convolutional neural network trained by generative adversarial networks. *BioMed Res Int* (2019) 2019:6051939. doi: 10.1155/2019/6051939

10. Chen X, Qi Q, Sun Z, Wang D, Sun J, Tan W, et al. Total nodule number as an independent prognostic factor in resected stage III non-small cell lung cancer: a deep learning-powered study. *Ann Transl Med* (2022) 10(2):33. doi: 10.21037/atm-21-3231

11. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* (2017) 542(7639):115–8. doi: 10.1038/nature21056

12. Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA* (2017) 318(22):2211–23. doi: 10.1001/jama.2017.18152

13. Zhu J, Wang J, Wang X, Gao M, Guo B, Gao M, et al. Prediction of drug efficacy from transcriptional profiles with deep learning. *Nat Biotechnol* (2021) 39(11):1444–52. doi: 10.1038/s41587-021-00946-z

14. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discovery* (2019) 18(6):463–77. doi: 10.1038/s41573-019-0024-5

15. Ekins S, Puhl AC, Zorn KM, Lane TR, Russo DP, Klein JJ, et al. Exploiting machine learning for end-to-end drug discovery and development. *Nat Mater* (2019) 18(5):435–41. doi: 10.1038/s41563-019-0338-z

16. Zhang M, Li X, Xu M, Li Q. Automated semantic segmentation of red blood cells for sickle cell disease. *IEEE J BioMed Health Inform.* (2020) 24(11):3095–102. doi: 10.1109/JBHI.2020.3000484

17. Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology. *Nat Rev Clin Oncol* (2019) 16(11):703–15. doi: 10.1038/s41571-019-0252-y

18. Niazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. *Lancet Oncol* (2019) 20(5):e253–e61. doi: 10.1016/S1470-2045(19)30154-8

19. Rajamani KT, Siebert H, Heinrich MP. Dynamic deformable attention network (DDANet) for COVID-19 lesions semantic segmentation. *J BioMed Inform.* (2021) 119:103816. doi: 10.1016/j.jbi.2021.103816

20. Zhang L, Zhang J, Li Z, Song Y. A multiple-channel and atrous convolution network for ultrasound image segmentation. *Med Phys* (2020) 47(12):6270–85. doi: 10.1002/mp.14512

21. Hu Y, Zhao H, Li W, Li J. Semantic image segmentation of brain MRI with deep learning. *Zhong Nan Da Xue Xue Bao Yi Xue Ban.* (2021) 46(8):858–64. doi: 10.11817/j.issn.1672-7347.2021.200744

22. Luo H, Xu G, Li C, He L, Luo L, Wang Z, et al. Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: a multicentre, case-control, diagnostic study. *Lancet Oncol* (2019) 20(12):1645–54. doi: 10.1016/S1470-2045(19)30637-0

23. Chen X, Wang Z, Qi Q, Zhang K, Sui X, Wang X, et al. A fully automated noncontrast CT 3-d reconstruction algorithm enabled accurate anatomical demonstration for lung segmentectomy. *Thorac Cancer.* (2022) 13(6):795–803. doi: 10.1111/1759-7714.14322

24. Gong C, Tao D, Liu W, Liu L, Yang J. Label propagation *via* teaching-to-Learn and learning-to-Teach. *IEEE Trans Neural Netw Learn. Syst* (2017) 28(6):1452–65. doi: 10.1109/TNNLS.2016.2514360

25. Qing Y, Zeng Y, Huang GB. Label propagation *via* local geometry preserving for deep semi-supervised image recognition. *Neural Netw* (2021) 143:303–13. doi: 10.1016/j.neunet.2021.06.007

26. Billmann F, Saracevic M, Schmidt C, Langan EA. Anatomical framework for pre-operative planning of laparoscopic left-sided colorectal surgery: Potential relevance of the distance between the inferior mesenteric artery and inferior mesenteric vein. *Ann Anat.* (2021) 237:151743. doi: 10.1016/j.aanat.2021.151743

27. Cui Z, Ding C, Li C, Song X, Chen J, Chen T, et al. Preoperative evaluation of the segmental artery by three-dimensional image reconstruction vs. thin-section multi-detector computed tomography. *J Thorac Dis* (2020) 12(8):4196–204. doi: 10.21037/jtd-20-1014

28. Yan J, Feng H, Wang H, Yuan F, Yang C, Liang X, et al. Hepatic artery classification based on three-dimensional CT. *Br J Surg* (2020) 107(7):906–16. doi: 10.1002/bjs.11458

frontiers | Frontiers in Oncology

TYPE Systematic Review
PUBLISHED 24 October 2022
DOI 10.3389/fonc.2022.987704

Check for updates

OPEN ACCESS

EDITED BY
Hongda Liu,
Nanjing Medical University, China

REVIEWED BY
Zehuan Liao,
Nanyang Technological University, Singapore
Faezeh Vakhshiteh,
Iran University of Medical Sciences, Iran

*CORRESPONDENCE
Julaiti Ainiwaer
julaiti0501@sina.com

†These authors share first authorship

SPECIALTY SECTION
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

RECEIVED 06 July 2022
ACCEPTED 03 October 2022
PUBLISHED 24 October 2022

CITATION
Yu W, Liu R, Miao Z, Zhang L, Sheyhidin I and Ainiwaer J (2022) The diagnostic significance of blood-derived circRNAs in NSCLC: Systematic review and meta-analysis. *Front. Oncol.* 12:987704. doi: 10.3389/fonc.2022.987704

COPYRIGHT
© 2022 Yu, Liu, Miao, Zhang, Sheyhidin and Ainiwaer. 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 diagnostic significance of blood-derived circRNAs in NSCLC: Systematic review and meta-analysis

Weijie Yu[1†], Ruixue Liu[2†], Zhoulin Miao[3], Liwei Zhang[2], Ilyar Sheyhidin[2] and Julaiti Ainiwaer[2*]

[1]Department of Cardio-Thoracic Surgery, The Fourth Affiliated Hospital of Xinjiang Medical University, Xinjiang, China, [2]Department of Thoracic Surgery, The First Affiliated Hospital of Xinjiang Medical University, Xinjiang, China, [3]Radiology Department, Qingdao Municipal Hospital, Qingdao, Shandong, China

Background: Using blood-derived circular RNAs (circRNAs) may be an efficient tool for noninvasive fluid biopsy in diagnosing non-small cell lung cancer (NSCLC). However, no relevant systemic meta-analysis has been conducted so far to support the diagnostic value of using blood-derived circRNAs in NSCLC clinically. The aim of this study is to clarify the issue through a meta-analysis.

Methods: A systematic search strategy was used to search relevant literature in the databases of PubMed, Web of Science, and Cochrane Library from 2017 to 2022. The relationship between the diagnostic accuracy of circRNAs and NSCLC was analyzed. For the purpose of evaluating the quality of the literature, Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) was used. Statistical analyses were assessed using Stata software (version 17.0) and META-DISC (version 1.4).

Results: The meta-analysis included 1,093 patients with NSCLC and 959 controls. Results are as follows: pooled sensitivity, 0.78 (95% CI = 0.71–0.83, I^2 = 71.86); pooled specificity, 0.76 (95% CI = 0.70–0.82, I^2 = 70.12); pooled positive likelihood ratio (PLR), 3.3 (95% CI = 2.6–4.2, I^2 = 37.56); pooled negative likelihood ratio (NLR), 0.29 (95% CI = 0.23–0.37, I^2 = 64.67); diagnostic odds ratio (DOR), 11.42 (95% CI = 7.88–16.56, I^2 = 99.05); area under the receiver operating characteristic curve (AUC), 0.84 (95% CI = 0.80–0.87). Based on the subgroup analysis, it appears that the heterogeneity is primarily caused by the NSCLC subgroup.

Conclusion: circRNAs are highly useful diagnostic biomarkers for NSCLC in China. Further prospective studies on the diagnostic value of circRNAs should be conducted in multiple countries.

Systematic review registration: https://www.crd.york.ac.uk/prospero/, identifier CRD42022323804.

KEYWORDS

circRNAs, NSCLC, diagnosis, blood-derived, oncology

Background

Circular RNAs (circRNAs) are a class of circular non-coding RNAs that are covalently formed by reverse 3′ and 5′ clipping (1). The original circRNAs were first located in the cytoplasm of HeLa cells in eukaryotic cells and then gradually found in humans and animals (2). This discovery has evolved into a current research hot spot. However, the initial understanding was not sufficient, and researchers once considered it as an error product in the transcription process (3). Until the emergence and development of high-throughput sequencing technology, researchers shifted their insights into circRNAs and gradually deepened their functional exploration. Now, circRNAs have been confirmed to play an essential role in the sponge adsorption of microRNAs (miRNAs), gene regulation of transcription, and protein translation (4–8). In addition, circRNAs are closely related to many malignant tumors, such as lung cancer and gastric cancer (9, 10). The abnormal expression of circRNAs also plays a vital role in the regulation of tumor cell invasion, metastasis, apoptosis, immune escape, etc. and is a promising biomarker and therapeutic target today (11–13). Moreover, circRNAs also have many characteristics. Due to their unique structure, circRNAs are not susceptible to ribonucleases, which provides them with strong structural stability. circRNAs have been demonstrated to be obtained from various sources, including tissues, and enter into the bloodstream by the exosomes. Aside from blood, urine and saliva can also be obtained. However, their expression varies by disease and tissue, indicating that they are tissue-specific (14, 15). It is also due to the characteristics of circRNAs that ensure circRNAs can be detected in blood as tumor markers, which provides a new direction for liquid biopsy in disease diagnosis (16, 17).

Lung malignancies represent a considerable proportion of today's tumor-related disease spectrum. In 2018, there were 2.1 million new cases of lung cancer worldwide, accounting for 11.6% of new tumor cases, and 1.8 million people died of lung cancer, accounting for 18.4% of cancer-related deaths (18). Moreover, non-small cell lung cancer (NSCLC) accounts for approximately 80%. A vast number of such cases are not only due to environmental factors but also due to the lack of obvious and specific symptoms of the disease, which make it difficult to attract the attention of patients, so they miss the best opportunity for treatment. Achieving early detection and early diagnosis to get medical treatment is significant for patients with NSCLC. It is widely regarded that the efficacy of tumor markers in NSCLC remains controversial, although blood-derived tumor markers such as carcinoembryonic antigen (CEA), neuron-specific enolase (NSE), cytokeratin 19 fragment antigen21 -1 (CYFRA21-1) have been listed as recommended tumor markers by many countries.

Huang et al. (19) explored the diagnostic efficacy of blood-derived hsa_circ_0070354 for NSCLC and found that it was superior to traditional tumor markers in the diagnosis. Wang et al. (20) evaluated the diagnostic performance of circRNAs for lung cancer by summarizing tissue-derived circRNAs, but in clinical practice, tissue must be obtained by puncture or surgical resection, which is meaningless for body fluid diagnosis. Therefore, high-level evidence of blood-derived circRNAs in the diagnosis of lung cancer is currently required.

In summary, this paper will explore the diagnostic value of blood-derived circRNAs in NSCLC and summarize the mechanism of the literature to provide new directions for clinical work and future scientific research.

Materials and methods

This meta-analysis was performed following the Preferred Reporting Items for Systematic reviews and Meta Analyses (PRISMA) statement.

Registration

The study has been registered on the PROSPERO website (ID: CRD42022323804).

Data source and search strategy

The electronic databases, including PubMed, Web of Science, Cochrane Library, are used to systematically search articles from January 2017 to January 2022. The following search terms or keywords are used: ("circRNAs or circular RNAs") and ("NSCLC or non-small cell lung cancer or Carcinoma, Non-Small Cell Lung or Lung squamous cell carcinoma, Adenocarcinoma, Lung or lung adenocarcinoma or Carcinoma, Non Small Cell Lung or Carcinomas, Non-Small-Cell Lung or Lung Carcinoma or Non-Small-Cell, Lung Carcinomas or Non-Small-Cell Lung Carcinomas or LCLC or large cell lung cancer") and ("biomarkers") and ("diagnosis") and ("serum or blood or plasma"). Some articles were manually searched for completeness and comprehensiveness. Two researchers (WY and RL) read titles, abstracts, and full-text articles to choose the right article. Reference lists of articles were reviewed and retrieved to further identify potentially relevant studies.

Inclusion and exclusion criteria

Inclusion criteria include the following: 1) studies analyzing the relationship between circRNAs and diagnosis of NSCLC; 2) studies providing data on the sensitivity (SEN), specificity (SPE), true positive (TP), false positive (FP), true negative (TN), or false

negative (FN); 3) All standard evidence must be the pathology; 4) Source of the circRNAs is blood-based including serum and plasma; 5) None of the included patients had a prior history of other cancers or metastatic cancer from other sites or had received chemotherapy or radiotherapy prior to plasma collection.

The exclusion criteria include the following: 1) repetitive research; 2) letters, editorials, commentaries, or abstracts; 3) studies involving ineligible patients or controls; 4) studies lacking data; 5) studies in a non-English language; 6) Source of the circRNAs is from tissue or other types; 7) In the absence of a control group research; 8) There was no comparative study. If the results or research cases are overlapped, only the first study or the complete study was included; 9) Patients had a prior history of other cancers or metastatic cancer from other sites or had received chemotherapy or radiotherapy prior to plasma collection.

Data extraction and quality assessment

Two reviewers (WY and RL) extracted the data. If necessary, the discrepancies were solved by the third reviewer (ZM). The following information is extracted from each article: first author name, publication time, area, impact factor, number of case, circRNA type, sample type or origination, cancer type, sensitivity, specificity, area under the receiver operating characteristic curve (AUC), testing method and differential expression of circRNAs. If the value of TP, FP, TN, or FN is omitted in the article, four indexes will be calculated from the sensitivity, specificity, and number of cases.

Study quality assessment

The Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) Evaluation Scale was used by two reviewers independently to assess the quality of every included article. A score beyond 9 is deemed as good quality.

Statistical analysis

The STATA 17.0 is used to analyze the diagnostic value of circRNAs. The sensitivity [TP/(FN+TP)], specificity [TN/(FP+TN)], positive likelihood ratio (PLR) [sen/(1-spe)], negative likelihood ratio (NLR) [(1-sen)/spe], diagnostic odds ratio (DOR) [PLR/NLR], 95% confidential intervals (95% CIs), summary receiver operating characteristic (SROC)curve, and AUC were plotted to estimate the diagnostic value of circRNAs. A two-sided $p < 0.05$ was considered statistically significant. A fixed-effects model is used if the heterogeneity is minimal ($I^2 < 50\%$); otherwise, a random-effects model was used in the significant heterogeneity ($I^2 > 50\%$). The possible source of heterogeneity is performed by threshold effect, regression analysis, subgroup analysis, and sensitivity analysis. The value of RDOR is calculated by the META-DISC 1.40.

Results

Search selection and characteristics

The flow diagram for literature research processes is shown in Figure 1. circRNAs derived from the blood are analyzed to evaluate the diagnostic value of NSCLC. The characteristics of the included studies are summarized (Table 1). A total of 1,093 patients with NSCLC and 959 controls, who are all from Asia, were analyzed in this article with 12 articles and 12 kinds of circRNAs. All circRNAs are detected in qRT-PCR.

All of the studies are independently scored based on QUADAS-2 score system.

Results of the meta-analysis

Obvious heterogeneity was assessed using the random-effects model ($I^2 > 50\%$). For the value of blood-derived circRNAs in diagnosing NSCLC, the result are as follows: pooled sensitivity, 0.78 (95% CI = 0.71–0.83, $I^2 = 71.86$); pooled specificity, 0.76 (95% CI = 0.70–0.82, $I^2 = 70.12$); pooled PLR, 3.3 (95% CI = 2.6–4.2, $I^2 = 37.56$); pooled NLR, 0.29 (95% CI = 0.23–0.37, $I^2 = 64.67$); DOR, 11.42 (95% CI = 7.88–16.56, $I^2 = 99.05$); AUC, 0.84 (95% CI = 0.80–0.87). Forest plots and SROC curves are shown in Figures 2A–D. As shown in Figure 2E, the result shows a pretest probability of 20%, and the posttest probability rates, given the positive and negative results, were respectively 45% and 7%. Among the 12 included studies (one article contains two types of circRNAs), 11 studies are located in the right lower quadrants, implying that the circRNAs are valuable in the diagnosis (Figure 2F).

Threshold effect

The threshold effect was calculated using the Spearman rank correlation. The Spearman correlation coefficient was 0.354 ($p = 0.259$), which suggests that the heterogeneity is not caused by the threshold effect.

Meta regression analysis

Because of the value of I^2 >50%, it suggests evident heterogeneity. From the baseline of study characteristics, cancer type, sample type, and expression situation may be the potential reasons for the heterogeneity. So, we further completed

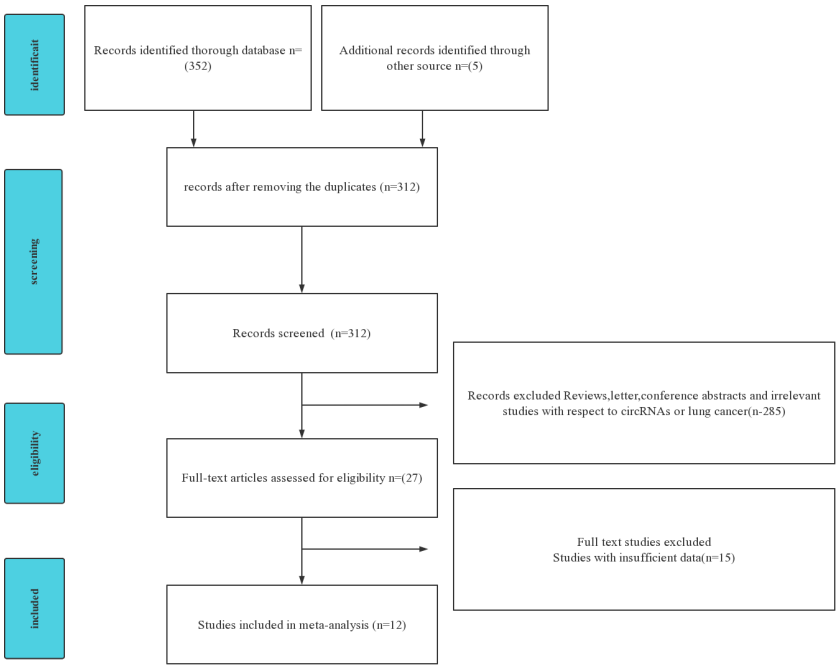


FIGURE 1
Preferred Reporting Items for Systematic reviews and Meta Analyses (PRISMA) flowchart of the included and excluded study selection process.

the meta-regression analysis. No significant causes of heterogeneity were found ($p > 0.05$). The results are shown in Table 2.

Sensitivity analysis

The value of I^2 >50% suggests apparent heterogeneity. Sensitivity analysis is further completed to discover the heterogeneity. The following chart conveys a conclusion that although there are some inconsistencies when successfully omitting one in the literature, the result is robust (Figure 3).

Subgroup analysis

In order to find out the reason for heterogeneity, a subgroup analysis was performed based on the cancer type, sample type, and expression condition. The AUC values for all three subgroups are >0.8, suggesting an excellent diagnosis value. Aside from the NSCLC subgroup, almost all of the subgroup analyses suggest that the I^2 of the subgroup is low (I^2 <50%). So, these subgroups may cause the heterogeneity. Moreover, results show that the diagnosis effectiveness of serum-based circRNAs is greater than that of the plasma-based circRNAs. Because of the paucity of literature on whole blood-based circRNAs, it was impossible to calculate I^2 and p values. So do downregulated circRNAs (Table 3).

Publication bias

The publication bias of the included studies was certified by STATA 17.0 software. The Egger test and funnel plot are used to assess the publication bias. The result of the Egger test shows 0.027 ($p > 0.05$) (Figure 4A). The funnel plot was constructed with the STATA 17.0 software, and the funnel plot is symmetrical and shows no publication bias (Figure 4B). Deeks' funnel plot is also used to evaluate the publication bias, and the result shows 0.88 ($p > 0.06$). So, Deeks' funnel plot shows no publication bias (Figure 4C).

Discussion

Nowadays, the discovery of various biomarkers and the exploration of body fluid diagnosis have gradually deepened. With the recognition of circRNAs, many kinds of circRNAs are explored, and mechanisms of diagnosis are identified (32, 33). Researchers know the importance of circRNAs. According to existing research, tissue-derived RNA has better diagnostic performance, but obtaining tissue means trauma to the

TABLE 1 Characteristics of circRNAs and quality of included studies in diagnosing.

First author	Country	Year	IF	circRNAs	Cancer type	circRNA regulation	Sample type	Sample size		Testing method	I–II/III–IV	TP	FP	FN	TN	Quality score
								No. of Cases	No. of control							
Guo-jun Lu (21)	Asia (China)	2020	4.147	hsa_circ_0001715	LA	Up	Plasma	60	57	qRT-PCR	31/29	53	16	7	41	9
Xiaoli Zhu (22)	Asia (China)	2017	5.542	hsa_circ_0013958	LA	Up	Plasma	30	30	qRT-PCR	-/-	20	2	10	28	9
Fan Yang (23)	Asia (China)	2021	3.269	circRNA_001846	NSCLC	Up	Serum	206	206	qRT-PCR	109/97	161	39	45	167	10
Chengzhi Ding (24)	Asia (China)	2020	4.599	circ-MEMO1	NSCLC	Up	Serum	30	25	qRT-PCR	-/-	17	1	13	24	9
Jie Shi (25)	Asia (China)	2021	11.61	circ-pvt1	LA	Up	Serum	104	110	qRT-PCR	50/54	95	43	9	67	9
Xiuyuan Li (26)	Asia (China)	2018	5.5	CircPVT1	NSCLC	Up	Serum	45	45	qRT-PCR	-/-	32	9	13	36	10
Xiaoxia Liu (27)	Asia (China)	2019	5.531	hsa_circ_0086414	LA	Down	Plasma	54	54	qRT-PCR	-/-	42	18	12	36	9
				hsa_circ_0005962	LA	Up	Plasma	54	54	qRT-PCR	-/-	39	15	16	39	9
Yong Zhou (28)	Asia (China)	2022	2.718	hsa_circ_0072309	NSCLC	Up	Serum	46	40	qRT-PCR	26/20	40	7	6	33	10
Dong Hang (29)	Asia (China)	2018	4.452	circFARSA	NSCLC	Up	Plasma	50	50	qRT-PCR	–	32	17	18	33	9
Ziyi Peng (30)	Asia (China)	2021	3.786	CircTOLLIP	NSCLC	Down	Blood	88	76	qRT-PCR	–	66	26	22	50	10
Yinzai He (31)	Asia (China)	2021	3.25	circ_0048856	NSCLC	Up	Serum	50	50	qRT-PCR	–	44	10	6	40	10

circRNA, circular RNA; NSCLC, non-small cell lung cancer; AUC, area under the receiver operating characteristic curve; LA, lung adenocarcinoma; FN, false negative; FP, false positive; TN, true negative; TP, true positive; qRT-PCR, quantitative real-time polymerase chain reaction.

10.3389/fonc.2022.987704

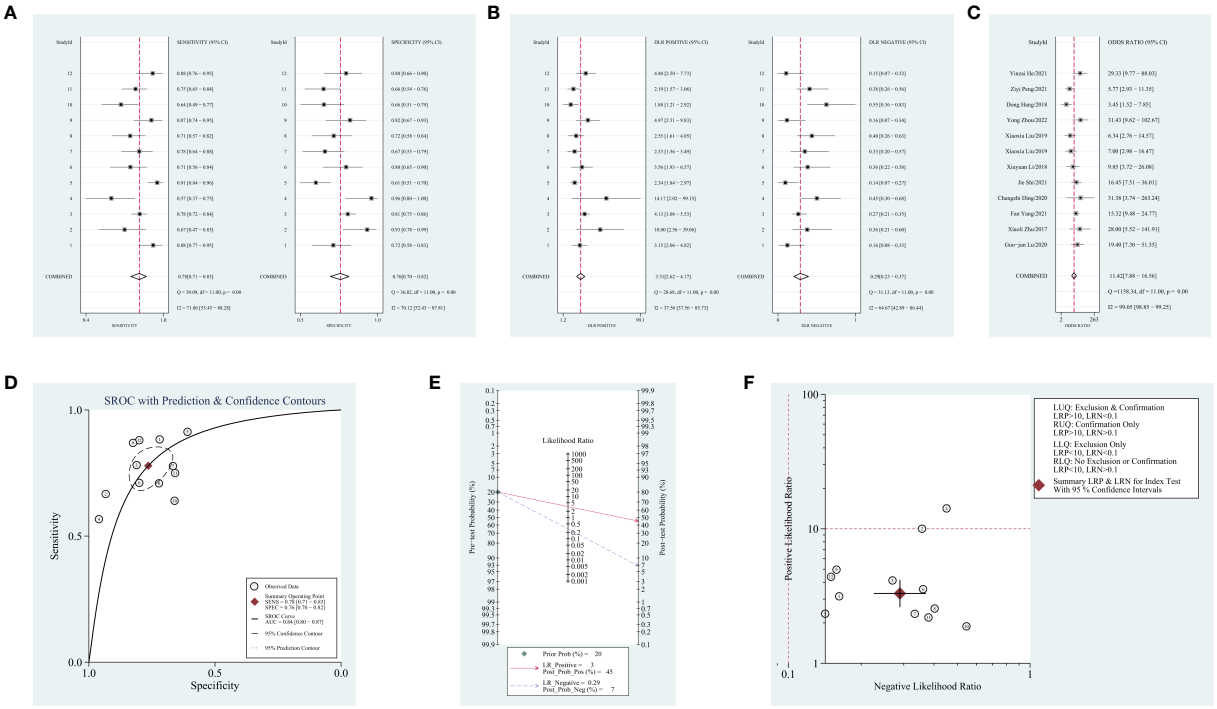


FIGURE 2
Forest plots of diagnostic accuracy index and summary receiver operating characteristic (SROC) curve and Fagan's nomogram for likelihood ratios, a likelihood ratio scattergram, publication bias. **(A)** Forest plots of sensitivity and specificity for circular RNAs (circRNAs) in diagnosis. **(B)** Forest plots of the positive likelihood ratio and negative likelihood ratio in diagnosis. **(C)** Forest plots of the diagnostic odds ratio in diagnosis. **(D)** SROC curve for circRNAs in diagnosing non-small cell lung cancer (NSCLC). **(E)** Fagan's nomogram for likelihood ratios. **(F)** A likelihood ratio scattergram.

patient. Nowadays, liquid biopsy in a non-invasion form is deemed a revolutionary tool in the diagnosis and has attracted much attention worldwide. As an excellent biomarker for diagnosis, it must hold several characteristics such as analytical validity (34), clinical validity, and utility. Combining the advantages of circRNAs, it may be a promising biomarker to distinguish normal, benign disease from malignant tumor. However, a blood-based meta-analysis of circRNAs with respect to diagnosis is absent. This article makes a supplement to this field and expresses the significance of circRNAs as biomarkers in liquid biopsy. A total of 12 articles are enrolled from 2017 to 2022, containing 12 types of circRNAs and more than 2,000 cases, in which 10 types are upregulated, and two types are downregulated. The results showed that the AUC was 0.84 (95% CI = 0.80–0.87), and the pooled sensitivity, specificity, DOR, pooled PLR, and pooled NLR were, respectively, 0.78 (95% CI = 0.71–0.83, I^2 = 71.86), 0.76 (95% CI = 0.70–0.82, I^2 = 70.12), 11.42 (95% CI = 7.88–16.56, I^2 = 99.05), 3.3 (95% CI = 2.6–4.2, I^2 = 37.56), and 0.29 (95% CI = 0.23–0.37, I^2 = 64.67). It is generally believed that the value of AUC between 0.7 and 0.9 is considered a good diagnostic value. So, the literature indicated the significant body fluid diagnostic value of circRNAs for NSCLC. Apart from this, stability and expression of different diseases will contribute to circRNAs being a more suitable biomarker in diagnosing and solving the problem of low organ specificity of existing markers. According to existing research, the AUC of tumor tissue-based circRNAs is between 0.75 and 0.85 (34–36), which is very close to the AUC of blood-based circRNAs and means that the non-invasion blood acquisition mode in diagnosing will be adapted to apply clinically.

TABLE 2 Meta regression analysis.

	Coeff	Std.Err	RDOR	P value	95% CI
Cancer type	-0.2529914	0.1375258	0.71	0.103	(-0.570126 to 0.0641436)
Sample type	0.3997855	0.2138254	1.95	0.098	(-0.093296 to 0.8928676)
Expression situation	-0.5589196	0.2791767	0.34	0.080	(0.820598–2.624503)

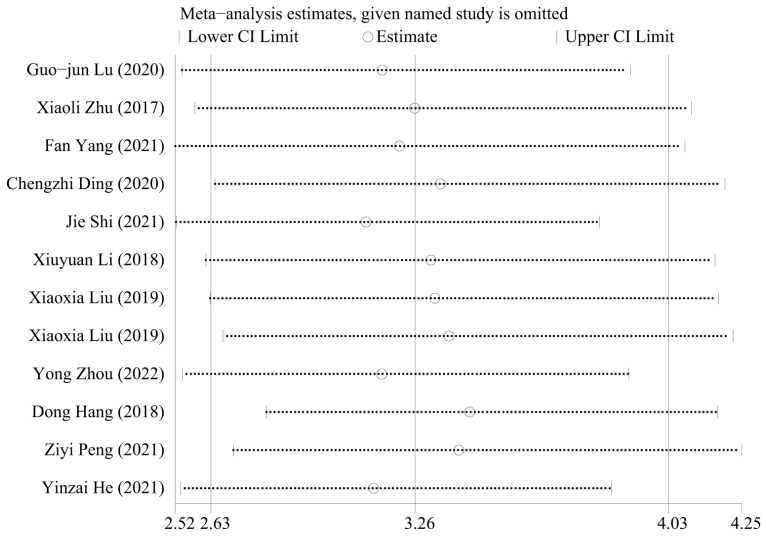


FIGURE 3
Sensitivity analyses for diagnosis analysis.

In addition, the result of the threshold effect calculated by the Spearman correlation coefficient shows that it is not responsible for the heterogeneity, and the outcome was 0.354 ($p > 0.05$). Next, the regression analysis, which involves sample type, lung cancer type, and differential expression of circRNAs, showed that these are not the reason for heterogeneity ($p > 0.05$). In order to find out the source of heterogeneity, results of the sensitivity analysis and subgroup analysis show that when the total cases are divided into three subgroups, the I^2 of the LA group is less than 50% ($p > 0.05$). However, the I^2 of NSCLC is greater than 50%. It means that the NSCLC subgroup might be the source of the heterogeneity. Further reading of the literature revealed that the types of disease and the case of patients in the literature might be responsible for the heterogeneity. The subgroup of serum-based circRNAs, which has an excellent AUC value, has a more extraordinary diagnosis performance than plasma-based circRNAs. Whether the difference is statistically significant is uncertain because the two subgroups contain different circRNAs and different diseases. However, Li et al. (37) believe that serum-derived RNA has higher diagnostic performance. The reason may be that serum exosomes contain a large number of stably expressed circRNAs. After dividing into three subgroups, the number of cases of some groups is fewer because some indexes cannot be calculated by the software. Using Egger test, funnel plot,

TABLE 3 Subgroup analysis of NSCLC diagnosis.

	Number of studies	I^2	p value	Sensitivity	Specificity	PLR	NLR	DOR	AUC
Cancer type									
LA	5	46.8%	0.111	0.81	0.72	2.9	0.27	11	0.82
NSCLC	7	64.3%	0.01	0.76	0.78	3.5	0.31	11.24	0.84
Sample type									
Plasma	5	48.9%	0.098	0.75	0.73	2.8	0.35	7.99	0.80
Serum	6	42.1%	0.124	0.8	0.8	4.1	2.4	17	0.87
Blood	1	–	–	0.9	0.902	9.183	0.111	82.837	0.95
Expression regulation									
Up	10	59.2%	0.009	0.77	0.80	3.9	0.28	14	0.86
Down	2	–	–	0.76	0.66	2.25	0.36	6.21	–

LA, lung adenocarcinoma; NSCLC, non-small cell lung cancer; PLR, positive likelihood ratio; NLR, negative likelihood ratio; DOR, diagnostic odds ratio; AUC, area under the receiver operating characteristic curve.
Publication bias was assessed in the end.

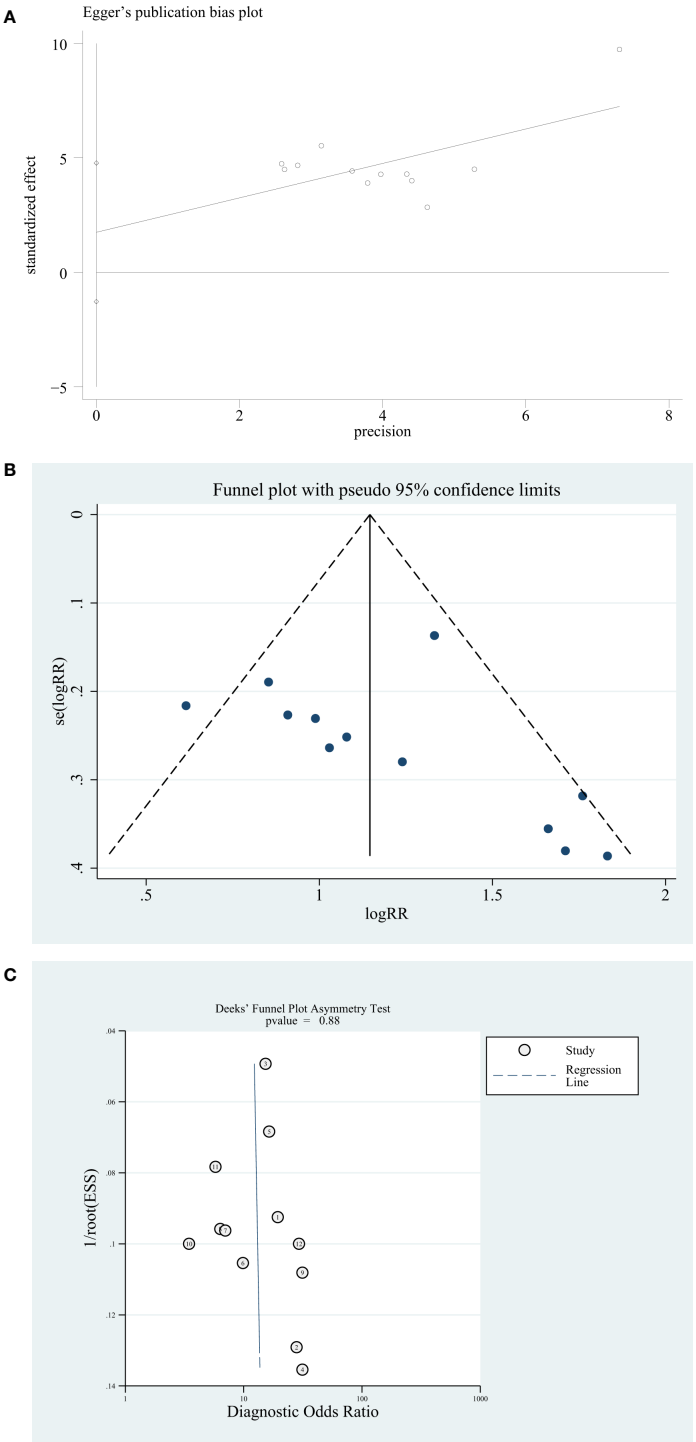


FIGURE 4
Publication bias analysis. **(A)** Egger's publication bias plot. **(B)** Funnel plot. **(C)** Deeks' funnel plot.

and Deeks' funnel plot to analyze publication bias, all three showed an insignificant publication bias ($p > 0.05$).

Similar findings have been found in a variety of malignancies, except for NSCLC. Hu et al. (38) validated that circGSK3 from plasma has a greater AUC value than 0.8 for the diagnosis of early esophageal squamous cell carcinoma (ESCC) in a cohort of 86 individuals and concluded that circGSK3β is a very excellent diagnostic marker for body fluids. Subsequent follow-up of patients in the experimental group and comparison of pre- and postoperative plasma-derived circGSK3β expression levels revealed that circGSK3β also played an excellent predictive function for postoperative recurrence and metastasis. Omid-Shafaat et al. (39) evaluated the diagnostic efficacy of circELP3 in combination with circFAF1 for breast cancer with AUC, sensitivity, and specificity of 0.891, 96%, and 62%, respectively, and concluded that it plays an excellent role as a non-invasive body fluid diagnostic marker for breast cancer. Yu et al. (40) compared plasma-derived circRNAs with α-fetoprotein (AFP) in the diagnosis of hepatitis B virus (HBV)-associated hepatocellular carcinoma and found that the combined AUC of hsa_circ_0000976, hsa_circ_0007750, and hsa_circ_0139897 was 0.843, which was much better than the AUC of 0.747 for conventional AFP. It also showed a significant effect in AFP-negative hepatocellular carcinoma. Shi et al. (41) similarly found that low expression of plasma-derived hsa_circ_001888 in gastric cancer may be a potential diagnostic marker.

To understand the role of circRNAs in the diagnosis field, the mechanism of circRNAs must be taken into consideration. Almost all abundance of circRNAs is related to disease stage and characteristics. The hsa_circ_0013958 is upregulated and promotes cell proliferation and invasion and inhibits cell apoptosis in lung adenocarcinoma. At the same time, it is also considered as a sponge of miR-134, and thus it upregulates oncogenic cyclin D1. The hsa_circ_0000190 is upregulated in the patients' plasma with lung cancer, and it has something to do with clinical characteristics (42). The circPVT1 is upregulated, and when it is absent, the invasion or proliferation is blocked. It controls the E2F transcription factor 2 (E2F2) signaling pathway to affect NSCLC (26).

It is important to take into account several limitations in this study. First, too few samples were included. Second, the type of circRNAs and sample size are limited. It is necessary to explore more types of circRNAs. Second, this type of research contains several different types and sources of circRNAs. The number of subtypes of NSCLC also varies significantly from article to article regarding the diagnosis of this disease. These are all potential causes. Third, the evidence level of enrolled literature is different. Fourth, the research objects are in China. It is worth noting that the expression of circRNAs may differ from other countries, and application of circRNAs as biomarkers will require further study in combination with data from those countries.

Conclusion

To sum up the above, this systematic review of data extracted from 12 theses based on patients' blood with NSCLC indicated the significant value of circRNAs in diagnosing NSCLC. Therefore, combining the features of circRNAs, the results mean that circRNAs, especially the serum-based circRNAs, play a role in non-invasion body fluid biopsy and may open up broad prospects for specific disease diagnostics in the future, which perhaps is not limited to NSCLC, depending on its expression and organ specificity. The noninvasive method of acquiring circRNAs facilitates their wide use clinically at a lower cost, and circRNAs will be a better choice than the tumor markers. Nonetheless, better designed and larger-scale research of multinational clinical trials are required to certify the results.

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/s.

Author contributions

All authors contributed to the article and approved the submitted version.

Funding

The project is supported by State Key Laboratory of Pathogenesis, Prevention and Treatment of High Incidence Diseases in Central Asia (No.SKL-HIDCA-2020-42).

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

1. Lei M, Zheng G, Ning Q, Zheng J, Dong D. Translation and functional roles of circular RNAs in human cancer. *Mol Cancer* (2020) 19(1):30. doi: 10.1186/s12943-020-1135-7

2. Sanger HL, Klotz G, Riesner D, Gross HJ, Kleinschmidt AK. Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures, proc. *Natl Acad Sci USA* (1976) 73(11):3852–6. doi: 10.1073/pnas.73.11.3852

3. Huang A, Zheng H, Wu Z, Chen M, Huang Y. Circular RNA-protein interactions: functions, mechanisms, and identification. *Theranostics* (2020) 10 (8):3503–17. doi: 10.7150/thno.42174

4. Liu T, Song Z, Gai Y. Circular RNA circ_0001649 acts as a prognostic biomarker and inhibits NSCLC progression *via* sponging miR-331-3p and miR-338-5p. *Biochem Biophys Res Commun* (2018) 503(3):1503–9. doi: 10.1016/j.bbrc.2018.07.070

5. Yang Y, Fan X, Mao M, Song X, Wu P, Zhang Y, et al. Extensive translation of circular RNAs driven by N(6)-methyladenosine. *Cell Res* (2017) 27(5):626–41. doi: 10.1038/cr.2017.31

6. Guo JU, Agarwal V, Guo H, Bartel DP. Expanded identification and characterization of mammalian circular RNAs. *Genome Biol* (2014) 15(7):409. doi: 10.1186/s13059-014-0409-z

7. Jeck WR, Sharpless NE. Detecting and characterizing circular RNAs. *Nat Biotechnol* (2014) 32(5):453–61. doi: 10.1038/nbt.2890

8. Zang J, Lu D, Xu A. The interaction of circRNAs and RNA binding proteins: An important part of circRNA maintenance and function. *J Neurosci Res* (2020) 98 (1):87–97. doi: 10.1002/jnr.24356

9. Wang C, Tan S, Li J, Liu WR, Peng Y, Li W. CircRNAs in lung cancer - biogenesis, function and clinical implication. *Cancer Lett* (2020) 492:106–15. doi: 10.1016/j.canlet.2020.08.013

10. Ghafouri-Fard S, Honarmand Tamizkar K, Jamali E, Taheri M, Ayatollahi SA. Contribution of circRNAs in gastric cancer. *Pathol Res Pract* (2021) 227:153640. doi: 10.1016/j.prp.2021.153640

11. Shi X, Wang B, Feng X, Xu Y, Lu K, Sun M. circRNAs and exosomes: A mysterious frontier for human cancer. *Mol Ther Nucleic Acids* (2020) 19:384–92. doi: 10.1016/j.omtn.2019.11.023

12. Zhang PF, Pei X, Li KS, Jin LN, Wang F, Wu J, et al. Circular RNA circFGFR1 promotes progression and anti-PD-1 resistance by sponging miR-381-3p in non-small cell lung cancer cells. *Mol Cancer* (2019) 18(1):179. doi: 10.1186/s12943-019-1111-2

13. Hong W, Xue M, Jiang J, Zhang Y, Gao X. Circular RNA circ-CPA4/ let-7 miRNA/PD-L1 axis regulates cell growth, stemness, drug resistance and immune evasion in non-small cell lung cancer (NSCLC). *J Exp Clin Cancer Res* (2020) 39 (1):149. doi: 10.1186/s13046-020-01648-1

14. Ma Z, Shuai Y, Gao X, Wen X, Ji J. Circular RNAs in the tumour microenvironment. *Mol Cancer* (2020) 19(1):8. doi: 10.1186/s12943-019-1113-0

15. Li Z, Cheng Y, Wu F, Wu L, Cao H, Wang Q, et al. The emerging landscape of circular RNAs in immunity: breakthroughs and challenges. *biomark Res* (2020) 8:25. doi: 10.1186/s40364-020-00204-5

16. Moldogazieva NT, Zavadskiy SP, Terentiev AA. Genomic landscape of liquid biopsy for hepatocellular carcinoma personalized medicine. *Cancer Genom Proteom.* (2021) 18(3 Suppl):369–83. doi: 10.21873/cgp.20266

17. Wang Y, Liu J, Ma J, Sun T, Zhou Q, Wang W, et al. Exosomal circRNAs: Biogenesis, effect and application in human diseases. *Mol Cancer* (2019) 18(1):116. doi: 10.1186/s12943-019-1041-z

18. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *C.A. Cancer J Clin* (2018) 68(6):394–424. doi: 10.3322/caac.21492

19. Huang Y, Qin S, Gu X, Zheng M, Zhang Q, Liu Y, et al. Comprehensive assessment of serum hsa_circ_0070354 as a novel diagnostic and predictive biomarker in non-small cell lung cancer, front. *Genet* (2021) 12:796776. doi: 10.3389/fgene.2021.796776

20. Wang C, Jiang Y, Lei Q, Wu Y, Shao J, Pu D, et al. Potential diagnostic and prognostic biomarkers of circular RNAs for lung cancer in China. *Biomed Res Int* (2019) 2019:8023541. doi: 10.1155/2019/8023541

21. Lu GJ, Cui J, Qian Q, Hou ZB, Xie HY, Hu W, et al. Overexpression of hsa_circ_0001715 is a potential diagnostic and prognostic biomarker in lung adenocarcinoma, onco. *Targets Ther* (2020) 13:10775–83. doi: 10.2147/ott.S274932

22. Zhu X, Wang X, Wei S, Chen Y, Chen Y, Fan X, et al. hsa_circ_0013958: a circular RNA and potential novel biomarker for lung adenocarcinoma. *Febs. J* (2017) 284(14):2170–82. doi: 10.1111/febs.14132

23. Yang F, Ma C, Qiu J, Feng X, Yang K. Identification of circRNA_001846 as putative non-small cell lung cancer biomarker. *Bioengineered* (2021) 12(1):8690–7. doi: 10.1080/21655979.2021.1991161

24. Ding C, Xi G, Wang G, Cui D, Zhang B, Wang H, et al. Exosomal circ-MEMO1 promotes the progression and aerobic glycolysis of non-small cell lung cancer through targeting MiR-101-3p/KRAS axis. *Front Genet* (2020) 11:962. doi: 10.3389/fgene.2020.00962

25. Shi J, Lv X, Zeng L, Li W, Zhong Y, Yuan J, et al. CircPVT1 promotes proliferation of lung squamous cell carcinoma by binding to miR-30d/e. *J Exp Clin Cancer Res* (2021) 40(1):193. doi: 10.1186/s13046-021-01976-w

26. Li X, Zhang Z, Jiang H, Li Q, Wang R, Pan H, et al. Circular RNA circPVT1 promotes proliferation and invasion through sponging miR-125b and activating E2F2 signaling in non-small cell lung cancer. *Cell Physiol Biochem* (2018) 51 (5):2324–40. doi: 10.1159/000495876

27. Liu XX, Yang YE, Liu X, Zhang MY, Li R, Yin YH, et al. A two-circular RNA signature as a noninvasive diagnostic biomarker for lung adenocarcinoma. *J Transl Med* (2019) 17(1):50. doi: 10.1186/s12967-019-1800-z

28. Zhou Y, Tong Z, Zhu X, Huang S, Dong Z, Ye Z. hsa_circ_0072309 expression profiling in non-small-Cell lung carcinoma and its implications for diagnosis and prognosis. *Front Surg* (2022) 9:842292. doi: 10.3389/fsurg.2022.842292

29. Hang D, Zhou J, Qin N, Zhou W, Ma H, Jin G, et al. A novel plasma circular RNA circFARSA is a potential biomarker for non-small cell lung cancer. *Cancer Med* (2018) 7(6):2783–91. doi: 10.1002/cam4.1514

30. Peng Z, Hu Q, Fang S, Zhang X, Hong X, Tao L, et al. Circulating circTOLLIP serves as a diagnostic biomarker for liquid biopsy in non-small cell lung cancer. *Clin Chim Acta* (2021) 523:415–22. doi: 10.1016/j.cca.2021.10.038

31. He Y, Liu Y, Cha N, Gao Y, Li F, Zhang M, et al. Exosomal circ_0048856 derived from non-small cell lung cancer contributes to aggressive cancer progression through downregulation of miR-1287-5p, pathol. *Res Pract* (2022) 232:153659. doi: 10.1016/j.prp.2021.153659

32. Chen F, Huang C, Wu Q, Jiang L, Chen S, Chen L. Circular RNAs expression profiles in plasma exosomes from early-stage lung adenocarcinoma and the potential biomarkers. *J Cell Biochem* (2020) 121(3):2525–33. doi: 10.1002/jcb.29475

33. Huang L, Rong Y, Tang X, Yi K, Wu J, Wang F. Circular RNAs are promising biomarkers in liquid biopsy for the diagnosis of non-small cell lung cancer. *Front Mol Biosci* (2021) 8:625722. doi: 10.3389/fmolb.2021.625722

34. Yang Q, Chen L, Yang L, Huang Y. Diagnostic and prognostic values of circular RNAs for lung cancer: A meta-analysis. *Postgrad. Med J* (2021) 97 (1147):286–93. doi: 10.1136/postgradmedj-2019-137178

35. Xiao Z, Chen X, Lu X, Zhong X, Ling Y. Accuracy evaluation of circular RNA in diagnosing lung cancer in a Chinese population. *Dis Markers* (2019) 2019:7485389. doi: 10.1155/2019/7485389

36. Wang C, Jiang Y, Lei Q, Wu Y, Shao J, Pu D, et al. Potential diagnostic and prognostic biomarkers of circular RNAs for lung cancer in China. *Biomed Res Int* (2019) 2019:8023541. doi: 10.1155/2019/8023541

37. Li Y, Zheng Q, Bao C, Li S, Guo W, Zhao J, et al. Circular RNA is enriched and stable in exosomes: A promising biomarker for cancer diagnosis. *Cell Res* (2015) 25(8):981–4. doi: 10.1038/cr.2015.82

38. Hu X, Wu D, He X, Zhao H, He Z, Lin J, et al. circGSK3β promotes metastasis in esophageal squamous cell carcinoma by augmenting β-catenin signaling. *Mol Cancer* (2019) 18(1):160. doi: 10.1186/s12943-019-1095-y

39. Omid-Shafaat R, Moayeri H, Rahimi K, Menbari MN, Vahabzadeh Z, Hakhamaneshi MS, et al. Serum circ-FAF1/Circ-ELP3: A novel potential biomarker for breast cancer diagnosis. *J Clin Lab Anal* (2021) 35(11):e24008. doi: 10.1002/jcla.24008

40. Yu J, Ding WB, Wang MC, Guo XG, Xu J, Xu QG, et al. Plasma circular RNA panel to diagnose hepatitis b virus-related hepatocellular carcinoma: A large-scale, multicenter study. *Int J Cancer* (2020) 146(6):1754–63. doi: 10.1002/ijc.32647

41. Shi P, Song H, Ding X. Reduced expression of circRNA hsa_circ_001888 in gastric cancer and its clinical significance, J. *Clin Lab Anal* (2021) 35(9):e23953. doi: 10.1002/jcla.23953

42. Luo YH, Yang YP, Chien CS, Yarmishyn AA, Ishola AA, Chien Y, et al. Plasma level of circular RNA hsa_circ_0000190 correlates with tumor progression and poor treatment response in advanced lung cancers. *Cancers (Basel)* (2020) 12 (7):1740. doi: 10.3390/cancers12071740

frontiers | Frontiers in Oncology

TYPE Review
PUBLISHED 27 October 2022
DOI 10.3389/fonc.2022.1027664

Check for updates

OPEN ACCESS

EDITED BY
Qingqing Zhu,
The First Affiliated Hospital of Soochow University, China

REVIEWED BY
Francesco Pepe,
University of Naples Federico II, Italy
Yoshiko Hashii,
Osaka University, Japan

*CORRESPONDENCE
Kaihua Lu
lukaihua@njmu.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

RECEIVED 25 August 2022
ACCEPTED 11 October 2022
PUBLISHED 27 October 2022

CITATION
Fang X, Yu S, Jiang Y, Xiang Y and Lu K (2022) Circulating tumor DNA detection in MRD assessment and diagnosis and treatment of non-small cell lung cancer.
Front. Oncol. 12:1027664.
doi: 10.3389/fonc.2022.1027664

COPYRIGHT
© 2022 Fang, Yu, Jiang, Xiang 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.

Circulating tumor DNA detection in MRD assessment and diagnosis and treatment of non-small cell lung cancer

Xiaoxu Fang†, Shaokun Yu†, Yingying Jiang, Yan Xiang and Kaihua Lu*

Department of Oncology, The First Affiliated Hosptial of Nanjing Medicial University, Nanjing, China

Circulating tumor DNA (ctDNA) has contributed immensely to the management of hematologic malignancy and is now considered a valuable detection tool for solid tumors. ctDNA can reflect the real-time tumor burden and be utilized for analyzing specific cancer mutations *via* liquid biopsy which is a non-invasive procedure that can be used with a relatively high frequency. Thus, many clinicians use ctDNA to assess minimal residual disease (MRD) and it serves as a prognostic and predictive biomarker for cancer therapy, especially for non-small cell lung cancer (NSCLC). Advanced methods have been developed to detect ctDNA, and recent clinical trials have shown the rationality and feasibility of ctDNA for identifying mutations and guiding treatments in NSCLC. Here, we have reviewed recently developed ctDNA detection methods and the importance of sequence analyses of ctDNA in NSCLC.

KEYWORDS

ctDNA, liquid biopsy, NSCLC, therapy monitoring, minimal residual disease (MRD)

Introduction

Lung cancer is the second most common cancer worldwide, responsible for the maximum number of cancer deaths (1). Non-small cell lung cancer (NSCLC) represents approximately 85% of diagnosed lung cancers; lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) are the two most common subtypes (2). However, the development and application of precision therapy, including targeted therapy and Immune checkpoint inhibitor (ICI) therapy, have fundamentally altered the management of NSCLC patients. Targeted therapy has shown potential in the treatment of patients with driver gene alterations such as epidermal growth factor receptor (EGFR) mutations, anaplastic lymphoma kinase (ALK) fusions, human epidermal growth factor receptor 2 (HER2) mutations, ROS1 fusions, MET

amplification, BRAF mutations, and RET fusions. It is now widely used in daily clinical practice (3). ICI therapy, which suppresses programmed cell death-1 (PD-1) or programmed cell death ligand-1 (PD-L1), has also been successful in prolonging the life of patients (4).

Clinical diagnosis requires a solid biopsy in order to determine tumor histology and staging. Compared with tissue biopsy, liquid biopsy is a non-invasive way to identify patients who might response to therapy, to dynamically monitor treatment effect and to unveil resistance mechanism. Liquid biopsy could typically detect circulating tumor cells (CTCs), circulating tumor DNA (ctDNA), exosomes, microRNAs (miRNA), peripheral blood circulating RNA, tumor-educated blood platelets (TEPs), and circulating tumor vascular endothelial cells (CTECs). ctDNA is one of the most commonly detected biomarkers (5).

Circulating cell-free DNAs (cfDNA) are DNA fragments ranging from 150 to 200 base pairs in length mainly derived from apoptotic or necrotic cells (6). Tumor cells also release circulating tumor DNA (ctDNA), accounting for <0.01% of total cfDNA, which need detection techniques with high sensitivity. Besides the traditional quantitative or real-time PCR (qPCR) and next-generation sequencing (NGS), other recently introduced methods to analyze ctDNA are advanced PCR-based techniques such as digital PCR (dPCR), droplet digital PCR (ddPCR), beads emulsion amplification magnetics (BEAMing), NGS-based techniques such as tagged amplicon deep sequencing (TAM-Seq), safe-sequencing (Safe-Seq), cancer personalized profiling by deep sequencing (CAPP-Seq), and Phased variant enrichment and detection sequencing (PhasED-seq). These will be briefly explained below.

Minimal residual disease (MRD) is a disease status in patients that escapes clinical observation by radiology. In oncology, MRD represents early tumor development and tumor relapse which needs to be urgently detected and assessed (7). In MRD detection, liquid biopsy of these tumor - derived factors plays an important role in clinical application. First of all, liquid biopsy can be used for early cancer screening which lacks detectable abnormalities found by radiology approaches. Secondly, liquid biopsy could monitor micrometastatic disease to assess the risk of disease recurrence after a radical treatment. Finally, the dynamic characterization of tumor burden and disease biological changes could clarify drug resistance mechanisms and guide the treatment strategies. (Figure 1) (8). More recently, ctDNA from a liquid biopsy has shown showing their potential to be a reliable plasma-based biomarker for MRD. Quantitative characterization of ctDNA *via* liquid biopsy has been associated with clinical and pathologic features of cancer, including stage, tumor burden, vascularization, and response to therapy. ctDNA can help detect the mutations and activity of different tumor sub-clones,which tissue biopsy cannot because of tumor heterogeneity (9). Moreover, the short half-life of ctDNA ensures that the detection results are in real-time. The molecular precision of longitudinal tumor surveillance *via* serial ctDNA measurement enables the identification of mutations that drive cancer progression and treatment resistance.

In this review, we will introduce several commonly used ctDNA detection approaches and discuss the clinical application of ctDNA-based MRD evalution.

Recently developed detection techniques

PCR assay

Detection and quantitation of specific nucleic acid sequences using PCR is fundamental to a large body of research and a growing number of molecular diagnostic tests. The first generation of PCR users performed end-point analysis by gel electrophoresis to obtain qualitative results. The advent of real-time PCR spawned a second generation. rtPCR is an analogue measurement based on monitoring amplification after each cycle of PCR using fluorescence probes. The point at which the reaction fluorescence crosses an intensity threshold is called the cycle threshold(Ct). As many factors can influence PCR efficiency and hence the Ct value, the accuracy and precision of real-time PCR can vary widely (10). Vogelstein and Kinzler introduced a new form of PCR called digital PCR (dPCR) in 1999 (11). Compared with those conventional PCRs, dPCR partitions samples into multiple parallel quantitative PCR reactions within separate compartments and therefore improves sensitivity, absolute quantification, and rare allele detection (12, 13). However, the large reaction volume and the limited number of compartments to minimize the dimensions of the chip have greatly limited its possible clinical applications (10, 14), and it is believed that ddPCR might overcome these limitations.

Droplet digital PCR (ddPCR) uses aqueous droplets with volumes ranging from a few femtoliters to nanoliters dispersed in oil to compartmentalize PCR reactions, having a theoretically unlimited number of compartments (10). In addition, ddPCR needs only a single reaction tube (10). Nanoliter-sized droplet technology paired with digital PCR (ddPCR) holds promise for highly precise, absolute nucleic acid quantification. Hindson et al. compared the microRNA quantification by ddPCR and real-time PCR which revealed greater precision (coefficients of variation decreased 37–86%) and improved day-to-day reproducibility (by a factor of seven) of ddPCR but with comparable sensitivity (15). Frank Diehl et al. found another droplet-based digital PCR named BEAMing (beads, emulsion, amplification, and magnetics) in 2006. BEAMing couples oligonucleotide primers to beads and distributes beads to compartments. After amplification, every compartment contains a bead binding to thousands of copies of the initial DNA molecule. The DNA can then be released from the beads and analyzed with flow cytometry or optical scanning instruments to locate mutation (16, 17). A recent LungBEAM study demonstrated the great potential of BEAMing in optimizing treatment in patients with advanced NSCLC (18).

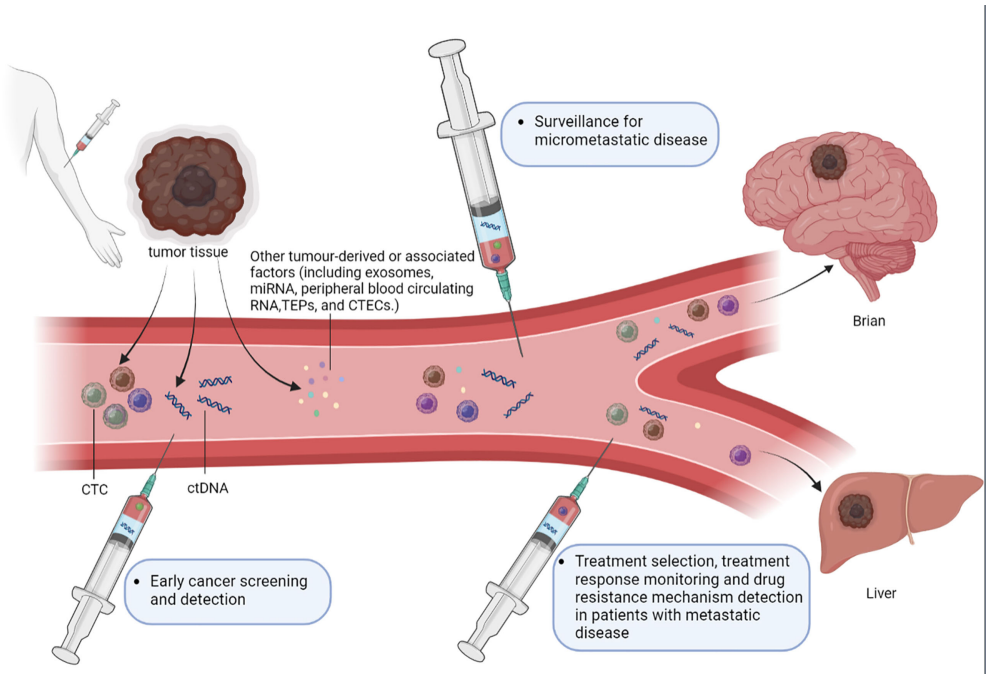


FIGURE 1
The role of liquid biopsy in MRD detection. Liquid biopsy typically detects circulating tumor cells (CTCs), circulating tumor DNA (ctDNA), exosomes, microRNAs (miRNA), peripheral blood circulating RNA, tumor-educated blood platelets (TEPs) and circulating tumor vascular endothelial cells (CTECs). Liquid biopsy assays of these tumor-derived factors can serve several purposes in the MRD detection. (1) Early cancer screening and detection, liquid biopsy approaches could also be used to further investigate abnormalities detected on imaging examinations. (2) Surveillance for micrometastatic disease following curative-intent treatment of a primary tumor, in order to evaluate the risk of disease recurrence and enable timely management of recurrent disease. (3) Guiding the selection of the most appropriate treatment, monitoring treatment responses and detecting the resistance mechanisms in patients with overt metastatic disease through dynamic characterization of changes in tumor burden and disease biology.

Presently, ddPCR and BEAMing are the two most commonly used PCR techniques in clinics to detect ctDNA, their reports must match. Ben O'Leary and his colleagues collected plasma from patients with advanced breast cancer and assessed ESR1 and PIK3CA mutations in ctDNA using both PCR techniques simultaneously. ESR1 mutation was calculated as 24.2% (88/363) with BEAMing and 25.3% (92/363) with ddPCR, (κ = 0.91; 95% CI, 0.85-0.95). The result for PIK3CA mutation was 26.2% (95/363) with BEAMing and 22.9% (83/363) with ddPCR, (κ = 0.87; 95% CI, 0.81– 0.93), showing consistency of results from BEAMing and ddPCR (19). Despite droplet-based digital PCR being highly sensitive, it can only detect known mutations and needs customed assays (20), which restrict its clinical applicability.

Targeted next generation sequencing approaches

The development of NGS including whole exome sequencing (WES) and whole genome sequencing (WGS), has facilitated cancer diagnosis over the past decade through samples from a tissue biopsy (21). However, the sensitivity of NGS for single nucleotide variants (SNV) detection is about 4% to 10% (22), which is enough for tissue samples but makes it hard to detect rare mutations in ctDNA for its extremely low percentage in cfDNA. To maintain NGS's covering broad areas across the genome and meanwhile improve the sensitivity, targeted NGS approaches that detect specific areas of the genome were developed (23).

Forshew et al. established a technique named tagged-amplicon deep sequencing (TAm-Seq; 2012) and successfully applied it to detect ctDNA mutations in patients with metastatic breast cancers and ovary cancers (24). TAm-Seq can detect cancer mutations with allele frequencies as low as 2% and sensitivity and specificity as high as 97% (24). Kinde et al. pioneered Safe Sequencing System (Safe-SeqS; 2011), in which they tagged each template molecule with a 12- or 14- base unique identifier (UID), then amplified the tagged molecules with two cycles of amplicon-specific PCR to create UID families, and sequenced the amplified product redundantly with NGS (25). This redundant sequencing approach makes Safe-SeqS detect rare mutations with high specificity. Moreover, its unique algorithm increases the accuracy of the base calling and reduces the error rate to an average of 2×10^{-4} errors/bp although efficacy is still limited by artifactual mutations

occurring during the PCR as well as any residual base-calling errors (25).

Newman et al. developed cancer personalized profiling by deep sequencing (CAPP-Seq; 2014) with ultrasensitive detection of ctDNA. CAPP-Seq utilizes DNA probes to hybridize and capture ctDNAs for its quantification and sequencing. These probes are designed for regions with high driver mutation frequencies in certain cancer types (26). This method can even detect 0.02% of cfDNA and ctDNA in patients with early or advanced stages of NSCLC (26). To further improve the efficiency of ctDNA detection, Newman et al. upgraded CAPP-Seq to integrate digital error suppression-enhanced CAPP-seq (iDES-enhanced CAPP-seq), which tags each template molecule with a UID just like the Safe-seqS to reach a detection limit of 0.001% and a specificity of 96% (27).

Phased variant enrichment and detection sequencing (PhasED-seq; 2021) is the most recent method that uses multiple somatic mutations in individual DNA fragments to improve the sensitivity of ctDNA detection. PhasED-seq can detect less than 0.0001% of tumor DNA, which is better than any earlier approaches (28). David et al. demonstrate that PhasED-seq can meaningfully improve detection of ctDNA in clinical samples both during therapy and before disease relapse. They analyzed serial samples from a participant with stage III lung adenocarcinoma treated with chemoradiotherapy. CAPP-seq and PhasED-seq detected similar ctDNA levels before therapy; however, three samples after treatment initiation had undetectable ctDNA by CAPP-seq before ctDNA re-emerged at the time of biopsy-confirmed recurrent disease. Using PhasED-seq, they observed molecular residual disease in 3/3 (100%) of samples that were undetected by single-nucleotide variants (SNVs), with tumor fraction as low as 0.00016% (28).

Clinical laboratories are increasingly developing and deploying NGS tests, ranging from targeted 'hotspot' panels to comprehensive genome-scale platforms. Ahmet et al.developed and implemented MSK-IMPACT, a hybridization capture–based NGS panel with distinct advantages over small panels for detecting all protein-coding mutations, copy number alterations (CNAs), and selected promoter mutations (29). Ivo et al. demonstrate that panel size is a critical parameter that influences confidence intervals (CIs) and cutoff values as well as important test parameters including sensitivity, specificity, and positive predictive value. Panels between 1.5 and 3 Mbp are ideally suited to estimate TMB with small CIs, whereas smaller panels tend to deliver imprecise TMB estimates for low to moderate TMB (0–30muts/Mbp) (30).

ctDNA and diagnosis of NSCLC

Early-stage diagnosis

According to the International Association for the Study of Lung Cancer (IASLC) lung cancer staging project, the 5-year survival of NSCLC diminishes rapidly as the disease stage progresses (82% for stage IA, 52% for stage IIA, 36% for stage IIIA and 6% for stage IV). Thus, the detection of early-stage NSCLC is urgent (31). However, early-stage NSCLC has few radiographic characteristics to be distinguished from benign solitary nodules, so the chances of false positives from radiology approaches are too high. Wong et al. assigned 10,061 candidates to the CANTOS (Canakinumab anti-inflammatory thrombosis outcome study), 71 of them developed lung cancer and each participant had deposited two plasma samples at two different time points during the study; one was at the baseline time point (collected at the beginning of the trial) and the other after the clinical diagnosis of lung cancer. The test of these samples indicated that patients with COSMIC (catalog of somatic mutations in Cancer) ctDNA mutations at baseline exhibited a shorter time to their lung cancer clinical diagnosis (407 days versus 837 days, P=0.011), indicating that mutations in ctDNA might predict an early-stage NSCLC to some extent (32).

A study analyzed ctDNA at different stages of NSCLC utilizing CAPP-seq and found that the diagnostic sensitivity of ctDNA was 64%, 82%, and 100% for tumor stages I, II, and III, respectively. A similar finding was reported from another study (57.9%, 66.7%, and 90% for tumor stages I, II and III, respectively), which implied a correlation between ctDNA levels and tumor volume and outlined the difficulty of early NSCLC detection (33, 34). However, Liang et al. established a method of DNA methylation profiling by high throughput DNA bisulfite sequencing that can distinguish malignant tumors from benign solitary nodules with a sensitivity of 79.5% (63.5%-90.7%) and a specificity of 85.2% (66.3% -95.8%) (35).

Detection of known mutations

The efficiency of targeted therapy depends on the precise detection of the driver gene mutations. Mack et al. tested plasma samples from 8388 patients and made a plasma-based comprehensive genomic profiling. Driver gene mutations were identified in 48% of patients, including EGFR mutations (26.4%), MET mutations (6.1%), BRAF mutations (2.8%), and fusions (ALK, RET, and ROS1; 2.3%) (36).

Although the golden standard guiding target therapy remains gene mutations detected from a tissue biopsy, non-invasive liquid biopsy utilizing ctDNA is sometimes preferred, and ctDNA increases the chances of identifying several targetable mutations, especially EGFR mutation (37, 38). However, it is crucial to clarify whether the mutations detected from ctDNA agree with those from tumor cell lesions (Table 1). A clinical trial study (NCT01203917) aimed to assess the efficacy and tolerability of gefitinib as first-line therapy for common EGFR mutations (19del, L858R, T790M) positive patients in stage III/IV NSCLC. Researchers found EGFR mutations to be

similar in tumor and plasma samples (ctDNA) with a sensitivity of 65.7% (95% CI: 55.8–74.7) and a specificity of 99.8% (95% CI: 99.0–100.0) (40), implying that plasma samples are useful to identify patients who might benefit from gefitinib when tumor tissue is unavailable. A similar comparison was designed in the FASTACT-2 study, and the sensitivity and specificity of the mutation detection were 75% and 96%, respectively (43). Cobas EGFR Mutation Test v2, a real-time PCR assay that can identify 42 different EGFR gene mutations, was the first approach approved by FDA to detect EGFR mutations in 2016 (52).

Pertaining to the oncogenic fusions, ctDNA reflects a high similarity with those detected in tissue samples. Horn et al. analyzed ALK fusions in tumor and plasma samples and found a concordance of 91% (20/22) between them (53). Wang et al. also detected ALK fusions in ctDNA from 19 out of 24 patients with ALK fusions in their tumor tissue, demonstrating a sensitivity of 79.2% (95%CI: 57.9%-92.9%). They could not detect ALK fusions in ctDNA from 36 patients without ALK fusions in their tumor tissue, implying that the specificity of the method was 100% (50). Plasma ROS1 fusions analysis also showed a 100% concordance with those observed in the tissue samples (54).

Vansteenkiste et al. found that the similarity in PIK3CA mutations between tissue and ctDNA samples was 55.3%. However, the concordance was 81.8% (9 of 11 samples) between ctDNA and metastatic tissue samples, compared with 44.4% (12 of 27 samples) between ctDNA and primary tissue, implying that ctDNA PI3K pathway mutations were more correlated with metastatic lesions than with primary tumor (55).

However, despite the high sensitivity of methods for detecting mutations in ctDNA, it is wise to retest tissue samples if the result is negative (56).

It is noteworthy that MRD detection can also be confounded by clonal hematopoiesis of indeterminate potential (CHIP). CHIP arises when age-dependent mutations accumulate in hematopoietic progenitor cells, leading to the formation of a genetically distinct subpopulation that contributes disproportionately to the population of mature blood cells. These distinct subclones have driver mutations and have been implicated in hematologic diseases. In the measurement of ctDNA, CHIP can result in false-positive results due to detection of non-reference variants in the blood plasma, which is especially problematic when the ctDNA mutant allele fraction is low in the setting of MRD detection. Thus, CHIP must be properly accounted for in order to specifically measure ctDNA, such as by sequencing matched PBMCs to similar depth, especially when using ultra-sensitive assays that are capable of achieving detection of low mutant allele fraction variants (57).

ctDNA and Treatment of NSCLC

Relapse after operation

The relapse after curative-intent resection has confused surgeons for years and ctDNA might be an early predictor of it. The DYNAMIC prospective study tested plasma ctDNA collected from 36 patients that underwent curative-intent lung resections 7 times, immediately before surgery (time A), after tumor resection [time B (5 minutes), time C (30 minutes), and time D (2 hours)] and after surgery [time P1 (1 day), time P2 (3 days), and time P3 (1 month)]. A rapid decrease in the content of ctDNA was found after the curative-intent resection (the mean mutant allele fraction at times A, B, C, and D was 2.72%, 2.11%, 1.14%, and 0.17%, respectively) which implied that the half-time of ctDNA is short and there is an association between ctDNA

TABLE 1 The mutation results obtained and compared between tumor and plasma DNA samples.

Approach	Gene/TMB	sensitivity	specificity	concordance	reference
ARMS (39)	EGFR (only exon 19 del, L858R mutation and T790M mutation)	65.7% (69/105)	99.8% (546/547)	94.3% (615/652)	(40)
ARMS SURVEYOR (41)	EGFR (only exon 19 del, L858R mutation and T790M mutation)	70% (21/30)	84.6% (11/13)	74% (32/43)	(42)
Cobas	41 EGFR mutations	75% (72/96)	96.5% (137/142)	87.8% (209/238)	(43)
real-time PCR	exon 19 del	64% (112/175)	96.4% (149/153)	79.6% (261/328)	(44)
	L858R mutation	55.1% (70/127)	99.5% (200/201)	82.3% (270/328)	
	T790M mutation	0 (0/1)	99.7% (326/327)	99.4% (326/328)	
ddPCR	EGFR (only exon 19 del, L858R mutation and T790M mutation)	70% (182/260)	93.9% (123/131)	78.0% (305/391)	(45)
BEAMing	EGFR (only exon 19 del and L858R mutation)	70.9% (78/110)	/	/	(18)
PNA-based real-time PCR	47 EGFR mutations	66.7% (34/51)	87.4% (125/142)	82.0% (159/193)	(46)
BEAMing	EGFR (only exon 19 del and L858R mutation)	81.5% (44/54)	/	/	(47)
	EGFR T790M mutation	83.7% (41/49)	/	/	
PNA-based real-time PCR	EGFR (only exon 19 del, L858R, G719X and L861Q mutation)	59.3% (48/81)	92.0% (230/250)	84.0% (278/331)	(48)
NGS	ALK fusions	54.2% (13/24)	99.8% (472/473)	97.6% (485/497)	(49)
capture-based NGS	ALK fusions	79.2% (19/24)	100% (36/36)	91.7% (55/60)	(50)
[a]Targeted NGS	bTMB	50.7% (77/152)	86.0% (172/200)	70.7% (249/352)	(51)

[a]The cut-off of bTMB is 20 mut/Mb and that of tTMB is 10 mut/Mb.

and tumor volume. In addition, the detection of ctDNA at time P2 (278 days versus 637 days, P=0.002) and time P3 (295 days versus 662 days, P=0.003) rather than time P1 (528 days versus 543 days, P=0.657) was negatively correlated with recurrence free survival (RFS) of patients; similar correlations were observed between ctDNA detection and overall survival (OS) (58).

Xia et al. analyzed ctDNA in another prospective, multicenter study (LUNGCA-1; 2021) on NSCLC surgery patients. They found that detectable ctDNA before operation (RFS; HR=4.2, 95%CI: 2.6-6.7; P < 0.001) or at 3 days and/or 1 month after operation (RFS; HR=11.1, 95%CI: 6.5-19.0; P < 0.001) was a robust predictor for relapse in patients with stage I–III NSCLC. Moreover, ctDNA status was tightly associated with the benefit of postoperative adjuvant therapy — ctDNA-positive patients who received adjuvant therapies had improved RFS over those that did not receive (RFS; HR=0.3, 95%CI: 0.1-0.8; P=0.008), while ctDNA-negative patients receiving adjuvant therapies had impaired RFS than those that did not receive (RFS; HR=3.1, 95%CI: 1.7-5.5; P < 0.001) (7).

Chaudhuri et al. prespecified "MRD landmark" as the ctDNA status following the first phlebotomy of curative-intent resection and within 4 months from the end of therapy, progression at 36 months after the MRD landmark was 100% and 7% in patients with detectable and undetectable ctDNA MRD (HR=43.4, 95%CI=5.7–341; P < 0.001), respectively (59).

Yilong Wu et al. elucidated the role of MRD monitoring in patients with stage I to IIIA NSCLC after definite surgical resection. Patients with undetectable MRD at landmark or longitudinal time points had better disease-free survival (DFS) than those with detectable MRD [landmark: unreached vs. 12.1 months (4.7–19.5); HR = 0.08; 95% CI, 0.02–0.33; longitudinal: unreached vs. 15.9 months (13.8–18.0); HR = 0.02; 95% CI, 0.01–0.05]. 96.8% of patients with longitudinal undetectable MRD were still disease-free at the last follow-up and had nothing to do with the clinical stage, thus it may represent the potentially cured population, which has important application value for the treatment of early lung cancer in the future. Because MRD status reflected the tumor load, adjuvant therapy was found to confer a survival benefit for patients with detectable MRD (P = 0.022; HR = 0.34; 95% CI, 0.12-0.88) (60).

Development of resistance due to targeted therapy

Driver gene mutations might exhibit changes during tumor development or treatment that can lead to resistance to the drugs, which limits the long-term use of targeted therapy. Thus, the new driver gene mutations need to be detected through a re-biopsy. Liquid biopsy can specifically detect the new gene mutations, and this can be used to predict targeted therapy resistance development in patients.

Approximately half or more NSCLC patients with EGFR mutations who develop resistance to the first- and second-generation EGFR-tyrosine kinase inhibitors (TKI) will develop a secondary EGFR T790M mutation in the tumor (61). Additionally, the ctDNA T790M mutation is more likely to be seen in patients with an initial EGFR del19 mutation compared with the EGFR L858R mutation (62). ctDNA analysis may help in predicting such resistance and directing the use of subsequent therapy such as the use of osimertinib, an oral, irreversible third-generation EGFR-TKI, approved by FDA in 2015 (42, 63). LiquidLung-O-Cohort 2 study screened ctDNA from patients with EGFR T790M mutation with a detection sensitivity of 56.8% (64). Serial monitoring of EGFR mutation in ctDNA is able to detect EGFR T790M mutation much earlier (range: 15-344 days) than clinical manifestation of the disease progression (65). However, since osimertinib is considered standard first-line therapy for NSCLC patients with EGFR mutations (66), some patients on osimertinib would inevitably develop new mutations (detected in ctDNA), including EGFR C797S mutation, MET amplification, HER2 exon 20 insertions, $BRAF^{V600E}$ mutation, PIK3CA mutation, and EGFR amplification and thus be resistant to the drug (67–71).

Dagogo et al. analyzed plasma and tumor samples from patients with progressed ALK-positive NSCLC treated with alectinib (2019). There was no difference in ALK mutation frequency (67% versus 63%), but ctDNA was more likely to harbor ≥2 ALK mutations (24% versus 2%, P=0.004). However, ALK L1196M, a gatekeeper mutation that leads to resistance to crizotinib, showed little prevalence between tumor DNA and ctDNA (2% versus 22%, P=0.008), which implies that ctDNA can predict ALK-TKI resistance sometimes. A similar phenomenon was found in those with lorlatinib, showing a promoted acquisition of ALK resistance mutations after sequential treatment with increasingly potent ALK-TKIs (72). Secondary ALK mutations such as ALK G1202R, ALK G1269A, and ALK L1196M were found in ctDNA through NGS, causing genetic resistance to first- and second-generation ALK-TKIs (50, 72–74).

Prognosis and treatment response after target therapy

ctDNA is commonly used to monitor the benefit of the treatment and predict progression *via* liquid biopsy. Several studies have found a significant association between the quantitative changes in ctDNA, the response of cancer to the targeted therapy, and the prognosis of NSCLC (Table 2).

High plasma cfDNA is associated with poor OS (16.0 months versus 28.6 months, P=0.030) and increased risk of death (HR=1.23, 95% CI: 1.01-1.50; P=0.045) (78). Bordi et al. defined a cut-off of 2200 copies/ml generated by means of ROC analysis and found a lower number of mutations (< 2200 copies/mL; at baseline) are associated with better progression free survival (PFS; 17.8 months versus 4.3 months, p=0.022) and OS (23.6 months versus 7.7 months, p= 0.016) (79).

Baseline EGFR T790M mutation detection in ctDNA might correlate with a larger baseline tumor size (56 mm for T790M

TABLE 2 ctDNA as a treatment response predictor.

Patients	Treatment	Standard	Test time	Approach	Findings	reference
Stage I– IIIA	curative-intent resections	RFS	Day 3	NGS	278d vs. 637d for ctDNA detected and EGFR undetected patients (HR=7.55; 95% CI: 2.09–27.27, P=0.002)	(58)
		OS			434d vs. 720d for ctDNA detected and EGFR undetected patients (HR=14.22; 95% CI: 1.58–128.15, P=0.018)	
ctDNA EGFR mut^{+} at baseline	gefitinib	PFS	Week 8	ddPCR	11.0m vs. 2.1m for ctDNA EGFR mut^{-} and EGFR mut^{+} patients (HR=0.14; 95% CI: 0.08–0.23, P < 0.0001)	(45)
ctDNA EGFR mut$^{+/-}$ at baseline	erlotinib	PFS	Week 8 (Cycle 2)	Cobas	11.0m vs. 5.7m for ctDNA EGFR mut^{-} and EGFR mut^{+} patients (HR=0.28; 95% CI: 0.15–0.52, P < 0.0001)	(75)
		OS			30.1m vs. 15.8m for ctDNA EGFR mut^{-} and EGFR mut^{+} patients (HR=0.35; 95% CI: 0.19–0.64, P < 0.0001)	
ctDNA EGFR mut^{+} at baseline	erlotinib	PFS	Cycle 3	Cobas	12.0m vs. 7.2m for ctDNA EGFR mut^{-} and EGFR mut^{+} patients (HR=0.32; 95% CI: 0.21–0.48, P < 0.0001)	(43)
		OS			31.9m vs. 18.2m for ctDNA EGFR mut^{-} and EGFR mut^{+} patients (HR=0.51; 95% CI: 0.31–0.84, P=0.0066)	
ctDNA T790M mut^{+} at baseline, progress on 1/2G EGFR-TKI	osimertinib	PFS	Cycle 4	ddPCR	4.9m vs. 15.9m for ctDNA EGFR MF-high and EGFR MF-low patients (HR=4.54; 95% CI: 2.23–9.23, P < 0.0001)	(71)
				Cobas	6.3m vs. 17.2m for ctDNA EGFR MF-high and EGFR MF-low patients (HR=2.87; 95% CI: 1.52–5.42, P=0.0012)	
				NGS	4.3m vs. 14.5m for ctDNA EGFR MF-high and EGFR MF-low patients (HR=6.00; 95% CI: 2.87–12.55, P < 0.0001)	
ctDNA EGFR mut^{+} at baseline	osimertinib + bevacizumab	PFS	Week 6	ddPCR	16.2m vs. 9.8m for ctDNA EGFR mut^{-} and EGFR mut^{+} patients P=0.04	(76)
		OS			NR vs. 10.1m for ctDNA EGFR mut^{-} and EGFR mut^{+} patients P=0.002	
StageIIIB/IV, progressed following two or more systemic treatments	durvalumab	PFS	Week 6	targeted NGS	5.6m vs. 1.9m for ctDNA dVAF<0 and dVAF≥0 HR 0.26 (95% CI, 0.12–0.54).	(77)
		OS			NR vs. 8.7m for ctDNA dVAF<0 and dVAF≥0 HR 0.23 (95% CI, 0.09–0.61).	

(+) versus 39 mm for T790M (-); P < 0.0001) and a higher probability of extra thoracic metastasis [58% M1b for T790M (+) versus 39% M1b for T790M (-); P = 0.002] (80). Moreover, tissue T790M positive patients without detectable T790M mutation in the ctDNA had a longer PFS, which might be attributed to a lower tumor burden (80).

Identification of EGFR mutation in ctDNA before the start of the treatment procedure helps to select patients who might benefit from EGFR-TKI treatment, and monitoring ctDNA consistently for further EGFR mutation helps to predict the outcome of current treatment and the patient's prognosis (45, 81). In patients treated with erlotinib and assessed to be stable disease (SD), undetectable ctDNA at week 8 is correlated with survival improvement (PFS: HR=0.27, 95%CI: 0.13-0.57, p<0.0001; OS: HR=0.40, 95% CI 0.20–0.80, p=0.009) (75). In NSCLC patients with progression after EGFR-TKI therapy, chest- or brain-limited disease has a significantly higher rate of ctDNA T790M mutation than the others (P<0.001). This showed that both ctDNA T790M mutation status and TKI treatment failure can predict prognosis (82). Furthermore, the persistence of EGFR mutation in ctDNA at 6 weeks in patients treated with osimertinib was associated with shorter PFS (9.8 months versus 16.2 months, P=0.04) (76), while loss of EGFR exon 19 deletion or L858R mutation post-treatment appears to correlate with longer PFS (14.7 months versus 5.5 months) (67). The monitering of ctDNA for EGFR mutations in NSCLC

patients at treatment cycle 4 is optimal for predicting the treatment outcomes for patients receiving osimertinib (71).

Patients with EML4-ALK fusion variants 1 detected in ctDNA at baseline exhibited longer PFS than those with EML4-ALK fusion variants 3 [8.2 months (95% CI: 2.1–11.7) versus 1.9 months (95% CI: 1.8-not estimable)] (53). In the ALTA-1L study, researchers found detectable baseline EML4-ALK fusion variant 3 rather than variant 1 in ctDNA, which was associated with poor PFS in patients treated with ALK TKI [crizotinib: HR: 3.42 (1.56–7.50), P=0.002; brigatinib: HR: 2.45 (1.07–5.60), P=0.033] (83).

Prognosis and treatment outcome after immunotherapy

Even though long-term positive responses have been observed in NSCLC patients receiving ICI therapy, the majority of them become refractory with an eventual unfavorable clinical outcome (84). Rapid as well as sensitive detection of dynamic changes in the ctDNA might help to identify NSCLC patients and plan appropriate immunotherapy for them (85). (Table 2)

Goldberg et al. defined ctDNA response as a >50% decrease in mutant allele fraction from its baseline (2018). In ICI therapy receiving patients with metastatic NSCLC, ctDNA response greatly agreed with the radiographic response (κ=0.753), and benefits could be assessed faster from ctDNA than radiographically (median 24.5 days versus median 72.5 days). Additionally, a ctDNA response is associated with long-term treatment benefit (205.5 days versus 69 days; P<0.001) as well as better prognosis (PFS: HR=0.29, 95%CI: 0.09–0.89, P=0.03; OS: HR=0.17, 95%CI: 0.05–0.62; P= 0.007) (86). Similar to target therapy, patients with undetectable levels of ctDNA were demonstrated to have significantly longer PFS (P=0.001) and OS (P=0.008) compared with those with no evidence of ctDNA clearance (85). Hellmann et al. tested ctDNA of 31 patients with advanced NSCLC and had achieved long-term benefit from ICI therapy (PFS≥12 months) at a median time of 26.7 months after the initiation of therapy. They found 25/27 (93%) patients with ctDNA negative remained progression-free, while in 4 patients with ctDNA positive, the desease eventually progressed (87).

Nabet et al. established an approach named DIREct-On (Durable Immunotherapy Response Estimation by immune profiling and ctDNA- On-treatment) to predict whether patients with NCSLC would show durable clinical response to ICI therapy. DIREct-On incorporated pre-treatment ctDNA and immune profiling with early on-treatment ctDNA response assessment and could get an accuracy of 92% to identify the potential patients obtaining benefit (88).

A high tumor mutation burden (TMB) and microsatellite instability (MSI) are demonstrated to correlate with a better response to immunotherapy in NSCLC (89–91).,TMB is measured from tumor tissue traditionally. Si et al. measured and compared TMB from tissue (tTMB) and plasma (bTMB) samples and found a positive correlation between bTMB (using a cutoff of 20 mut/Mb) and tTMB (using a cutoff of 10 mut/Mb) values (P<0.0001, χ^2 test). They also found that higher bTMB was also associated with the clinical benefits of immunotherapy (51). When the bTMB cut-off point was set to 6, patients with higher bTMB showed superior PFS (NR versus 2.9 m; HR=0.39, 95%CI: 0.18-0.84, P=0.01) (92). Goldberg et al. found that bTMB is independently predictive of the immunotherapy outcome benefits without association with high PD-L1 expression. They further discovered that bTMB=16 mut/Mb is a clinically meaningful cut-off point in NSCLC and patients with bTMB≥16 mut/Mb benefited from a second-line immunotherapy rather than chemotherapy (PFS was 4.2m in the atezolizumab arm and 2.9m in the docetaxel arm, HR=0.57, 95%CI: 0.33–0.99; OS was 13.0m and 7.4m, HR=0.56, 95%CI: 0.31–0.99) (93). Similarly, Georgiadis et al. defined ctDNA bTMB≥10 mut/Mb of the whole exome as bTMB-high and demonstrated that bTMB-high before immunotherapy predicted a better PFS (HR=0.23, 95% CI, 0.07–0.63, P=0.003) and OS (HR=0.26, 95% CI, 0.08–0.72, P=0.008) in pan-cancer. Additionally, patients with blood MSI also had a better PFS (HR=0.21, 95% CI, 0.08–0.54, P=0.001) and OS (HR=0.41, 95% CI, 0.16–1.05, P=0.063) than those with microsatellite stability (MSS) (94).

Variant allele frequencies (VAF) in ctDNA can also predict immunotherapy response as an alteration of TMB. Patients with decreased VAF at week 6 of the treatment had a mean reduced tumor volume by 39%, while those with a high VAF had a mean increased tumor volume by 36% (77). Additionally, a decrease of VAF at week 6 of the immunotherapy implied a longer PFS and OS (77).

Discussion and limitation

Several clinical trials have demonstrated definite correlations between ctDNA levels and NSCLC patients' medical status, including tumor sizes, recurrence post operations, choice of treatments, treatment response and prognosis. Thus, ctDNA could help to guide clinicians in selecting appropriate therapies for each patient: whether to utilize adjuvant therapies after curative-intent resections, how to select a treatment that could benefit a patient maximumly and how to evaluate treatment efficiency and diagnose drug resistance on time. The short half-life of ctDNA enables it to exhibit a real-time status of a dynamic disease and overcome temporal tumor heterogeneity. Additionally, as ctDNA could represent spatial heterogeneity better than primary tumors or metastatic lesions biopsy, ctDNA has an inherent advantage in monitoring a patient's condition. Undoubtedly, ctDNA can serve as a predictor of MRD and can be frequent applied in the medical management of NSCLC patients. At the same time, there were a lot of research about MRD in predicting the recurrence trajectory of early lung cancer

and the curative effect of consolidation immunotherapy, which show great clinical application prospect.

Based on this, we put forward some ideas. For lung cancer patients with driver gene mutations after radical resection, ctDNA-based MRD monitoring to guide the use of targeted drugs, rather than continuous drug use mode, can theoretically delay the development of drug-resistant clones of tumor targeted therapy, thereby delaying drug resistance? At the same time, the treatment burden of patients can be reduced; For patients with inoperable locally advanced NSCLC, after radical therapy, ctDNA-based MRD monitoring can be used to guide immunodrug maintenance therapy, which can not only predict the population benefiting from ICI consolidation therapy, but also reduce the treatment burden. Starting from a population of patients with advanced targeted therapy, MRD monitoring should be used to guide the use of targeted drugs after patients have achieved complete remission or local treatment for oligometrics.

However, some limitations can hamper the wide use of ctDNA. Early-stage NSCLC remains tough to be detected by ctDNA mainly because of its extremely low concentrations due to small tumor sizes. Thus, highly sensitive methods need to be developed. Moreover, most of the trials that utilize ctDNA to plan the targeted therapy and predict treatment response or prognosis focused on the most common driver gene mutation, EGFR mutations. The use of ctDNA in patients with ALK fusions, MET amplification, HER2 mutations, and other rare mutations still needs to be studied in detail.

To conclude, a transformation in the management of NSCLC patients by analyzing ctDNA is afoot.

Author contributions

XF, SY, YJ, YX, KL drafted the manuscript, conceived and designed the study, and accomplished the revision of manuscript for important intellectual content. KL obtained funding. All authors contributed to the article and approved the submitted version.

Funding

This study is supported by the National Natural Science Foundation of China (81902327, 82172708); and the Jiangsu National Natural Science Foundation (BK20191064).

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

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. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, et al. The 2015 world health organization classification of lung tumors: Impact of genetic, clinical and radiologic advances since the 2004 classification. *J Thorac Oncol* (2015) 10:1243–60. doi: 10.1097/jto.0000000000000630

3. Kris MG, Johnson BE, Berry LD, Kwiatkowski DJ, Iafrate AJ, Wistuba II, et al. Using multiplexed assays of oncogenic drivers in lung cancers to select targeted drugs. *Jama* (2014) 311:1998–2006. doi: 10.1001/jama.2014.3741

4. Doroshow DB, Sanmamed MF, Hastings K, Politi K, Rimm DL, Chen L, et al. Immunotherapy in non-small cell lung cancer: Facts and hopes. *Clin Cancer Res* (2019) 25:4592–602. doi: 10.1158/1078-0432.Ccr-18-1538

5. Li W, Liu J-B, Hou L-K, Yu F, Zhang J, Wu W, et al. Liquid biopsy in lung cancer: significance in diagnostics, prediction, and treatment monitoring. *Mol Cancer* (2022) 21:25. doi: 10.1186/s12943-022-01505-z

6. Corcoran RB, Chabner BA. Application of cell-free DNA analysis to cancer treatment. *N Engl J Med* (2018) 379:1754–65. doi: 10.1056/NEJMra1706174

7. 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* (2021) 28: 3308–17. doi: 10.1158/1078-0432.Ccr-21-3044

8. Ignatiadis M, Sledge GW, Jeffrey SS. Liquid biopsy enters the clinic - implementation issues and future challenges. *Nat Rev Clin Oncol* (2021) 18:297–312. doi: 10.1038/s41571-020-00457-x

9. Murtaza M, Dawson S-J, Pogrebniak K, Rueda OM, Provenzano E, Grant J, et al. Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer. *Nat Commun* (2015) 6:8760. doi: 10.1038/ncomms9760

10. Hindson BJ, Ness KD, Masquelier DA, Belgrader P, Heredia NJ, Makarewicz AJ, et al. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal Chem* (2011) 83:8604–10. doi: 10.1021/ac202028g

11. Vogelstein B, Kinzler KW. Digital PCR. *Proc Natl Acad Sci U.S.A.* (1999) 96:9236–41. doi: 10.1073/pnas.96.16.9236

12. Huggett JF, Whale A. Digital PCR as a novel technology and its potential implications for molecular diagnostics. *Clin Chem* (2013) 59:1691–3. doi: 10.1373/clinchem.2013.214742

13. Zonta E, Garlan F, Pécuchet N, Perez-Toralla K, Caen O, Milbury C, et al. Multiplex detection of rare mutations by picoliter droplet based digital PCR: Sensitivity and specificity considerations. *PloS One* (2016) 11:e0159094. doi: 10.1371/journal.pone.0159094

14. Perkins G, Lu H, Garlan F, Taly V. Droplet-based digital PCR: Application in cancer research. *Adv Clin Chem* (2017) 79:43–91. doi: 10.1016/bs.acc.2016.10.001

15. Hindson CM, Chevillet JR, Briggs HA, Gallichotte EN, Ruf IK, Hindson BJ, et al. Absolute quantification by droplet digital PCR versus analog real-time PCR. *Nat Methods* (2013) 10:1003–5. doi: 10.1038/nmeth.2633

16. Dressman D, Yan H, Traverso G, Kinzler KW, Vogelstein B. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc Natl Acad Sci U.S.A.* (2003) 100:8817–22. doi: 10.1073/pnas.1133470100

17. Diehl F, Li M, He Y, Kinzler KW, Vogelstein B, Dressman D. BEAMing: single-molecule PCR on microparticles in water-in-oil emulsions. *Nat Methods* (2006) 3:551–9. doi: 10.1038/nmeth898

18. Garrido P, Paz-Ares L, Majem M, Morán T, Trigo JM, Bosch-Barrera J, et al. LungBEAM: A prospective multicenter study to monitor stage IV NSCLC patients with EGFR mutations using BEAMing technology. *Cancer Med* (2021) 10:5878–88. doi: 10.1002/cam4.4135

19. O'Leary B, Hrebien S, Beaney M, Fribbens C, Garcia-Murillas I, Jiang J, et al. Comparison of BEAMing and droplet digital PCR for circulating tumor DNA analysis. *Clin Chem* (2019) 65:1405–13. doi: 10.1373/clinchem.2019.305805

20. Wan JCM, Massie C, Garcia-Corbacho J, Mouliere F, Brenton JD, Caldas C, et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat Rev Cancer* (2017) 17:223–38. doi: 10.1038/nrc.2017.7

21. Yohe S, Thyagarajan B. Review of clinical next-generation sequencing. *Arch Pathol Lab Med* (2017) 141:1544–57. doi: 10.5858/arpa.2016-0501-RA

22. D'Haene N, Le Mercier M, De Nève N, Blanchard O, Delaunoy M, El Housni H, et al. Clinical validation of targeted next generation sequencing for colon and lung cancers. *PloS One* (2015) 10:e0138245. doi: 10.1371/journal.pone.0138245

23. Larribère L, Martens UM. Advantages and challenges of using ctDNA NGS to assess the presence of minimal residual disease (MRD) in solid tumors. *Cancers (Basel)* (2021) 13:5698. doi: 10.3390/cancers13225698

24. Forshew T, Murtaza M, Parkinson C, Gale D, Tsui DW, Kaper F, et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci Transl Med* (2012) 4:136ra168. doi: 10.1126/scitranslmed.3003726

25. Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B. Detection and quantification of rare mutations with massively parallel sequencing. *Proc Natl Acad Sci U.S.A.* (2011) 108:9530–5. doi: 10.1073/pnas.1105422108

26. Newman AM, Bratman SV, To J, Wynne JF, Eclov NC, Modlin LA, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat Med* (2014) 20:548–54. doi: 10.1038/nm.3519

27. Newman AM, Lovejoy AF, Klass DM, Kurtz DM, Chabon JJ, Scherer F, et al. Integrated digital error suppression for improved detection of circulating tumor DNA. *Nat Biotechnol* (2016) 34:547–55. doi: 10.1038/nbt.3520

28. Kurtz DM, Soo J, Co Ting Keh L, Alig S, Chabon JJ, Sworder BJ, et al. Enhanced detection of minimal residual disease by targeted sequencing of phased variants in circulating tumor DNA. *Nat Biotechnol* (2021) 39:1537–47. doi: 10.1038/s41587-021-00981-w

29. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med* (2017) 23:703–13. doi: 10.1038/nm.4333

30. Buchhalter I, Rempel E, Endris V, Allgäuer M, Neumann O, Volckmar A-L, et al. Size matters: Dissecting key parameters for panel-based tumor mutational burden analysis. *Int J Cancer* (2019) 144:848–58. doi: 10.1002/ijc.31878

31. Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WE, et al. The IASLC lung cancer staging project: Proposals for revision of the TNM stage groupings in the forthcoming (Eighth) edition of the TNM classification for lung cancer. *J Thorac Oncol* (2016) 11:39–51. doi: 10.1016/j.jtho.2015.09.009

32. Wong CC, Baum J, Silvestro A, Beste MT, Bharani-Dharan B, Xu S, et al. Inhibition of IL1β by canakinumab may be effective against diverse molecular subtypes of lung cancer: An exploratory analysis of the CANTOS trial. *Cancer Res* (2020) 80:5597–605. doi: 10.1158/0008-5472.Can-19-3176

33. Chabon JJ, Hamilton EG, Kurtz DM, Esfahani MS, Moding EJ, Stehr H, et al. Integrating genomic features for non-invasive early lung cancer detection. *Nature* (2020) 580:245–51. doi: 10.1038/s41586-020-2140-0

34. Chen K, Zhang J, Guan T, Yang F, Lou F, Chen W, et al. Comparison of plasma to tissue DNA mutations in surgical patients with non-small cell lung cancer. *J Thorac Cardiovasc Surg* (2017) 154:1123–1131.e1122. doi: 10.1016/j.jtcvs.2017.04.073

35. Liang W, Zhao Y, Huang W, Gao Y, Xu W, Tao J, et al. Non-invasive diagnosis of early-stage lung cancer using high-throughput targeted DNA methylation sequencing of circulating tumor DNA (ctDNA). *Theranostics* (2019) 9:2056–70. doi: 10.7150/thno.28119

36. Mack PC, Banks KC, Espenschied CR, Burich RA, Zill OA, Lee CE, et al. Spectrum of driver mutations and clinical impact of circulating tumor DNA analysis in non-small cell lung cancer: Analysis of over 8000 cases. *Cancer* (2020) 126:3219–28. doi: 10.1002/cncr.32876

37. Aggarwal C, Thompson JC, Black TA, Katz SI, Fan R, Yee SS, et al. Clinical implications of plasma-based genotyping with the delivery of personalized therapy in metastatic non-small cell lung cancer. *JAMA Oncol* (2019) 5:173–80. doi: 10.1001/jamaoncol.2018.4305

38. Leighl NB, Page RD, Raymond VM, Daniel DB, Divers SG, Reckamp KL, et al. Clinical utility of comprehensive cell-free DNA analysis to identify genomic biomarkers in patients with newly diagnosed metastatic non-small cell lung cancer. *Clin Cancer Res* (2019) 25:4691–700. doi: 10.1158/1078-0432.Ccr-19-0624

39. Kimura H, Suminoe M, Kasahara K, Sone T, Araya T, Tamori S, et al. Evaluation of epidermal growth factor receptor mutation status in serum DNA as a predictor of response to gefitinib (IRESSA). *Br J Cancer* (2007) 97:778–84. doi: 10.1038/sj.bjc.6603949

40. Douillard JY, Ostoros G, Cobo M, Ciuleanu T, McCormack R, Webster A, et al. First-line gefitinib in Caucasian EGFR mutation-positive NSCLC patients: a phase-IV, open-label, single-arm study. *Br J Cancer* (2014) 110:55–62. doi: 10.1038/bjc.2013.721

41. Jänne PA, Borras AM, Kuang Y, Rogers AM, Joshi VA, Liyanage H, et al. A rapid and sensitive enzymatic method for epidermal growth factor receptor mutation screening. *Clin Cancer Res* (2006) 12:751–8. doi: 10.1158/1078-0432.Ccr-05-2047

42. Kuang Y, Rogers A, Yeap BY, Wang L, Makrigiorgos M, Vetrand K, et al. Noninvasive detection of EGFR T790M in gefitinib or erlotinib resistant non-small cell lung cancer. *Clin Cancer Res* (2009) 15:2630–6. doi: 10.1158/1078-0432.Ccr-08-2592

43. Mok T, Wu YL, Lee JS, Yu CJ, Sriuranpong V, Sandoval-Tan J, et al. Detection and dynamic changes of EGFR mutations from circulating tumor DNA as a predictor of survival outcomes in NSCLC patients treated with first-line intercalated erlotinib and chemotherapy. *Clin Cancer Res* (2015) 21:3196–203. doi: 10.1158/1078-0432.Ccr-14-2594

44. Wu YL, Sequist LV, Hu CP, Feng J, Lu S, Huang Y, et al. EGFR mutation detection in circulating cell-free DNA of lung adenocarcinoma patients: analysis of LUX-lung 3 and 6. *Br J Cancer* (2017) 116:175–85. doi: 10.1038/bjc.2016.420

45. Wang Z, Cheng Y, An T, Gao H, Wang K, Zhou Q, et al. Detection of EGFR mutations in plasma circulating tumour DNA as a selection criterion for first-line gefitinib treatment in patients with advanced lung adenocarcinoma (BENEFIT): a phase 2, single-arm, multicentre clinical trial. *Lancet Respir Med* (2018) 6:681–90. doi: 10.1016/s2213-2600(18)30264-9

46. Han JY, Choi JJ, Kim JY, Han YL, Lee GK. PNA clamping-assisted fluorescence melting curve analysis for detecting EGFR and KRAS mutations in the circulating tumor DNA of patients with advanced non-small cell lung cancer. *BMC Cancer* (2016) 16:627. doi: 10.1186/s12885-016-2678-2

47. Krug AK, Enderle D, Karlovich C, Priewasser T, Bentink S, Spiel A, et al. Improved EGFR mutation detection using combined exosomal RNA and circulating tumor DNA in NSCLC patient plasma. *Ann Oncol* (2018) 29:700–6. doi: 10.1093/annonc/mdx765

48. Park CK, Lee SY, Lee JC, Choi CM, Lee SY, Jang TW, et al. Phase II open-label multicenter study to assess the antitumor activity of afatinib in lung cancer patients with activating epidermal growth factor receptor mutation from circulating tumor DNA: Liquid-Lung-A. *Thorac Cancer* (2021) 12:444–52. doi: 10.1111/1759-7714.13763

49. Mondaca S, Lebow ES, Namakydoust A, Razavi P, Reis-Filho JS, Shen R, et al. Clinical utility of next-generation sequencing-based ctDNA testing for common and novel ALK fusions. *Lung Cancer* (2021) 159:66–73. doi: 10.1016/j.lungcan.2021.06.018

50. Wang Y, Tian PW, Wang WY, Wang K, Zhang Z, Chen BJ, et al. Noninvasive genotyping and monitoring of anaplastic lymphoma kinase (ALK) rearranged non-small cell lung cancer by capture-based next-generation sequencing. *Oncotarget* (2016) 7:65208–17. doi: 10.18632/oncotarget.11569

51. Si H, Kuziora M, Quinn KJ, Helman E, Ye J, Liu F, et al. A blood-based assay for assessment of tumor mutational burden in first-line metastatic NSCLC treatment: Results from the MYSTIC study. *Clin Cancer Res* (2021) 27:1631–40. doi: 10.1158/1078-0432.Ccr-20-3771

52. Kwapisz D. The first liquid biopsy test approved. is it a new era of mutation testing for non-small cell lung cancer? *Ann Transl Med* (2017) 5:46. doi: 10.21037/atm.2017.01.32

53. Horn L, Whisenant JG, Wakelee H, Reckamp KL, Qiao H, Leal TA, et al. Monitoring therapeutic response and resistance: Analysis of circulating tumor DNA in patients with ALK+ lung cancer. *J Thorac Oncol* (2019) 14:1901–11. doi: 10.1016/j.jtho.2019.08.003

54. Dagogo-Jack I, Rooney M, Nagy RJ, Lin JJ, Chin E, Ferris LA, et al. Molecular analysis of plasma from patients with ROS1-positive NSCLC. *J Thorac Oncol* (2019) 14:816–24. doi: 10.1016/j.jtho.2019.01.009

55. Vansteenkiste JF, Canon JL, De Braud F, Grossi F, De Pas T, Gray JE, et al. Safety and efficacy of buparlisib (BKM120) in patients with PI3K pathway-activated non-small cell lung cancer: Results from the phase II BASALT-1 study. *J Thorac Oncol* (2015) 10:1319–27. doi: 10.1097/jto.0000000000000607

56. Goto K, Ichinose Y, Ohe Y, Yamamoto N, Negoro S, Nishio K, et al. Epidermal growth factor receptor mutation status in circulating free DNA in serum: from IPASS, a phase III study of gefitinib or carboplatin/paclitaxel in non-small cell lung cancer. *J Thorac Oncol* (2012) 7:115–21. doi: 10.1097/JTO.0b013e3182307f98

57. Chin R-I, Chen K, Usmani A, Chua C, Harris PK, Binkley MS, et al. Detection of solid tumor molecular residual disease (MRD) using circulating tumor DNA (ctDNA). *Mol Diagn Ther* (2019) 23:311–31. doi: 10.1007/s40291-019-00390-5

58. Chen K, Zhao H, Shi Y, Yang F, Wang LT, Kang G, et al. Perioperative dynamic changes in circulating tumor DNA in patients with lung cancer (DYNAMIC). *Clin Cancer Res* (2019) 25:7058–67. doi: 10.1158/1078-0432.Ccr-19-1213

59. Chaudhuri AA, Chabon JJ, Lovejoy AF, Newman AM, Stehr H, Azad TD, et al. Early detection of molecular residual disease in localized lung cancer by circulating tumor DNA profiling. *Cancer Discov* (2017) 7:1394–403. doi: 10.1158/2159-8290.Cd-17-0716

60. Zhang JT, Liu SY, Gao W, Liu SM, Yan HH, Ji L, et al. Longitudinal undetectable molecular residual disease defines potentially cured population in localized non-small cell lung cancer. *Cancer Discov* (2022) 12:1690–701. doi: 10.1158/2159-8290.Cd-21-1486

61. Balak MN, Gong Y, Riely GJ, Somwar R, Li AR, Zakowski MF, et al. Novel D761Y and common secondary T790M mutations in epidermal growth factor receptor-mutant lung adenocarcinomas with acquired resistance to kinase inhibitors. *Clin Cancer Res* (2006) 12:6494–501. doi: 10.1158/1078-0432.Ccr-06-1570

62. Remon J, Caramella C, Jovelet C, Lacroix L, Lawson A, Smalley S, et al. Osimertinib benefit in EGFR-mutant NSCLC patients with T790M-mutation detected by circulating tumour DNA. *Ann Oncol* (2017) 28:784–90. doi: 10.1093/annonc/mdx017

63. Yang JC, Ahn MJ, Kim DW, Ramalingam SS, Sequist LV, Su WC, et al. Osimertinib in pretreated T790M-positive advanced non-Small-Cell lung cancer: AURA study phase II extension component. *J Clin Oncol* (2017) 35:1288–96. doi: 10.1200/jco.2016.70.3223

64. Park CK, Cho HJ, Choi YD, Oh IJ, Kim YCA. Phase II trial of osimertinib in the second-line treatment of non-small cell lung cancer with the EGFR T790M mutation, detected from circulating tumor DNA: LiquidLung-O-Cohort 2. *Cancer Res Treat* (2019) 51:777–87. doi: 10.4143/crt.2018.387

65. Sorensen BS, Wu L, Wei W, Tsai J, Weber B, Nexo E, et al. Monitoring of epidermal growth factor receptor tyrosine kinase inhibitor-sensitizing and resistance mutations in the plasma DNA of patients with advanced non-small cell lung cancer during treatment with erlotinib. *Cancer* (2014) 120:3896–901. doi: 10.1002/cncr.28964

66. Ramalingam SS, Vansteenkiste J, Planchard D, Cho BC, Gray JE, Ohe Y, et al. Overall survival with osimertinib in untreated, EGFR-mutated advanced NSCLC. *N Engl J Med* (2020) 382:41–50. doi: 10.1056/NEJMoa1913662

67. Yu HA, Paz-Ares LG, Yang JC, Lee KH, Garrido P, Park K, et al. Phase I study of the efficacy and safety of ramucirumab in combination with osimertinib in advanced T790M-positive EGFR-mutant non-small cell lung cancer. *Clin Cancer Res* (2021) 27:992–1002. doi: 10.1158/1078-0432.Ccr-20-1690

68. Thress KS, Paweletz CP, Felip E, Cho BC, Stetson D, Dougherty B, et al. Acquired EGFR C797S mutation mediates resistance to AZD9291 in non-small cell lung cancer harboring EGFR T790M. *Nat Med* (2015) 21:560–2. doi: 10.1038/nm.3854

69. Guibert N, Hu Y, Feeney N, Kuang Y, Plagnol V, Jones G, et al. Amplicon-based next-generation sequencing of plasma cell-free DNA for detection of driver and resistance mutations in advanced non-small cell lung cancer. *Ann Oncol* (2018) 29:1049–55. doi: 10.1093/annonc/mdy005

70. Ramalingam SS, Yang JC, Lee CK, Kurata T, Kim DW, John T, et al. Osimertinib as first-line treatment of EGFR mutation-positive advanced non-Small-Cell lung cancer. *J Clin Oncol* (2018) 36:841–9. doi: 10.1200/jco.2017.74.7576

71. Sakai K, Takahama T, Shimokawa M, Azuma K, Takeda M, Kato T, et al. Predicting osimertinib-treatment outcomes through EGFR mutant-fraction monitoring in the circulating tumor DNA of EGFR T790M-positive patients with non-small cell lung cancer (WJOG8815L). *Mol Oncol* (2021) 15:126–37. doi: 10.1002/1878-0261.12841

72. Dagogo-Jack I, Rooney M, Lin JJ, Nagy RJ, Yeap BY, Hubbeling H, et al. Treatment with next-generation ALK inhibitors fuels plasma ALK mutation diversity. *Clin Cancer Res* (2019) 25:6662–70. doi: 10.1158/1078-0432.Ccr-19-1436

73. Sánchez-Herrero E, Serna-Blasco R, Ivanchuk V, García-Campelo R, Dómine Gómez M, Sánchez JM, et al. NGS-based liquid biopsy profiling identifies mechanisms of resistance to ALK inhibitors: a step toward personalized NSCLC treatment. *Mol Oncol* (2021) 15:2363–76. doi: 10.1002/1878-0261.13033

74. Lin YT, Chiang CL, Hung JY, Lee MH, Su WC, Wu SY, et al. Resistance profiles of anaplastic lymphoma kinase tyrosine kinase inhibitors in advanced non-small-cell lung cancer: a multicenter study using targeted next-generation sequencing. *Eur J Cancer* (2021) 156:1–11. doi: 10.1016/j.ejca.2021.06.043

75. Kok PS, Lee K, Lord S, John T, Marschner I, Wu YL, et al. Incorporating circulating tumor DNA detection to radiographic assessment for treatment response in advanced EGFR-mutant lung cancer. *Lung Cancer* (2022) 163:14–8. doi: 10.1016/j.lungcan.2021.11.010

76. Yu HA, Schoenfeld AJ, Makhnin A, Kim R, Rizvi H, Tsui D, et al. Effect of osimertinib and bevacizumab on progression-free survival for patients with metastatic EGFR-mutant lung cancers: A phase 1/2 single-group open-label trial. *JAMA Oncol* (2020) 6:1048–54. doi: 10.1001/jamaoncol.2020.1260

77. Raja R, Kuziora M, Brohawn PZ, Higgs BW, Gupta A, Dennis PA, et al. Early reduction in ctDNA predicts survival in patients with lung and bladder cancer treated with durvalumab. *Clin Cancer Res* (2018) 24:6212–22. doi: 10.1158/1078-0432.Ccr-18-0386

78. Lee YJ, Yoon KA, Han JY, Kim HT, Yun T, Lee GK, et al. Circulating cell-free DNA in plasma of never smokers with advanced lung adenocarcinoma receiving gefitinib or standard chemotherapy as first-line therapy. *Clin Cancer Res* (2011) 17:5179–87. doi: 10.1158/1078-0432.Ccr-11-0400

79. Bordi P, Del Re M, Minari R, Rofi E, Buti S, Restante G, et al. From the beginning to resistance: Study of plasma monitoring and resistance mechanisms in a cohort of patients treated with osimertinib for advanced T790M-positive NSCLC. *Lung Cancer* (2019) 131:78–85. doi: 10.1016/j.lungcan.2019.03.017

80. Papadimitrakopoulou VA, Han JY, Ahn MJ, Ramalingam SS, Delmonte A, Hsia TC, et al. Epidermal growth factor receptor mutation analysis in tissue and plasma from the AURA3 trial: Osimertinib versus platinum-pemetrexed for T790M mutation-positive advanced non-small cell lung cancer. *Cancer* (2020) 126:373–80. doi: 10.1002/cncr.32503

81. Fukuhara T, Saito H, Furuya N, Watanabe K, Sugawara S, Iwasawa S, et al. Evaluation of plasma EGFR mutation as an early predictor of response of erlotinib plus bevacizumab treatment in the NEJ026 study. *EBioMedicine* (2020) 57:102861. doi: 10.1016/j.ebiom.2020.102861

82. Zhang S, Zhu L, Xia B, Chen E, Zhao Q, Zhang X, et al. Epidermal growth factor receptor (EGFR) T790M mutation identified in plasma indicates failure sites and predicts clinical prognosis in non-small cell lung cancer progression during first-generation tyrosine kinase inhibitor therapy: a prospective observational study. *Cancer Commun (Lond)* (2018) 38:28. doi: 10.1186/s40880-018-0303-2

83. Camidge DR, Kim HR, Ahn MJ, Yang JCH, Han JY, Hochmair MJ, et al. Brigatinib versus crizotinib in ALK inhibitor-naive advanced ALK-positive NSCLC: Final results of phase 3 ALTA-1L trial. *J Thorac Oncol* (2021) 16:2091–108. doi: 10.1016/j.jtho.2021.07.035

84. Horn L, Spigel DR, Vokes EE, Holgado E, Ready N, Steins M, et al. Nivolumab versus docetaxel in previously treated patients with advanced non-Small-Cell lung cancer: Two-year outcomes from two randomized, open-label, phase III trials (CheckMate 017 and CheckMate 057). *J Clin Oncol* (2017) 35:3924–33. doi: 10.1200/jco.2017.74.3062

85. Anagnostou V, Forde PM, White JR, Niknafs N, Hruban C, Naidoo J, et al. Dynamics of tumor and immune responses during immune checkpoint blockade in non-small cell lung cancer. *Cancer Res* (2019) 79:1214–25. doi: 10.1158/0008-5472.Can-18-1127

86. Goldberg SB, Narayan A, Kole AJ, Decker RH, Teysir J, Carriero NJ, et al. Early assessment of lung cancer immunotherapy response *via* circulating tumor DNA. *Clin Cancer Res* (2018) 24:1872–80. doi: 10.1158/1078-0432.Ccr-17-1341

87. Hellmann MD, Nabet BY, Rizvi H, Chaudhuri AA, Wells DK, Dunphy MPS, et al. Circulating tumor DNA analysis to assess risk of progression after long-term response to PD-(L)1 blockade in NSCLC. *Clin Cancer Res* (2020) 26:2849–58. doi: 10.1158/1078-0432.Ccr-19-3418

88. Nabet BY, Esfahani MS, Moding EJ, Hamilton EG, Chabon JJ, Rizvi H, et al. Noninvasive early identification of therapeutic benefit from immune checkpoint inhibition. *Cell* (2020) 183:363–376.e313. doi: 10.1016/j.cell.2020.09.001

89. Samstein RM, Lee CH, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet* (2019) 51:202–6. doi: 10.1038/s41588-018-0312-8

90. Chan TA, Yarchoan M, Jaffee E, Swanton C, Quezada SA, Stenzinger A, et al. Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. *Ann Oncol* (2019) 30:44–56. doi: 10.1093/annonc/mdy495

91. Dudley JC, Lin MT, Le DT, Eshleman JR. Microsatellite instability as a biomarker for PD-1 blockade. *Clin Cancer Res* (2016) 22:813–20. doi: 10.1158/1078-0432.Ccr-15-1678

92. Wang Z, Duan J, Cai S, Han M, Dong H, Zhao J, et al. Assessment of blood tumor mutational burden as a potential biomarker for immunotherapy in patients with non-small cell lung cancer with use of a next-generation sequencing cancer gene panel. *JAMA Oncol* (2019) 5:696–702. doi: 10.1001/jamaoncol.2018.7098

93. Gandara DR, Paul SM, Kowanetz M, Schleifman E, Zou W, Li Y, et al. Blood-based tumor mutational burden as a predictor of clinical benefit in non-small-cell lung cancer patients treated with atezolizumab. *Nat Med* (2018) 24:1441–8. doi: 10.1038/s41591-018-0134-3

94. Georgiadis A, Durham JN, Keefer LA, Bartlett BR, Zielonka M, Murphy D, et al. Noninvasive detection of microsatellite instability and high tumor mutation burden in cancer patients treated with PD-1 blockade. *Clin Cancer Res* (2019) 25:7024–34. doi: 10.1158/1078-0432.Ccr-19-1372

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 27 October 2022
DOI 10.3389/fonc.2022.991120

Check for updates

OPEN ACCESS

EDITED BY
Ping Zhan,
Nanjing University School of Medicine, China

REVIEWED BY
Andrea Moreno Manuel,
Fundación de Investigación del Hospital General Universitario de Valencia, Spain
Martyna Filipska,
Fundació Institut Mar d'Investigacions Mèdiques (IMIM), Spain

*CORRESPONDENCE
Jian Tajbakhsh
jian.tajbakhsh@cshs.org

†PRESENT ADDRESS
Fariborz Mortazavi,
Ventura County Hematology Oncology Specialists, Oxnard, CA, United States

SPECIALTY SECTION
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

RECEIVED 11 July 2022
ACCEPTED 17 October 2022
PUBLISHED 27 October 2022

CITATION
Tajbakhsh J, Mortazavi F and Gupta NK (2022) DNA methylation topology differentiates between normal and malignant in cell models, resected human tissues, and exfoliated sputum cells of lung epithelium. *Front. Oncol.* 12:991120. doi: 10.3389/fonc.2022.991120

COPYRIGHT
© 2022 Tajbakhsh, Mortazavi and Gupta. 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.

DNA methylation topology differentiates between normal and malignant in cell models, resected human tissues, and exfoliated sputum cells of lung epithelium

Jian Tajbakhsh[1,2,3]*, Fariborz Mortazavi[4,5]† and Nirdesh K. Gupta[3]

[1]Department of Surgery, Cedars-Sinai, Los Angeles, CA, United States, [2]Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai, Los Angeles, CA, United States, [3]3rd Street Diagnostics, Cedars-Sinai, Los Angeles, CA, United States, [4]Division of Hematology/Oncology, West Los Angeles Veterans Affairs (VA) Medical Center, Los Angeles, CA, United States, [5]Department of Medicine, University of California, Los Angeles, CA, United States

Background: Global DNA hypomethylation is a prominent feature of cancer cells including lung cancer, that has not been widely explored towards cancer diagnosis. In this study we assess the comparative distribution of global DNA methylation in normal cells versus cancer cells in various specimen models.

Methods: We used *in situ* immunofluorescence labeling of overall 5-methylcytosine (5mC) and covisualization of global DNA (gDNA) by 4',6-diamidino-2-phenylindole (DAPI), confocal microscopy and 3D image analysis to derive 5mC/DAPI colocalization patterns in human cell lines (BEAS-2B, A549, H157) and upper respiratory epithelial cells derived from various sources (i.e., sputum from healthy and cancer patients, and resected tissues from normal parenchyma and lung tumors).

Results: By introducing 5mC/DAPI colocalization index as a metric we could distinguish between normal epithelial cells and aberrantly hypomethylated cancer cells. Cultured lung cancer cells (H157 and A549) had significantly lower indices compared to normal cells (BEAS-2B). Furthermore, we were able to identify such extensively hypomethylated low-index cells in tumor tissues and the matching sputum from cancer patients. In contrast, the indices of cells derived from sputum of healthy individuals had more similarity to epithelial cells of normal parenchyma and the phenotypically normal BEAS-2B cells.

Conclusions: The results suggest that 5mC topology using high-resolution image cytometry shows potential for identifying hypomethylated cancerous cells in human tissues and amongst normal cells in matching sputum, which may render a valuable surrogate for biopsied tissues. This promising feature deserves further validation in more comprehensive studies.

KEYWORDS

global DNA methylation, 5-methylcytosine, sputum, lung cancer, confocal microscopy, 3D image cytometry, single-cell analysis, nuclear topology

Introduction

In cancer cells two types of aberrant DNA methylation features coexist: 1) promoter hypermethylation of a few genes and 2) global hypomethylation, mostly attributed to severe hypomethylation of repeat sequences that comprise more than two-thirds of the human genome (1–3). The most prevalent types of repetitive sequences include long interspersed nucleotide element 1 (LINE-1) and *Alu* sequences that contribute to around 30% of the genome. The analysis of DNA hypomethylation has been largely remained unexploited although it has been known for decades that global hypomethylation is prevalent in tumors compared to normal cells (4, 5). Generally, cancer cells show a net hypomethylation, containing 20–60% less genomic 5mC than their normal counterparts. Global DNA hypomethylation occurs in many of the major cancer types, including the tumors of the breast, colon, head and neck, bladder, esophagus, liver, prostate, stomach, and lung (6, 7). Thus, the global methylation status is a unique feature of cells and tissues and global hypomethylation is a common epigenetic process in cancer, which may progressively evolve during multistage carcinogenesis.

Because of its high frequency in the genome, methylation in LINE-1 has shown to be a good indicator of the global DNA methylation level within a cell (7–9). LINE-1 is heavily methylated in normal human tissues. Hypomethylation of LINE-1 repetitive elements has been described as one of the key hallmarks of tumorigenesis. This shift was also shown in lung tumor cells (7, 10–12) and in blood cells of lung cancer patients (13, 14). Moreover, the degree of LINE-1 hypomethylation is associated with clinical data and survival prognosis (8, 15). The selection of literature amongst numerous other publications are a proof of evidence that global hypomethylation largely exists in both subtypes of non-small cell lung cancer (NSCLC), squamous cell carcinoma (SCC) as well as in adenocarcinoma. Chalitchagorn et al. (7) evaluated the differential level of LINE-1 methylation between normal tissues and cancers in a broad panel of malignancies including NSCLC (7). The investigators detected significantly greater hypomethylation in most cancers including carcinomas of the lung. Daskalos et al. (2009) reported that LINE-1 and Alu methylation indices in primary tumors strongly correlated with each other (12). However, clinicopathological parameters such as age, gender, T status (size and extension of the tumor), differentiation and nodal metastasis did not correlate with LINE-1 and *Alu* methylation. Notably LINE-1 hypomethylation was found more frequent in SSC than in adenocarcinoma, however only at borderline significance ($p = 0.052$). Suzuki et al. (2013) disclosed that through accurate measurement of methylation levels using pyrosequencing, hypomethylation of LINE-1 was frequently detected in NSCLC and associated with various clinical features (11). Tumor tissues showed significantly lower levels of LINE-1 methylation when compared with matched nonmalignant lung tissues. A Study by Ikeda et al. (2013), also using pyrosequencing, revealed that methylation levels of resected lung cancer tissue were significantly lower than that of matched normal lung tissues (8). The association between clinical characteristics and methylation levels of lung adenocarcinoma tissues has revealed that higher histologic grade and positive findings for vascular invasion were significantly associated with stronger hypomethylation. Furthermore, previous studies brought to light that hypomethylation is related to worse prognosis of NSCLC, that is significantly shorter disease-free intervals after curative resection. Along the same lines, the methylation rate by LINE-1 contribution was significantly lower in squamous cell carcinoma than in adenocarcinoma (11, 15).

To date, differential DNA methylation analysis has been quantitatively assessed mostly by molecular approaches including electrophoretic, chromatographic, polymerase chain reaction (PCR) based, array based, and sequencing technologies (16, 17). Despite tremendous improvement in specificity,

sensitivity, and the inherent single-base resolution of these methods, they remain technically and economically challenging in the high-throughput analysis of single cells (18). These include the limitation of PCR-based approaches in multiplexing and the challenging sensitivity and cost issues of whole-genome sequencing, especially for the interrogation of repetitive elements. Considering the prevalence and load of DNA methylation imbalances —especially hypomethylation of repeat sequences— imaging-based assessment of changes in global nuclear 5mC patterns provides a powerful alternative for the massively parallel analysis of cells. The reason being that DNA hypomethylation at such scales involves large-scale chromatin reorganization that can be made visible by light microscopy (19–21). Beyond *in situ* methods, the dynamics of global DNA methylation has been successfully visualized by live-cell imaging using constructed reporters (22–24).

We had previously introduced an image-cytometric approach termed three-dimensional quantitative DNA methylation imaging (3D-qDMI), a nondestructive *in situ* method that entails the parallel quantitative measurement of 5mC load and spatial nuclear distribution to be used for the characterization of cells and tissues (25, 26). 3D-qDMI combines immunofluorescence, high-resolution confocal microscopy and 3D image analysis, and allows for the rapid, parallel, single-cell phenotyping of thousands of cells within heterogeneous samples. 3D-qDMI has been successfully applied in the characterization of mouse and human cells and tissues of various origin (27–33). Utilizing this high-content tool, in this study we explored the comparative 5mC topology in normal epithelial and cancer cell models of the lung, as well as in cells of resected human tissues and exfoliated upper respiratory epithelial cells derived from matching sputum samples from cancer patients and healthy donors.

Materials and methods

Cultured cells

A549 and H157 cells (ATCC) were routinely cultured in RPMI supplemented with antibiotics and 10% heat-inactivated fetal bovine serum (FBS) (Omega Scientific). BEAS-2B cells (ATCC) were cultured in bronchial epithelial cell basal medium (BEBM) supplemented with all 1x BEGM (bronchial epithelial growth medium) SingleQuots kit additives (all from Lonza): 2ml of bovine pituitary extract (BPE), 0.5 ml of hydrocortisone, 0.5 ml of human epidermal growth factor (hEGF), 0.5 ml of epinephrine, 0.5 ml of transferrin, 0.5 ml of insulin, 0.5 ml of retinoic acid, triiodothyronine, and 0.5 ml gentamicin/gentamicin-B. According to the manufacturer's protocol to formulate 500 ml of BEGM, the entire volume of each additive in the kit was added to 500 ml of BEBM. All cells were grown to 70% confluency in 5% CO_2 at 37°C.

Patient materials

The study utilized pre-surgical sputum samples and post-surgical tissue from three patients with stage I-II NSCLC. Sputum samples were collected following obtaining the patients' written consents. Surgically resected specimens were provided in a deidentified manner and were exempt from patients' consents. Since this study was categorized as basic research, information that could lead to patient identity and patient characteristics such as age, gender, and clinical information were masked, and each sample received a research code.

Sputum collection and processing

Patients were given a cup of water to gurgle, immediately before sputum induction, to minimize the contribution of oral cavity saliva in the sample. The actual sputum induction was performed by inhalation of hypertonic saline (NaCl 4.5%). Aerosols were generated by an ultrasonic nebulizer, with an output at 1.5 mL/min. The subjects inhaled the saline solution aerosols for a fixed period of 15 min. Then subjects were encouraged to expectorate sputum. Samples were collected in a plastic container and kept at 2–8°C until processing for cell extraction. For purification of cellular components from mucus and other contaminants, samples were processed twice as follows. Samples were diluted with phosphate-buffered saline solution (PBS) containing 10 mM dithiothreitol (DTT), then centrifuged at 400g for 10 min at 4°C to separate cellular and fluid phases. The ultimate cell pellet was resuspended in phosphate-buffered saline (PBS) containing 20 mM ethylenediamine tetra-acetic acid (EDTA) and 2% human serum albumin. Cell counts were performed on centrifuged samples and a sample of about $5x10^5$ cells was spread on a microscope slide. Subsequently, cells were fixed in 4% paraformaldehyde for 15 minutes. Then the slide was kept in PBS at 2–8°C until further processing for immunofluorescence.

Immunofluorescence assay

Immunofluorescence staining was performed in four-chamber microplates (ThermoFisher Scientific) according to previously established protocols (25, 28, 33). The primary and secondary antibody set included unconjugated mouse anti-5-methylcytosine monoclonal antibody (AMM99021, Aviva Systems Biology) at 1 mg/ml and Alexa488-linked donkey anti-mouse IgG (A-21202, ThermoFisher Scientific) at 5 mg/ml final concentrations. The cells were subsequently delineated with the cytoplasmic marker Cell Mask Red (ThermoFischer Scientific) and cell nuclei counterstained with DAPI. The

specificity/dynamic range of the anti-5mC antibody was tested as previously reported in (29) (data not shown in here). Formalin-fixed tissue sections at 5 μm thickness were kept floating in 10% formalin at 2–8°C until immunofluorescence staining. Prior to staining, tissues were transferred to microwell plates, washed in PBS at room temperature, then stained as floating tissues using the same protocol that was applied to fixed cells.

Confocal imaging and 3D image analysis

Confocal imaging of labeled slides was performed using a TCS SP5 X Supercontinuum microscope (Leica Microsystems, Mannheim Germany). The system provides full freedom and flexibility in excitation and emission within the continuous range of 470 to 670 nm within 1 nm increment. The TCS SP5 X system was coupled with a 405nm diode laser line for excitation of DAPI fluorescence. Serial optical sections were collected at increments of 250–500 nm with a Plan-Apo 63×1.3 glycerol immersion lens. The pinhole size was consistently 1.0 airy unit. The output file format was a series of TIFF images that were utilized for 3D-image analysis. The typical image size was 1576×1576 with a respective voxel size of 189 nm ×189 nm × 500 nm (x, y, and z axes), and a dynamic intensity range of 12 bits per pixel in all four channels. All biomarker signals from optical sections were recorded into separate channels. All images were acquired under nearly identical conditions and modality settings. The drift of the settings during acquisition was considered minimal and therefore neglected. 3D image analysis was performed using a dedicated algorithm for multi-parametric high-content analysis, as previously described in (25, 26, 28). This image analysis tool operates in three steps. First cells (within imaged populations) were processed for 3D segmentation. Then fluorescence 5mC and DAPI signals were recorded for each cell nucleus, and subsequently two parameters were generated per cell: (a) the global 5mC intensity of the entire nucleus, and (b) the codistribution (2D scatter plots) of 5mC signals and gDNA (DAPI) signals. The angle under the regression line of the 5mC/DAPI codistribition plot is automatically calculated as the second parameter. The two parameters were output as DAT files. The results can be traced back for each imaged epithelial cell through a numerical identifier that was generated by the software. In addition, the software also calculated the similarity of 5mC/DAPI codistribution patterns between cells based on Kullback-Leibler (K-L) divergence and generated similarity maps of cells as previously described (25, 26). The pseudo-colors within a similarity map represent the four K-L characteristics known as soft-qualifiers: green (similar), blue (likely similar), yellow (unlikely similar), and red (dissimilar). The soft-qualifiers for each cell were generated through comparison of the 5mC/DAPI distribution pattern of the individual cell with the composite 5mC/DAPI pattern of the entire imaged cell population.

Results

Nuclear DNA methylation patterns distinguish between normal and transformed cells

The notion of our study was to explore differential three-dimensional (3D) nuclear 5mC patterns in normal versus cancerous human upper respiratory cells. To establish these differences, we started with three cultured human cell lines comprising the immortalized normal human epithelial cell line BEAS-2B, and the two NSCLC cell lines A549 (alveolar basal epithelial cells) and H157 (highly invasive lung carcinoma cells). BEAS-2B are epithelial cells that were established through isolation from normal human bronchial epithelium, obtained from autopsy of a noncancerous individual (34). These cells have been widely used as an *in vitro* cell model representing normal lung epithelial cells in a large variety of studies associated with respiratory diseases including lung carcinogenesis. A549 cells are adenocarcinoma human alveolar basal epithelial cells established through an explant culture of adenocarcinoma lung tissue of a 58-year-old Caucasian male (35). This cell line is categorized as NSCLC. On that note, NSCLC accounts for 85-88% of all cases of lung cancer. The A549 cell line is widely used as a model of lung adenocarcinoma, as well as an *in vitro* model for type II pulmonary epithelial cells (35). The more aggressive H157 cell line was established by A.F. Gazdar, H.K. Oie, J.D. Minna and associates in 1979 from a SCC of the buccal mucosa, from cells recovered from pleural effusion obtained from a patient prior to therapy.

Using K-L divergence, cell-similarity maps of 5mC/DAPI codistribution were generated as described in (25). Basically, each cell nucleus receives a numerical score which is converted into the four classifiers “similar”, “likely similar”, “unlikely similar” and “dissimilar”, with respective color-codes as explained in the Methods section. For clarity, the cell-similarity maps only contain cell nuclei extracted from fluorescence images (by our image analysis software). The more a cell population contains similar and likely similar cells the more homogenous it is regarding the cells’ 5mC/DAPI codistribution. We observed a high degree of homogeneity in 5mC patterns among cells for all three cell lines. This observation agrees with the fact that cultured cells typically align with the more dominant cell features, as represented by similar and likely similar cells in Figure 1. Thus, we conclude that the observed patterns are in fact dominant patterns for each cell line. Figure 1 depicts 5mC/DAPI codistribution patterns for each of the three cell lines as scatter plots on the cell-population

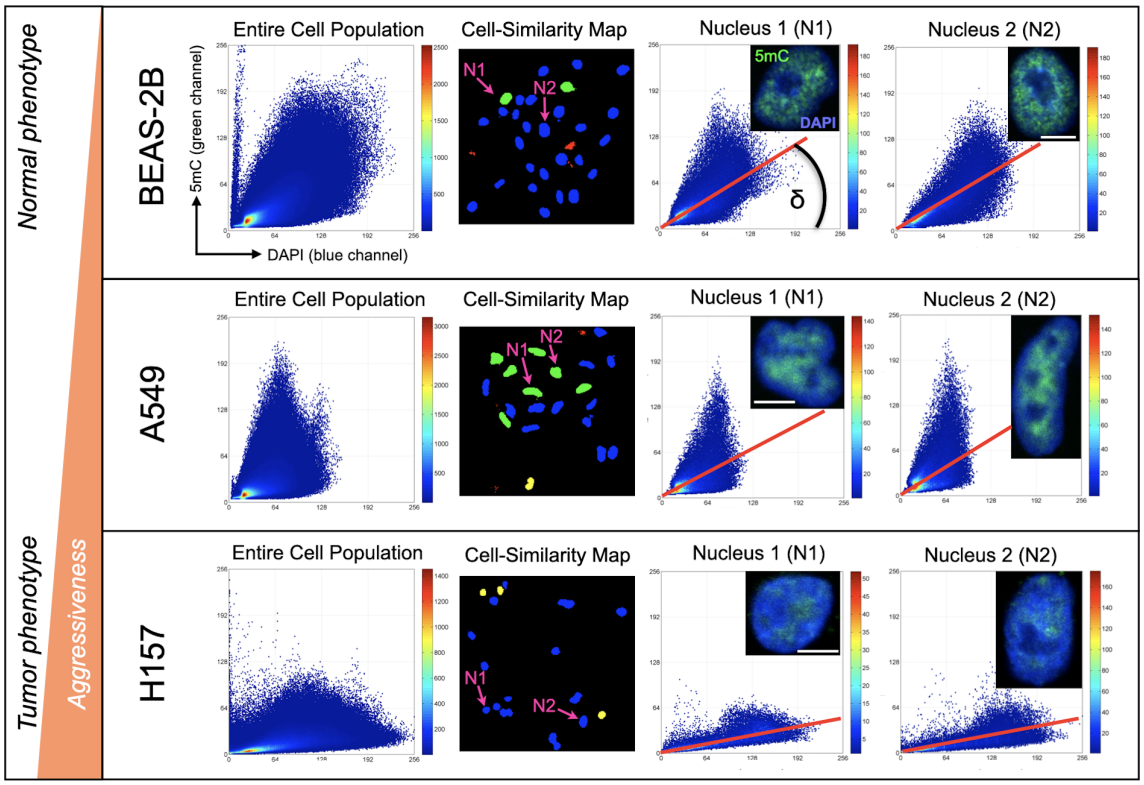


FIGURE 1
Global DNA methylation phenotyping of cultured cells with 3D-qDMI. The method was able to distinguish between the different cell types based on differential 5mC/DAPI distribution patterns calculated and displayed as individual heat map scatter plots (DAPI = x-axis, 5mC = y-axis). Plots were generated for the entire cell population (composite 5mC/DAPI codistribution of all measured cell nuclei) as the reference plot, and for each cell nucleus. Cell-similarity maps based on K-L divergence indicate a high degree of homogeneity in 5mC/DAPI topology for all three cell lines. This is due to most cells being similar (green nuclei) and likely similar (blue nuclei). Selected cell nuclei N1 and N2 for each cell line selected from the cell-similarity maps represent the most dominant 5mC topology for each cell line. The angle δ under the regression line (false-colored red) equals the 5mC/DAPI colocalization index for each nucleus. White bars are 5 μm.

level and for individual representative cell nuclei (N1 and N2). Because of the high degree in feature homogeneity within a cell line, N1 and N2 are very similar in 5mC topology across all cell lines. We observed common global DNA methylation patterns amongst healthy cells that significantly differ from the 5mC/DAPI patterns of cancerous cells. We introduced the angle δ under the regression line of the plot as a measurable descriptor of 5mC/DAPI codistribution for each cell type, calling it the 5mC/DAPI codistribution index or simply the 5mC/DAPI index. This index is differential between the cell lines and corresponds to the aggressiveness of the cells. NSCLC cell line A549 displays slightly reduced (~20% on average) and H157 significantly reduced (> than 60% on average) 5mC loads compared to BEAS-2B cells. The same proportional reduction could be experienced for 5mC/DAPI colocalization indices. H157 cells, which are reported to have more metastatic potential than A549 cells, are even more hypomethylated displaying a flatter curve and in conjunction a smaller 5mC/DAPI index. These features indicate that a large portion of the highly condensed repeat sequences in the nuclei (as represented by DAPI-staining) is hypomethylated. From these initial results we glean that the method was able to successfully distinguish between the different cell types, specifically between the more normal and transformed phenotypes and in correlation with aggressiveness, based on differential 5mC/DAPI codistribution patterns (scatter plots).

Normal and malignant tissue cells display differential 5mC phenotypes found in cultured cells

Next, we tried to verify the observed 5mC patterns in surgically removed tissues from lung cancer patients and

adjacent normal lung tissue, as well as in cells from matching sputum of the lung cancer patients versus sputum cells of the healthy donors (with no history of cancer). Figure 2 shows a similar differential 5mC load in the fluorescently labeled section of normal parenchyma and from surgically resected lung tumor: a substantial degree of hypomethylation was seen by eye under the microscope in the nuclei of epithelial cells residing within the tumor area compared to their normal counterparts. A confirmation of the differential 5mC phenotypes was obtained using 3D image analysis. The same comparative relation as for the cell lines could be found in tumor tissue from lung cancer patients and in adjacent normal lung tissue: the normal lung section was populated by an absolute majority of epithelial cells with normal methylation patterns (5mC loads) and high 5mC/DAPI colocalization indices (Figure 3). In stark contrast, the cancerous tissue showed most cells having drastically reduced 5mC signal and significantly lower 5mC/DAPI indices. Again, in Figure 3, for the displayed tissue sections we selected only nuclei (N1 and N2) of cells that represented the most prevalent 5mC features within the epithelial compartment: that is either similar (blue) or likely similar (green) nuclei.

Sputum of cancer patients contain cells resembling malignant 5mC features

Notably, using 3D-qDMI the same differential 5mC phenotypes were also present in cells from matching sputum of the lung cancer patients versus sputum cells of the healthy donors (with no history of cancer) (Figure 3). In the sputum from normal (healthy) donors we only found nuclei of epithelial cells with high resemblance to the most dominant cytometric 5mC/DAPI indices found in BEAS-2B cells and cells in the phenotypically normal area of the lung tissue section. Thus, the selected representing nuclei N1 and N2 are very similar in their 5mC features. In comparison, the sputum from cancer patients contained severely hypomethylated cells amongst normally methylated cells. For this case Figure 3 is showing the selected cell nuclei that differ in 5mC/DAPI indices: N1 represents the normally methylated cell and N2 constitutes an aberrantly hypomethylated cell. Please note that the displayed high-resolution images only exhibit a portion of the whole slide. The aberrant cells in the cancer sputum share a high likeliness in 5mC/DAPI indices with hypomethylated tumor cells in biopsied

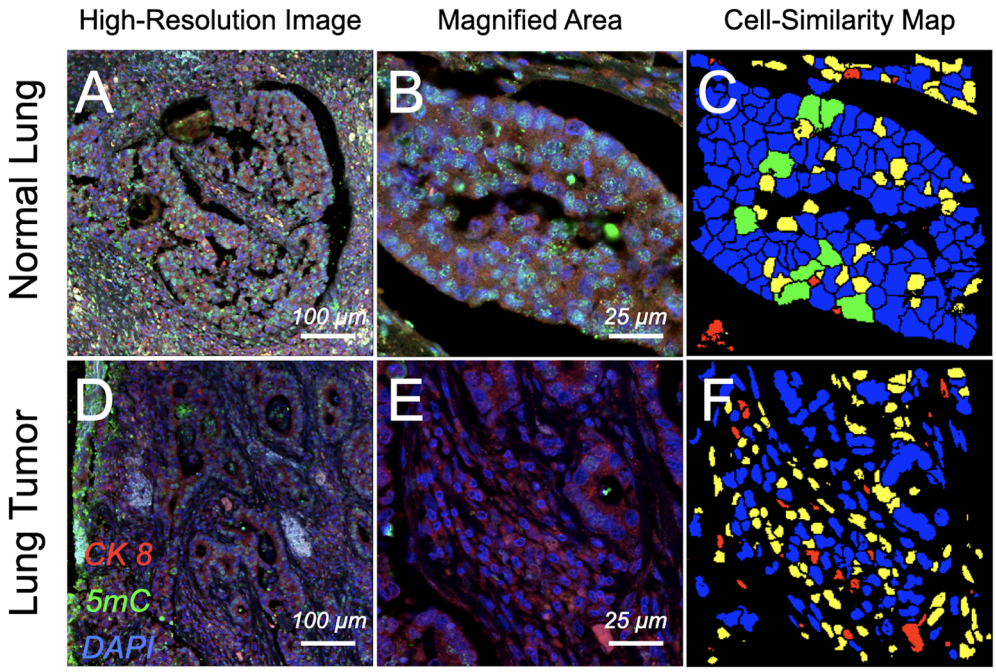


FIGURE 2
Confocal images of immunofluorescence-labeled lung cancer and adjacent normal tissue section. Surgically resected normal parenchyma and tumoral regions from patients that were diagnosed with lung cancer were *in situ* labeled for differential patterns of global DNA methylation. Cell nuclei containing global DNA marked by DAPI (false-colored blue) in normal lobules **(A)** and a magnified subarea on the same section **(B)** show higher degree of DNA methylation (5mC, false-colored green) compared with severely hypomethylated nuclei in ductal regions of the tumor **(D)** and a respectively magnified subarea **(E)**; cytokeratin 8 (CK 8) (red) was used as a marker to delineate the epithelial compartments. Cell-similarity maps **(C, F)** illustrate higher heterogeneity in cell composition for the tumor area compared to normal tissue, illustrated by the higher number of unlikely similar (yellow) and dissimilar (red) cells. In comparison normal tissue is populated by a majority of similar (green) and likely similar (blue) cells, thus presenting a high degree of cellular homogeneity.

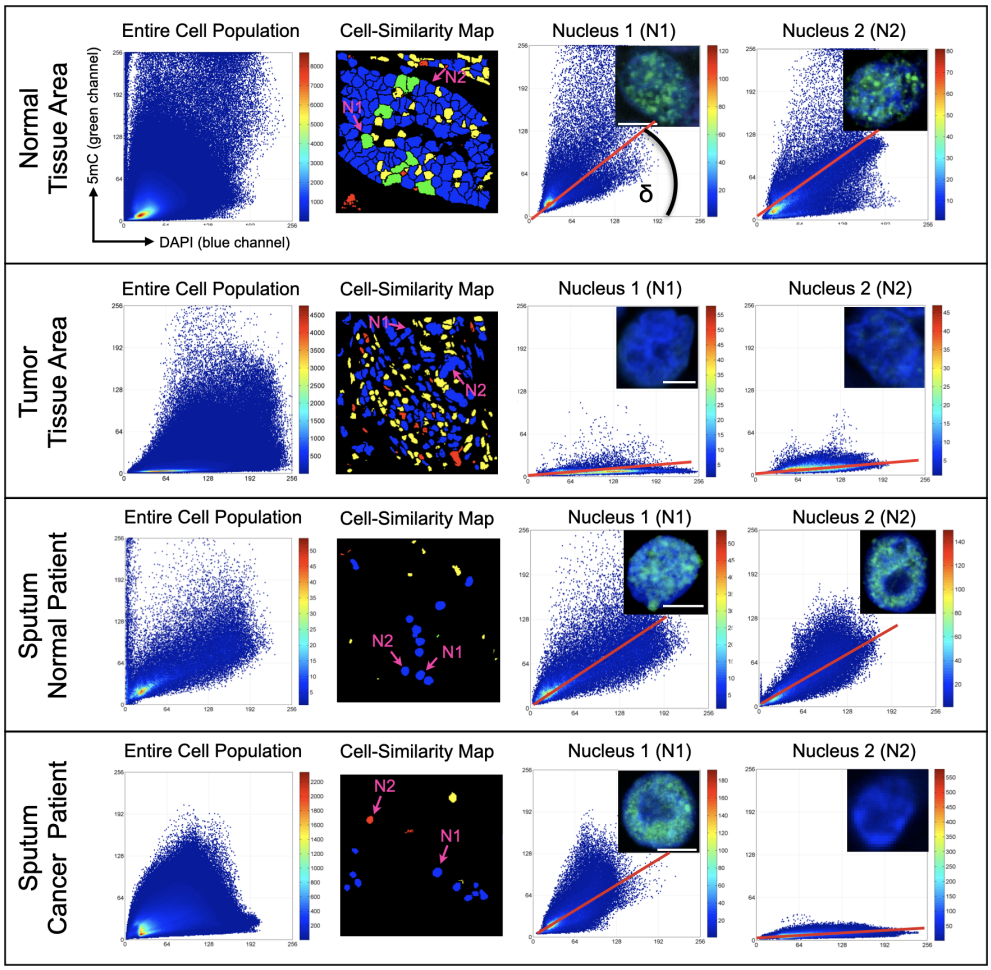


FIGURE 3
Comparative DNA methylation phenotyping of patient tissues and matching sputum. The normal lung epithelium as well as the normal sputum contain a majority of epithelial cells that display nuclear 5mC features with a high degree of similarity to the patterns seen in BEAS-2B cells. The two cell sources appear quite homogeneous in terms of 5mC topology. In contrast, the tumoral region appears more heterogeneous and contains a majority of severely hypomethylated cells. The selected nuclei N1 and N2 in the first three respective panels represent the most prevalent 5mC topology for the normal tissue and sputum as well as for the tumor area. The sputum from cancer patients contained severely hypomethylated cells amongst normally methylated cells. For this case the selected nuclei differ in 5mC indices: cell nucleus N1 represents the normally methylated cell whereas N2 constitutes an aberrantly hypomethylated cell. White bars are 5 μm.

cancer tissues and the more aggressive H157 cell line (with higher metastatic potential). These sputum cells are characterized by an extremely flat regression line ($\delta < 15°$).

Discussion

Measuring alterations in DNA methylation is a valuable method for detecting cancer cells. This phenomenon correlates with early events in carcinogenesis and tumor progression and can serve as a signature in early cancer detection and for therapeutic monitoring.

Cytometric analysis of global DNA in conjunction with cell imaging was introduced for tissue characterization towards the end of the 1990ies but did not gain much popularity in comparison to contemporaneously developed molecular methods. These methods are based on different platforms such as high-pressure liquid chromatography (HPLC), liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS), the more recent capillary electrophoresis (CE), and more simplified assays: such as the luminometric methylation assay (LUMA), enzyme-linked immunosorbent assay (ELISA) based assays, as well as pyrosequencing and the real-time PCR based MethyLight (9, 36–43). The latter two methods

approximate global DNA methylation by quantifying LINE-1 and *Alu* DNA.

Nevertheless, cytometric analyses such as image cytometry and flow cytometry have the advantage of being nondestructive and requiring no error-prone DNA extraction steps, while providing information at single-cell resolution in a highly paralleled and throughput fashion. However, image cytometry was applied in combination with radio-labeled or enzymatic reporters for detection, which either lack sensitivity, multiplexing capability or affect repeatability (consistency) of the assay, and failed to provide enough significance in differential results due to low image resolution. Furthermore, flow-cytometry provides either average 5mC measurements across a large population of cells or only quantifies mean 5mC intensity values in cell nuclei. In the meantime, molecular methods have advanced to also measure analyte contents at single-cell resolution. Yet, both technologies still lack to produce the quality of information that *in situ* methods yield regarding subcellular target localization and distribution.

In contrast to these methods including previous cell imaging-based attempts, 3D-qDMI leverages the extraction of differential 5mC topology by considering secondary effects of DNA methylation imbalances that occur throughout cellular transformation, especially hypomethylation of global DNA. In particular, the latter mechanism elicits reorganization of the genome within cell nuclei, affecting nuclear architecture (20). This phenomenon is well described in basic cell biological research but has not yet been exploited towards more clinical application. The approach we developed covers this gap and displays the relevant changes as intensity distribution of the two types of signals that reflect said phenomena: (a) 5mC signals created through immunofluorescence targeting using a sandwich assay and (b) gDNA represented by DAPI-signals that are generated by subsequent counterstaining of the same cells, as DAPI intercalates into AT-rich DNA, the main component of highly repetitive and compact heterochromatic sequences. The process results in images that represent maps of interrogated cells with a spectrum of differential DNA methylation phenotypes represented by 5mC/DAPI texture features. One such texture feature that we used to characterize cells is the 5mC/DAPI colocalization index.

Using the 5mC/DAPI index we were able to distinguish between the different cell types based on their differential 5mC/DAPI distribution patterns (scatter plots). In all comparisons between normal and cancerous cells, from cultured cell models over human tissues and epithelial cells derived from patient sputum, we basically saw the same differential 5mC/DAPI distributions and resulting 5mC/DAPI indices. The significant reduction of global 5mC in cancerous cells versus normal epithelial cells, specifically in areas of higher DNA density — delineating more compact genomic regions that predominantly harbor repeat sequences— leads to a shift in the nuclear colocalization of 5mC and gDNA constituted by a lower 5mC/DAPI index. Especially the resemblance between the cell signatures of the more aggressive H157 cells and the hypomethylated N2-type cells found in the sputum of cancer patients and typical tumor tissue cells indicates that the sputum of a cancer patient contains exfoliated epithelial cells of the tumor that can be detected based on aberrant global 5mC content and nuclear distribution.

The results of our analyses are very intriguing and can play a central role in detecting abnormal cells in sputum samples. Nuclear DNA methylation topology may serve as a novel biomarker for the noninvasive detection of malignant cells of the respiratory tract. The fact that epigenetic markers such as 5mC change early in tumor development makes 3D-qDMI in conjunction with noninvasive sputum cytology an attractive approach to be assessed for early lung cancer detection.

Final conclusions

This study provides proof-of-concept that normal and cancerous cells can be distinguished by their 5mC/DAPI topology as represented by the 5mC/DAPI colocalization index. This method could also differentiate between normal and tumor tissue of the lung and identify exfoliated aberrant cells in the sputum of cancer patients. These results encourage further feasibility and testing of 5mC/DAPI codistribution as a biomarker in noninvasive early lung cancer detection.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by VA Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

Author contributions

JT and FM provided concept and design of the study. FM contributed with providing cultured cells, tissue and sputum samples. JT performed confocal imaging, image analysis, and data analysis with the support of NG. JT wrote the initial draft of the manuscript. FM and NG contributed to manuscript revision

and manuscript submission. All authors contributed to the article and approved the submitted version.

Funding

This study was supported in part by a grant from the Department of Surgery and in part by Technology Ventures (both Cedars-Sinai).

Acknowledgments

We thank Dr. Gholamhossein Pezeshkpoor (at the time of the study with the Department of Pathology, West Los Angeles VA) for providing deidentified lung tissue.

Conflict of interest

The Authors have submitted a patent application relevant to the outcome of the study reported in the manuscript.

The authors declare that the research was conducted in the absence of any other 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. Jones PA, Baylin SB. The epigenomics of cancer. *Cell* (2007) 128:683–92. doi: 10.1016/j.cell.2007.01.029

2. Ehrlich M. DNA Hypomethylation in cancer cells. *Epigenomics* (2009) 1:239–59. doi: 10.2217/epi.09.33

3. de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DD. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet* (2011) 7:e1002384. doi: 10.1371/journal.pgen.1002384

4. Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* (1983) 301:89–92. doi: 10.1038/301089a0

5. Gama-Sosa MA, Slagel VA, Trewyn RW, Oxenhandler R, Kuo KC, Gehrke CW, et al. The 5-methylcytosine content of DNA fom human tumors. *Nucleic Acid Res* (1983) 11:6883–94. doi: 10.1093/nar/11.19.6883

6. Feinberg AP, Ohlsson R, Henikoff S. The epigenetic progenitor origin of human cancer. *Nat Rev Genet* (2006) 7:21–33. doi: 10.1038/nrg1748

7. Chalitchagorn K, Shuangshoti S, Hourpai N, Kongruttanachok N, Tangkijvanich P, Thong-ngam N, et al. Distinctive pattern of LINE-1 methylation level in normal tissues and the association with carcinogenesis. *Oncogene* (2004) 23:8841–46. doi: 10.1038/sj.onc.1208137

8. Ikeda K, Shiraishi K, Eguchi A, Shibata H, Yoshimoto K, Mori T, et al. Long interspersed nucleotide element 1 hypomethylation is associated with poor prognosis of lung adenocarcinoma. *Ann Thorac Surg* (2013) 96:1790–94. doi: 10.1016/j.athoracsur.2013.06.035

9. Weisenberger DJ, Campan M, Long TI, Kim M, Woods C, Fiala E, et al. Analysis of repetitive element DNA methylation by MethyLight. *Nucleic Acids Res* (2005) 33:6823–36. doi: 10.1093/nar/gki987

10. Gainetdinov IV, Kapitskaya KY, Rykova EY, Ponomaryova AA, Cherdyntseva NV, Vlassov VV, et al. Hypomethylation of human-specific family of LINE-1 retrotransposons in circu-lating DNA of lung cancer patients. *Lung Cancer* (2016) 99:127–30. doi: 10.1016/j.lungcan.2016.07.005

11. Suzuki M, Shiraishi K, Eguchi A, Ikeda K, Mori T, Yoshimoto K, et al. Aberrant methylation of LINE-1, SLIT2, MAL and IGFBP7 in non-small cell lung cancer. *Oncol Rep* (2013) 29:1308–14. doi: 10.3892/or.2013.2266

12. Daskalos A, Nikolaidis G, Xinarianos G, Savvari P, Cassidy A, Zakopoulou R, et al. Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer. *Int J Cancer* (2009) 124:81–7. doi: 10.1002/ijc.23849

13. Kitkumthorn N, Tuangsintanakul T, Rattanatanyong P, Tiwawech D, Mutirangura A. LINE-1 methylation in the peripheral blood mononuclear cells of cancer patients. *Clin Chim Acta* (2012) 413:869–74. doi: 10.1016/j.cca.2012.01.024

14. Zhu ZZ, Sparrow D, Hou L, Tarantini L, Bollati V, Litonjua AA, et al. Repetitive element hypomethylation in blood leukocyte DNA and cancer incidence, prevalence, and mortality in elderly individuals: The normative aging study. *Cancer Causes Control* (2011) 22:437–47. doi: 10.1007/s10552-010-9715-2

15. Saito K, Kawakami K, Matsumoto I, Oda M, Watanabe G, Minamoto T. Long interspersed nuclear element 1 hypomethylation is a marker of poor prognosis in stage IA non-small cell lung cancer. *Clin Cancer Res* (2010) 16:2418–26. doi: 10.1158/1078-0432.CCR-09-2819

16. Kurdyukov S, Bullock M. DNA Methylation analysis: Choosing the right method. *Biol (Basel)* (2016) 5:3. doi: 10.3390/biology5010003

17. Li S, Tollefsbol TO. DNA Methylation methods: Global DNA methylation and methylomic analyses. *Methods* (2021) 187:28–43. doi: 10.1016/j.ymeth.2020.10.002

18. Laird PW. Principles and challenges of genome-wide DNA methylation analysis. *Nat Rev Genet* (2010) 11:191–203. doi: 10.1038/nrg2732

19. Arney KL, Fisher AG. Epigenetic aspects of differentiation. *J Cell Sci* (2004) 117:4355–63. doi: 10.1242/jcs.01390

20. Espada J, Esteller M. Epigenetic control of nuclear architecture. *Cell Mol Life Sci* (2007) 64:449–57. doi: 10.1007/s00018-007-6358-x

21. Gilbert N, Thomson I, Boyle S, Allan J, Ramsahoye B, Bickmore WA. DNA Methylation affects nuclear organization, histone modifications, and linker histone binding but not chromatin compaction. *J Cell Biol* (2007) 177:401–11. doi: 10.1083/jcb.200607133

22. Kobayakawa S, Miike K, Nakao M, Abe K. Dynamic changes in the epigenomic state and nuclear organization of differentiating mouse embryonic stem cells. *Genes Cells* (2007) 12:447–60. doi: 10.1111/j.1365-2443.2007.01063.x

23. Yamagata K. DNA Methylation profiling using live-cell imaging. *Methods* (2010) 52:259–66. doi: 10.1016/j.ymeth.2010.04.008

24. Kumar N, Hori Y, Kikuchi K. Live-cell imaging of DNA methylation based on synthetic-Molecule/Protein hybrid probe. *Chem Rec* (2018) 18:1672–80. doi: 10.1002/tcr.201800039

25. Gertych A, Wawrowsky KA, Lindsley E, Vishnevsky E, Farkas DL, Tajbakhsh J. Automated quantification of DNA demethylation effects in cells *via* 3D mapping of nuclear signatures and population homogeneity assessment. *Cytometry A* (2009) 75:569–83. doi: 10.1002/cyto.a.20740

26. Gertych A, Farkas DL, Tajbakhsh J. Measuring topology of low-intensity DNA methylation sites for high-throughput assessment of epigenetic drug-induced effects in cancer cells. *Exp Cell Res* (2010) 316:3150–60. doi: 10.1016/j.yexcr.2010.08.013

27. Tajbakhsh J, Gertych A, Fagg WS, Hatada S, Fair JH. Early *in vitro* differentiation of mouse definitive endoderm is not correlated with progressive maturation of nuclear DNA methylation patterns. *PLoS One* (2011) 6:e21861. doi: 10.1371/journal.pone.0021861

28. Oh JH, Gertych A, Tajbakhsh J. Nuclear DNA methylation and chromatin condensation phenotypes are distinct between normally proliferating/aging, rapidly growing/immortal, and senescent cells. *Oncotarget* (2013) 4:474–93. doi: 10.18632/oncotarget.942

29. Gertych A, Oh JH, Wawrowsky KA, Weisenberger DJ, Tajbakhsh J. 3-d DNA methylation phenotypes correlate with cytotoxicity levels in prostate and

liver cancer cell models. *BMC Pharmacol Toxicol* (2013) 14:11. doi: 10.1186/2050-6511-14-11

30. Tajbakhsh J, Stefanovski D, Tang G, Wawrowsky K, Liu N, Fair JH. Dynamic heterogeneity of DNA methylation and hydroxymethylation in embryonic stem cell populations captured by single-cell 3D high-content analysis. *Exp Cell Res* (2015) 332:190–201. doi: 10.1016/j.yexcr.2015.02.004

31. Stefanovski D, Tang G, Wawrowsky K, Boston RC, Lambrecht N, Tajbakhsh J. Prostate cancer diagnosis using epigenetic biomarkers, 3D high-content imaging and probabilistic cell-by-cell classifiers. *Oncotarget* (2017) 8:57278–301. doi: 10.18632/oncotarget.18985

32. Edderkaoui M, Chheda C, Soufi B, Zayou F, Hu RW, Ramanujan VK, et al. An inhibitor of GSK3B and HDACs kills pancreatic cancer cells and slows pancreatic tumor growth and metastasis in mice. *Gastroenterology* (2018) 155:1985–98. doi: 10.1053/j.gastro.2018.08.028

33. Tajbakhsh J. Covisualization of global DNA Methylation/Hydroxymethylation and protein biomarkers for ultrahigh-definition epigenetic phenotyping of stem cells. *Methods Mol Biol* (2020) 2150:79–92. doi: 10.1007/7651_2019_276

34. Reddel RR, Ke Y, Gerwin BI, McMenamin MG, Lechner JF, Su RT, et al. Transformation of human bronchial epithelial cells by infection with SV40 or adenovirus-12 SV40 hybrid virus, or transfection *via* strontium phosphate coprecipitation with a plasmid containing SV40 early region genes. *Cancer Res* (1988) 48:1904–09.

35. Giard DJ, Aaronson SA, Todaro GJ, Arnstein P, Kersey JH, Dosik H, et al. *In vitro* cultivation of human tumors: Establishment of cell lines derived from a series of solid tumors. *J Natl Cancer Inst* (1973) 51:1417–23. doi: 10.1093/jnci/51.5.1417

36. Piyathilake CJ, Johanning GL, Frost AR, Whiteside MA, Manne U, Grizzle WE, et al. Immunohistochemical evaluation of global DNA methylation: Comparison with *in vitro* radiolabeledmethyl incorporation assay. *Biotech Histochem* (2000) 75:251–58. doi: 10.3109/10520290009085128

37. Piyathilake CJ, Frost AR, Bell WC, Oelschlager D, Weiss H, Johanning GL, et al. Altered global methylation of DNA: An epigenetic difference in susceptibility for lung cancer is associated with its progression. *Hum Pathol* (2001) 32:856–62. doi: 10.1053/hupa.2001.26471

38. Soares J, Pinto AE, Cunha CV, André S, Barão I, Sousa JM, et al. Global DNA hypomethylation in breast carcinoma: Correlation with prognostic factors and tumor progression. *Cancer* (1999) 85:112–18. doi: 10.1002/(SICI)1097-0142(19990101)85:1<112::AID-CNCR16>3.0.CO;2-T

39. Kuo KC, McCune RA, Gehrke CW, Midgett R, Ehrlich M. Quantitative reversed-phase high performance liquid chromatographic determination of major and modified deoxyribonucleosides in DNA. *Nucleic Acids Res* (1980) 8:4763–76. doi: 10.1093/nar/8.20.4763

40. Song L, James SR, Kazim L, Karpf AR. Specific method for the determination of genomic DNA methylation by liquid chromatography-electrospray ionization tandem mass spectrometry. *Anal Chem* (2005) 77:504–10. doi: 10.1021/ac0489420

41. Stach D, Schmitz OJ, Stilgenbauer S, Benner A, Döhner H, Wiessler M, et al. Capillary electrophoretic analysis of genomic DNA methylation levels. *Nucleic Acids Res* (2003) 31(2):E2. doi: 10.1093/nar/gng002

42. Karimi M, Johansson S, Stach D, Corcoran M, Grandér D, Schalling M, et al. LUMA (LUminometric methylation assay)–a high throughput method to the analysis of genomic DNA methylation. *Exp Cell Res* (2006) 312:1989–95. doi: 10.1016/j.yexcr.2006.03.006

43. Ogino S, Nosho K, Kirkner GJ, Kawasaki T, Chan AT, Schernhammer ES, et al. A cohort study of tumoral line-1 hypomethylation and prognosis in colon cancer. *J Natl Cancer Inst* (2008) 100:1734–38. doi: 10.1093/jnci/djn359

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 01 November 2022
DOI 10.3389/fonc.2022.1019442

Check for updates

OPEN ACCESS

EDITED BY
Miao Liu,
Harvard Medical School, United States

REVIEWED BY
Jing Wang,
Nanjing Drum Tower Hospital, China
Xuefeng Wang,
Moffitt Cancer Center, United States

*CORRESPONDENCE
Wu Yin
wyin@nju.edu.cn
Yan Chen
amandacy@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

RECEIVED 15 August 2022
ACCEPTED 17 October 2022
PUBLISHED 01 November 2022

CITATION
Li Z, Zeng T, Zhou C, Chen Y and Yin W (2022) A prognostic signature model for unveiling tumor progression in lung adenocarcinoma. *Front. Oncol.* 12:1019442. doi: 10.3389/fonc.2022.1019442

COPYRIGHT
© 2022 Li, Zeng, Zhou, Chen and Yin. 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 prognostic signature model for unveiling tumor progression in lung adenocarcinoma

Zijian Li[1†], Tao Zeng[1†], Chong Zhou[1], Yan Chen[2*] and Wu Yin[1*]

[1]State Key Lab of Pharmaceutical Biotechnology, College of Life Sciences, Nanjing University, Nanjing, Jiangsu, China, [2]Department of Chinese Medicine, Jiangsu Cancer Hospital & Jiangsu Institute of Cancer Research & The Affiliated Cancer Hospital of Nanjing Medical University, Nanjing, Jiangsu, China

A more accurate prognosis is important for clinical treatment of lung adenocarcinoma. However, due to the limitation of sample and technical bias, most prognostic signatures lacked reproducibility, and few were applied to clinical practice. In addition, understanding the molecular driving mechanism is indispensable for developing more promising therapies for lung adenocarcinoma. Here, we built an unbiased prognostic significance model to perform an integrative analysis, including differentially expressed genes and clinical data with lung adenocarcinoma patients from TCGA. Multivariable Cox proportional hazards model with the Lasso penalty and 10-fold cross-validate were used to identify the best gene signature. We generated a 17-gene signature for prognostic risk prediction based on the overall survival time of lung adenocarcinoma patients. To further test the model's predictive ability, we have applied an independent GEO database to verify the predictive ability of prognostic signature. The model can more objectively describe several biological processes related to tumors and reveal important molecular mechanisms in tumor development by GO and KEGG analysis. Furthermore, differential expression analysis by GSEA revealed that tumor microenvironments such as ER stress, exosome, and immune microenvironment were enriched. Using single-cell RNA sequence technology, we found that risk score was positively correlated with lung adenocarcinoma marker genes and copy number variation but negatively correlated with lung epithelial marker genes. High-risk cell populations with the model had stronger cancer stemness and tumor-related pathway activation. As we expected, the risk score was in accordance with the malignancy of each cluster from tumor progression. In conclusion, the risking model established in this study is more reliable than others in evaluating the prognosis of LUAD patients.

KEYWORDS

lung adenocarcinoma, tumor microenvironment, single-cell RNA sequence, risk score, tumor progress

Introduction

Lung adenocarcinoma (LUAD) is the most common subtype of lung cancer and a significant cause of cancer-related death worldwide (1, 2). Most LUAD patients are diagnosed in advanced or metastatic stages, which is the primary cause of mortality in lung cancer (3). However, LUAD patients' prognosis is far from satisfying, and its associated microenvironments remain poorly understood (1). Even stage I lung cancer has a poor prognosis with 5-year overall survival (OS) after surgical resection (4), revealing the need for treatment escalation, such as adjuvant therapy. Notably, LUAD is a complex disease involving multiple pathways in pathogenesis. Thus, an in-depth understanding of the driven molecular mechanisms of LUAD is indispensable for developing more promising therapies.

Investigators have continued to seek prognostic signature that are predictive of survival benefit, as it is the basis for developing personalized approaches to improve the survival of early-stage lung cancer patients (5, 6). Many studies have proposed genomic signatures for risk score and survival prediction in lung cancer patients (7–9). However, most prognostic signatures lacked reproducibility due to problematic issues such as limited sample size, individual heterogeneity, and technical bias, few prognostic signatures were applied to routine clinical practice (6). The study built a significant prognostic model to perform an integrative analysis including differentially expressed genes (DEGs) and clinical data with lung adenocarcinoma patients from The Cancer Genome Atlas (TCGA) Program. Multivariable Cox proportional hazards model with the Lasso penalty and ten-fold cross-validation were used to identify the best gene signatures among different gene categories. We generated a 17-gene signature for prognostic risk prediction based on OS time with LUAD patients.

Efforts for understanding lung cancer progression have primarily focused on the profiling of cancer cells with genetic aberrations (10, 11). However, progression also can be influenced by complex and dynamic features from the tumor surroundings (12). For learning more about tumor progression, genetically engineered mice (GEM) with somatic mutation of Kras-G12D with or without TP53 deletion in alveolar type 2 cells (termed 'K' mice and 'KP' mice) can spontaneously suffer lung adenoma (13, 14), and the adenoma in 'KP' mice can progress into advanced LUAD over more than 12 weeks (15–17). This technique can control the tumor progress with increased accuracy. Single-cell RNA sequence (scRNA-seq) can offer more details about tumor, while bulk RNA sequence can only offer an overview description. Additionally, it can be a powerful tool to characterize each cell in a tumor, which could help us understand more about tumor characteristics (18–20). To validate the survival, scRNA-seq and GEM were used to identify more details regarding tumor progression through the prognostic signature.

Material and methods

Data collection

RNA sequence and clinical data with lung adenocarcinoma patients from TCGA were downloaded and prepared using the R 'TCGAbiolinks' package (21). The gene expression profiles include 520 primary tumor samples, which could correspond with clinical data and 59 normal tissue. Gene expression omnibus (GEO) database including GSE43458 (110 samples), GSE10072 (107 samples), GSE32863 (116 samples), GSE31210 (246 samples) and GSE50081 (181 samples) were downloaded by 'GEOquery' R package (22). Probs in GSE43458 (110 samples), GSE10072 (107 samples) and GSE32863 (116 samples) were annotated by 'hgu133a.db' and 'annotate' R packages; Probs in GSE31210 and GSE50081 were annotated by 'hgu133plus2.db' and 'annotate' R package (23). When a gene was mapped to different probes, the genic expression value was calculated by the average expression value. Furthermore, gene expression data from the GEO (https://www.ncbi.nlm.nih.gov/geo/) and GDC databases (https://cancergenome.nih.gov/) were z-score transformed for survival analysis. The clinical features of the TCGA and GEO patients is shown in Table S1 and Table S2.

Construction prognostic signature

The 'limma' R package performed differential expression analysis on primary tumor and normal tissues (24). P-values and fold changes were controlled for false discovery rate (FDR< 0.05 and |logFC| > 0.5). The 'survival' R package performed univariate Cox analyses of OS to identify the prognostic signature (25). The "Coxph" function was used to build a Univariate Cox model and calculate the p-value and C-index (consistency index). In addition, the "Cox.zph" function was used to test the proportional hazards assumption for a Cox regression model fit. (p< 0.05 and C-index > 0.58 and p-value of Proportional Hazards Assumption larger than 0.4).

Multivariable Cox regression analysis was performed to analyze the overall probability of survival. Lasso regression was performed by the 'glmnet' R package to reduce the number of genes, and a 10-fold cross-validation was performed to set proper lambda, and finally, 17 genes were left (26).

$$RiskScore = \sum_{i \in RiskGenes} w_i * mRNA_i$$

w represents the Lasso coefficient index of risk genes, *mRNA* represent gene expression, and the gene expression level of each gene, respectively.

The differential between high risk (HR > 1) and low risk (LR< 1) was confirmed by Kaplan-Meier method (log-rank test). The 'timeROC' R package performed a receiver operating

characteristic curve (ROC) analysis to assess the predictive efficiency of the prognostic signature (27). C-index was used to evaluate the model's predictive ability. Multivariable Cox regression was used to integrate different predictive factors (upper and lower limits of 95% confidence interval). The GEO database, including GSE31210 and GSE50081, validated the model. C20orf197 in GSE31210 and GSE50081 is missing, and the expression of C20orf197 in two databases was assigned 0.

Function and enrichment analysis

Endoplasmic reticulum (ER) stress-related genes were collected from 'Msigdb' (28), and exosome-related genes were collected from ExoCarta (29, 30). The 'GSVA' R package was applied to perform a single-sample gene set enrichment analysis (ssGSEA) to quantify the ssGSEA score of immune signatures and endoplasmic reticulum stress and exosome-related genes (31). Differential genes in LUAD samples ad. pavlue< 0.05 and logFC > 0.5 or logFC< -0.5. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed by the 'clusterProfiler' R package. Gene sets were collected from 'Msigdb' based on the DEGs, which was done by the 'limma' R package between high-risk and low-risk groups (28, 32). To estimate the activation of hallmark pathways, the GSEA analysis was applied with standard settings (32). To analyze the tumor immune microenvironment of the LUAD samples, the 'GSVA' R package was used to perform ssGSEA to quantify the ssGSEA score of 29 immune cell-related gene sets (33, 34). The Wilcox test tested the differential expression analysis of immune-related signatures, and the correlation between risk score and immune signatures was performed by Spearman correlation analysis. Differentially expressed genes about immune signatures were done by 'limma' R package (P.Value< 0.005). The same methods are performed in ER stress and exosome analysis.

Single-cell analysis

Single-cell RNA sequence data (after depth normalization) was downloaded from GEO database (GSE154989). Seurat v4.0.4 was used for single-cell analysis (35–38). The count for the genes in each cell was log normalized, and the ScaleData function was used for scaling. SCTransform function was used for correcting different animals and plates, and the top 40 principal components were used to construct the SNN graph and embedding. FindClusters function was used for clustering cells at 0.5 resolution. AUCell v1.15.0 was used for scoring the ER stress and exosome gene sets in scRNA-seq (39). InferCNV v1.9.1 was used for the copy number variation (CNV) assignment with default parameters (40). Raw CNV score for each cell was collected from 'infercnv.observations.txt' file and transformed into 3-level scoring (0.7~1.3 assign 0; 0~0.3 or larger than 1.5 assigns 2; 0.3 ~ 0.7 or 1.3~1.5 assigns 1). Cluster 11, the earliest cell type T, was considered as the control and had no CNV. The CNV score for each cell was the sum of every gene's CNV Score.

For estimating the trajectory of scRNA-seq, Monocle v2.21.1 was used to estimate and order cells in pseudotime along a trajectory (41). Several ablation experiments were done to select proper genes to estimate pseudotime. High variable genes calculated by Monocle and stemness genes collected from 'Msigdb' were used to estimate pseudotime. Cluster 11, which appeared in the early stage of LUAD, was set to the original point of pseudotime. Monocle V2 was used to obtain highly variable genes, DDRtree was used to establish the minimum spanning tree after dimensionality reduction.

Statistical analysis

Multivariable and univariate Cox regression were used to analyze the probability of OS, and KM method was used to test the difference between high-risk and low-risk patients. Chi-squared test tested differences between risk scores and clinical information. ER stress scores, exosome scores, and immune signatures between the two groups were examined using Wilcox test and Spearman correlation analysis. Statistical analyses were performed using R v.4.1.1. The detailed analysis methods in the website (https://github.com/ZengTaox/xlw/blob/main/upload.R).

Results

DEGs identification and construction of a prognostic model in TCGA cohort

Five hundred ninety-two LUAD and adjacent non-tumorous samples from the TCGA database were included in DEGs and prognostic genes for OS. Meanwhile, 520 LUAD samples from the TCGA database and 427 patients from two GEO cohorts were incorporated into the following study about prognostic model.

Firstly, compared tumor and adjacent non-tumorous tissues from TCGA and get DEGs. Secondly, we use univariate Cox regression analysis in tumor tissue from TCGA to get prognostic genes (p< 0.05), 90 of 205 prognostic genes were DEGs (Figures 1A, B). To further construct a risk scoring model for predicting possibility of OS in LUAD patients, LASSO Cox regression was used to build a prognostic model, which included *PLEK2*, *PTPRH*, *OGFRP1*, *CHRNA5*, *CBFA2T3*, *SMIM15*, *AVEN*, *MELTF*, *KRT8*, *RGS20*, *FAM207A*, *SOWAHC*, *ELF5*, *LSP1P4*, *C20orf197*, *C11orf16* and *DNALI1* (Figure 1C). The univariate Cox regression analysis suggests that these genes can be used as prognostic genes (Figure 1D), and can also be used to distinguish tumor from adjacent tissues. In addition, we evaluated three extra GEO

10.3389/fonc.2022.1019442

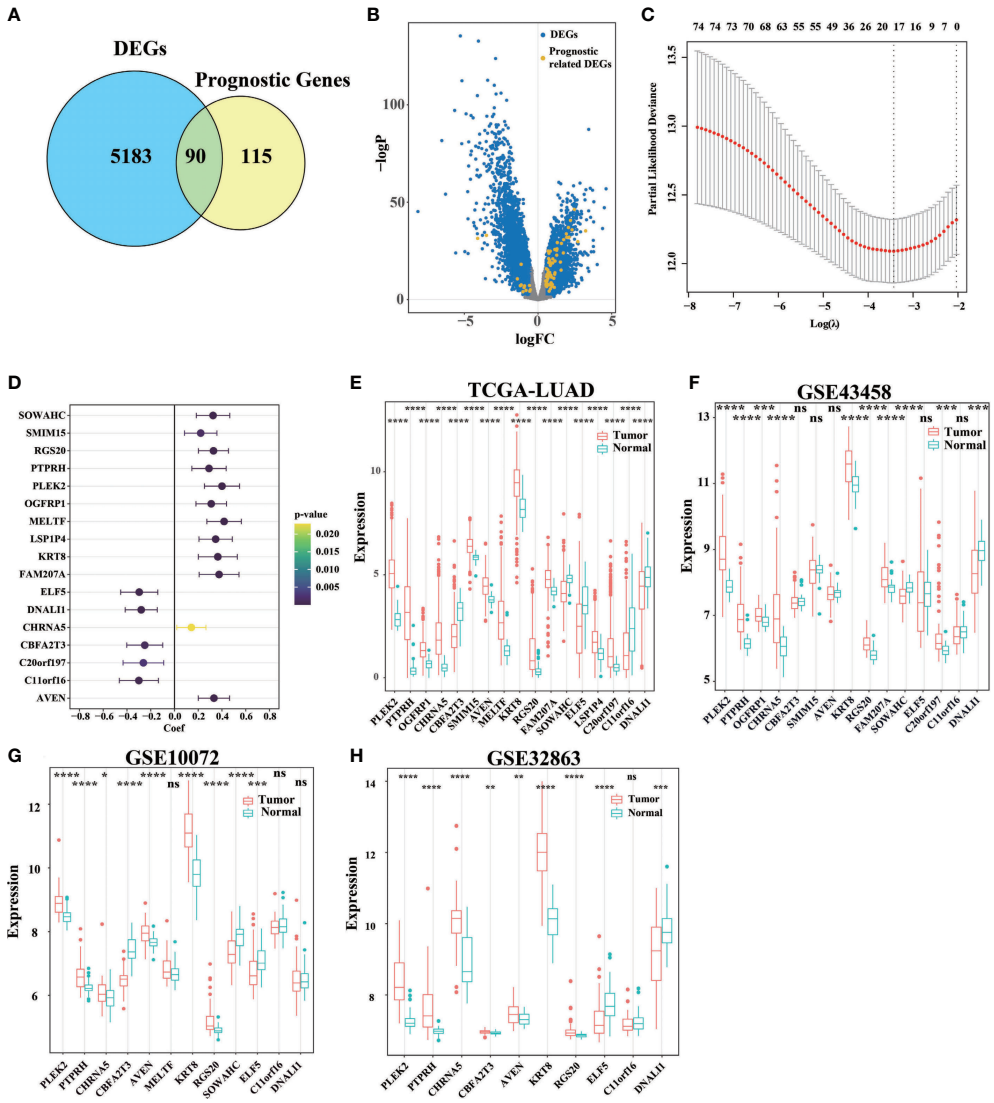


FIGURE 1
Identification of signature genes and establishment of a survival model. **(A)** A total of 5273 DEGs with adj. pvalue< 0.05 and logFC > 0.5 or logFC< -0.5 were selected by limma. 205 genes related to survival were found by stationarity test and univariate Cox regression analysis. Of these, 90 genes were associated with both survival and differential between the cancer and para-cancer groups. **(B)** Use volcano map to describe the distribution of DEGs in all genes. **(C)** LASSO Cox regression model was used to establish a multivariable model, and 10-fold cross validation was used to calculate the total deviation corresponding to differe nt penalty coefficients. **(D)** Univariate Cox regression test **(C)** The influence of the gene on survival (upper and lower 95% confidence intervals, color significance). **(E)** Analysis of gene expression differences between primary cancer (n = 533) and adjacent tissue (n = 59) in TCGA-LUAD. **(F–H)** DEGs in tumor tissue and para-cancer tissue in GEO databases (ns was non-significant, *P< 0.05, **P< 0.01, ***P< 0.001, ****P< 0.0001).

cohorts (Figures 1F–H), the prognostic genes in the above GEO databases are consistent with TCGA databases (Figure 1E).

Predictive performance and prognostic value of the model

For predictive performance of the 17-gene prognostic model in TCGA cohort, the area under the curve (AUC) in the time-dependent ROC analysis reached 0.69, 0.72, 0.75, and 0.78 (Figure 2A), indicating excellent specificity and sensitivity of the risk score for predicting OS. According to the median value of risk score, the patients were divided into high-risk and low-risk groups. PCA analysis confirmed that patients in two groups were stratified into two directions (Figure 2B). KM analysis indicated that worse prognosis and significantly poorer OS were detected in high-risk patients (two-stage test P< 0.0001, log-rank test P< 0.001, Figure 2C). Compared with low-risk group, the

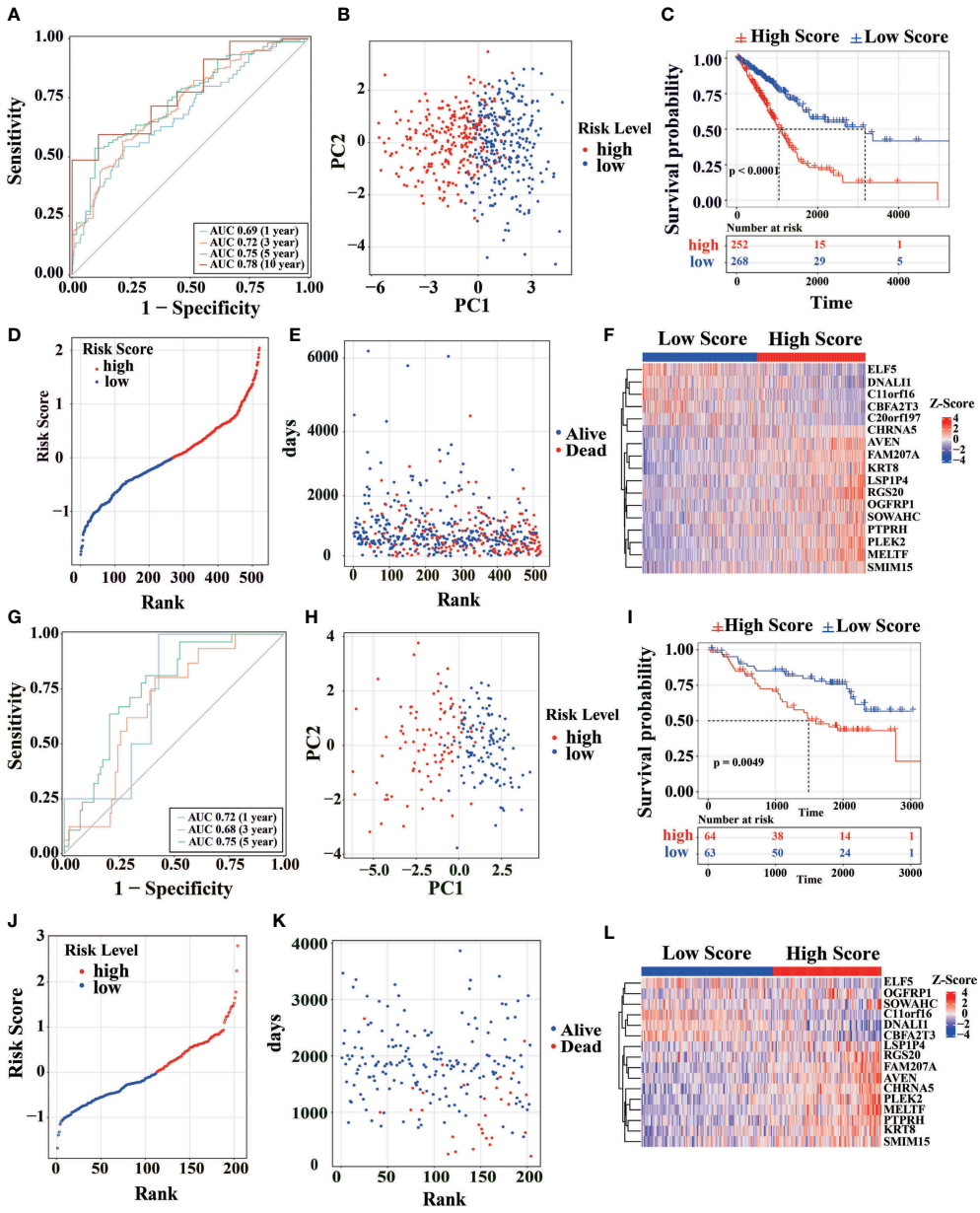


FIGURE 2
Survival analysis of the model in TCGA cohort. **(A)** Using TCGA-LUAD primary cancer samples and corresponding clinical records (n = 520) to calculate the AUC of time-dependent ROC curves of the risk score at 1, 3, 5 and 10 years. **(B)** PCA analysis was performed using the genes in Figure 1C of the primary LUAD sample in TCGA cohort. **(C)** The survival rates of LUAD patients were tested to test by KM survival analysis in the TCGA database. **(D)** Distribution map of different survival risks of LUAD patients in TCGA database. **(E)** Time distribution of LUAD patients with different survival risks and death in TCGA database (ordinate: follow-up time, abscissa: risk ranking). **(F)** Gene expression trends in Figure 1C of LUAD patients in TCGA database (top note: blue: low-risk, red: high-risk; Heat map below: Red: high expression, blue: low expression). **(G-L)** Patients from the GEO cohort (GSE31210) were analyzed similarly to the above analysis in TCGA cohort.

patients had a higher proportion of death and shorter survival time in high-risk group (Figures 2D, E). *ELF5*, *DNALI1*, *C11orf16*, *CBFA2T3*, and *C20orf197* had higher expression levels in low-risk group, while the other genes in the prognostic signature had higher expression levels in high-risk group (Figure 2F).

To further estimate the model's generalization performance, we have validated the predictive ability of prognostic signatures in the GEO database (GSE31210, Figures 2G–L; GSE50081, Figures S1A–F). In GSE31210 cohort, the AUC was 0.72 at 1 year, 0.68 at 3 years, and 0.75 at 5 years (Figure 2G). Similar to TCGA, the patients from GSE31210 cohort were divided into

different directions by PCA analysis (Figure 2H). KM survival analysis indicated that high-risk patients had a higher proportion of death (log-rank test P< 0.0049, Figure 2I) and shorter survival time (Figures 2J, K). The ELF5, DNALI1, C11orf16, and CBFA2T3 genes also had higher expression levels in the low-risk group in GSE31210 cohort (Figure 2L). The same results are shown in GSE50081 cohort (Figures S1A–F). These similar results show that the signature has good generalization performance and can potentially predict prognosis for LUAD patients.

The signature score as an independent prognostic factor in clinical features

After controlling for confounding variables, risk score of the signature remained statistically significant for OS. The clinical characteristics analysis of the cohorts is summarized in Table S1 and Table S2. As shown in Figure 3A, univariate Cox survival analysis indicated that risk score (p< 0.0001), invasion depth (T stage, p< 0.0001), distant metastasis (M stage, p< 0.05), lymph node metastasis (N stage, p< 0.0001) and clinical staging (TNM stage, p< 0.0001) were significant parameters that affect the prognosis of LUAD patients. Multivariable Cox survival analysis revealed that risk score was independent predictors of unfavorable prognosis in LUAD patients (p=4.33e-11 Figure 3B). Additionally, with LUAD TNM staging progress, the risk score increased in different degrees (Figure 3C–F). These results suggest that high-risk score might imply worse clinical symptoms regarding invasion depth, distant metastasis, and lymph node metastasis. By analyzing the risk score, we found that the model had a higher C-index, indicating that the risk score was more accurate than the traditional clinical stage (Figure 3G).

We also found the same result in GEO database (GSE31210, Figures 3H–J; GSE50081, Figures S1G–K), with LUAD TNM staging progress, the risk score increased in different degrees. These results indicate the risk scoring model, as an independent prognostic factor in clinical features, can more accurately evaluate prognosis of patients.

Functional enrichment analyses of the prognostic signature

To analyze the prognostic signature, base on differential analysis between high-risk and low-risk groups (Figure 4A), GSEA score on 50 hallmark pathways were displayed. Notably, we found some tumor-related pathways were activated in high-risk patients, such as epithelial-mesenchymal transition, mTORC1, and PI3K/Akt/mTOR pathways (TCGA, Figure 4B). The activation of mTORC1 signaling and PI3K/Akt/mTOR pathways promotes glucose metabolism and growth regulation (42). It is also worth noting that LUAD patients with epithelial-mesenchymal transition, glycolysis, and highly proliferative state were associated with poorer survival (43). The imbalance of these pathways may be related to tumor progression, indicating the poor prognosis in high-risk patients.

In TCGA and GEO cohorts (TCGA, Figure 4B; GSE31210, Figure 4C; GSE50081, Figure 4D), E2F transcription factors and c-Myc signal were activated in the high-risk group. The c-Myc signal regulates differentiation and proliferation through activated transcription and amplification of target genes in various tumors (44). In addition to mediating the cell cycle, E2F transcription factors play critical roles in tumor development and metastasis, including angiogenesis, extracellular matrix remodeling, tumor cell survival, and epithelial-mesenchymal transition (44, 45). With the activation of hypoxia-related pathways in high-risk group, unfolded protein responses are markedly activated (Figures 4B–D), and tumor cells may produce tumor-specific exosomes. Increased unfolded protein responses in high-risk group may contribute to ER stress. These factors jointly promote tumor proliferation, drug resistance, metastasis, and invasion and might even cause an immunosuppressive microenvironment (46).

Furthermore, "clusterProfiler" R package was used to conduct GO and KEGG enrichment analyses between high-risk and low-risk groups in TCGA and GEO cohorts (Figures 4E, F). Interestingly, KEGG pathway analysis indicated that the signature was associated with P53 signaling pathway, immune response, DNA replication, extracellular matrix remodeling (ECM)-receptor interaction, and glycolysis (Figure 4E). Similarly, the overlapped GO functional pathways between the three cohorts were predominantly enriched in tumor microenvironments associated with multivesicular bodies, mitotic activity, immune response, ECM, focal adhesion, and others (Figure 4F). These results imply that there might be differences in ER stress, exosome pathway, immune response, ECM, cell cycle, and proliferation between high-risk and low-risk groups.

Analyze tumor microenvironment and validation differences biological process

To further explore the differences in survival, we compared the effects of immune response, exosome pathway, and ER stress between two groups. It was observed that high-risk patients were associated with significantly higher ER stress scores (TCGA, Figure 5A; GEO, Figure S2). We also found that high-risk scores were related to some ER stress pathways (Figure 5A). Boxplots depicting ER stress-related ssGSEA scores showed that patients in high-risk group had higher scores than those in low-risk group (Figure 5B). To explore more detail about ER stress, a heatmap of ER stress-related genes was utilized to confirm the difference in both groups (Figure 5C). *Tbl2*, which is positively

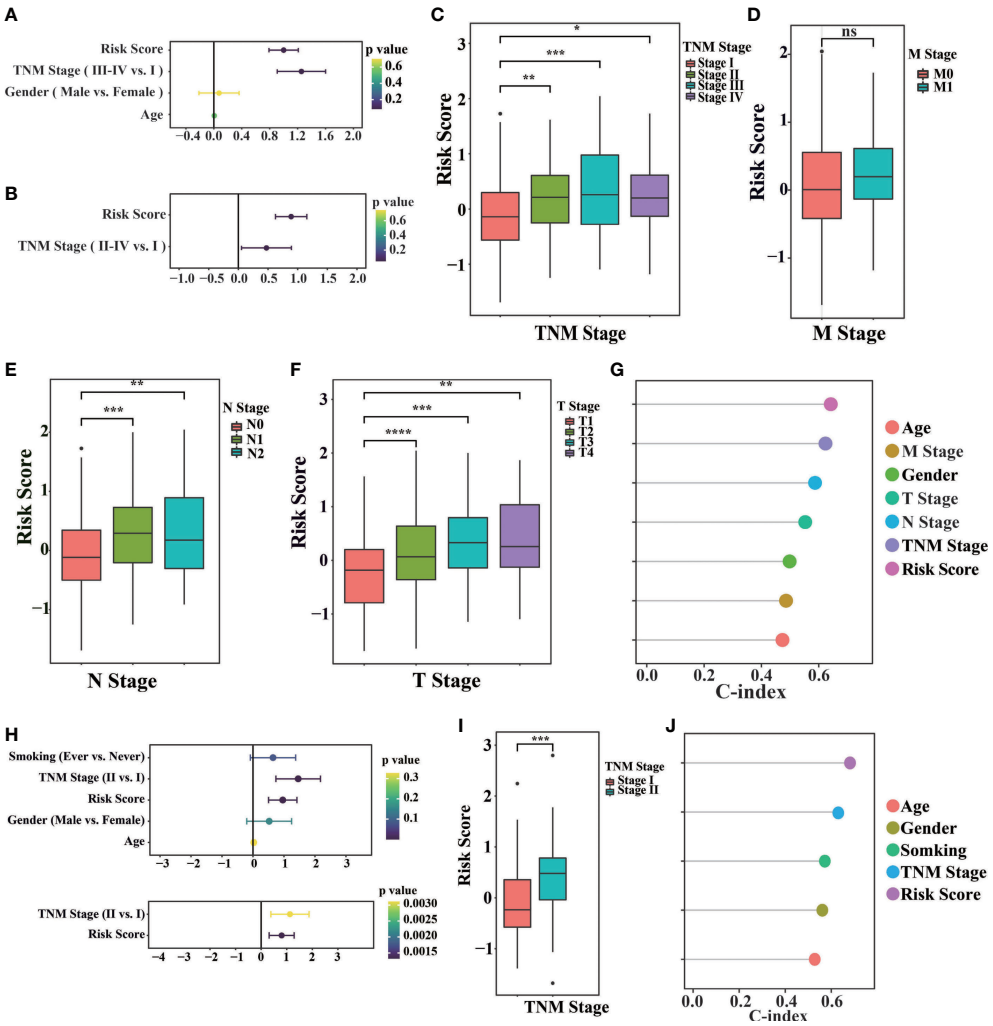


FIGURE 3
Prognostic values of the model in TCGA and GEO cohorts. **(A)** The TCGA-LUAD primary cancer samples, corresponding clinical records and corresponding risk score (n = 520) were used for univariate Cox regression. **(B)** Perform multivariable Cox regression on the significant factors in univariate Cox regression. **(C–F)** TCGA-LUAD samples with different tumor traditional clinical stage (TNM stage, M stage, N stage, T stage) and their corresponding risk scores. **(G)** Consistency index of TCGA-LUAD primary cancer samples, corresponding clinical records and corresponding risk score (n = 520) calculated by univariate Cox regression. **(H)** Using univariate Cox regression (above) in the GSE31210 cohort (n = 246), the significant factors in univariate Cox regression are performed in multivariable Cox regression (below). **(I)** samples with different TNM stages in GSE31210 cohort and their corresponding risk scores. **(J)** The consistency index of primary cancer samples, corresponding clinical records and corresponding risk scores in GSE31210 cohort was calculated by univariate Cox regression. ("ns" is non-significant, *P< 0.05, **P< 0.01, ***P< 0.001, ****P< 0.0001).

correlated with risk score, can cause ER stress through PERK-eIF2α-ATF4 axis (47) (Figures 5C, S2B). We also found that erlin1 and *EIF4EBP1* were positively correlated with risk scores (Figure 5C). ATF4-mediated induction of *erlin1* and EIF4EBP1 contributes to ER stress (48). As ER stress indicator, highly expressed *erlin1* indicates the increased ER stress in high-risk group (49) (Figure S2C). Therefore, ER stress is significantly involved in high-risk patients.

Meanwhile, we also found significant differences in exosome pathways. Compared with low-risk group, high-risk patients had higher exosome-related ssGSEA scores in TCGA (Figure 5D, additional data in GEO database, Figure S3). Cancer-derived exosome are implicated in various carcinogenesis processes, including malignant transformation, angiogenesis, immunosuppression, invasion, and treatment resistance (50, 51). There is also a significant correlation between exosome-related ssGSEA score and risk score in TCGA database (Figure 5E). Exosome-related genes also demonstrated different expression levels between two groups (Figure 5F). As a scaffold protein of exosomes, *PTGFRN* is positively correlated

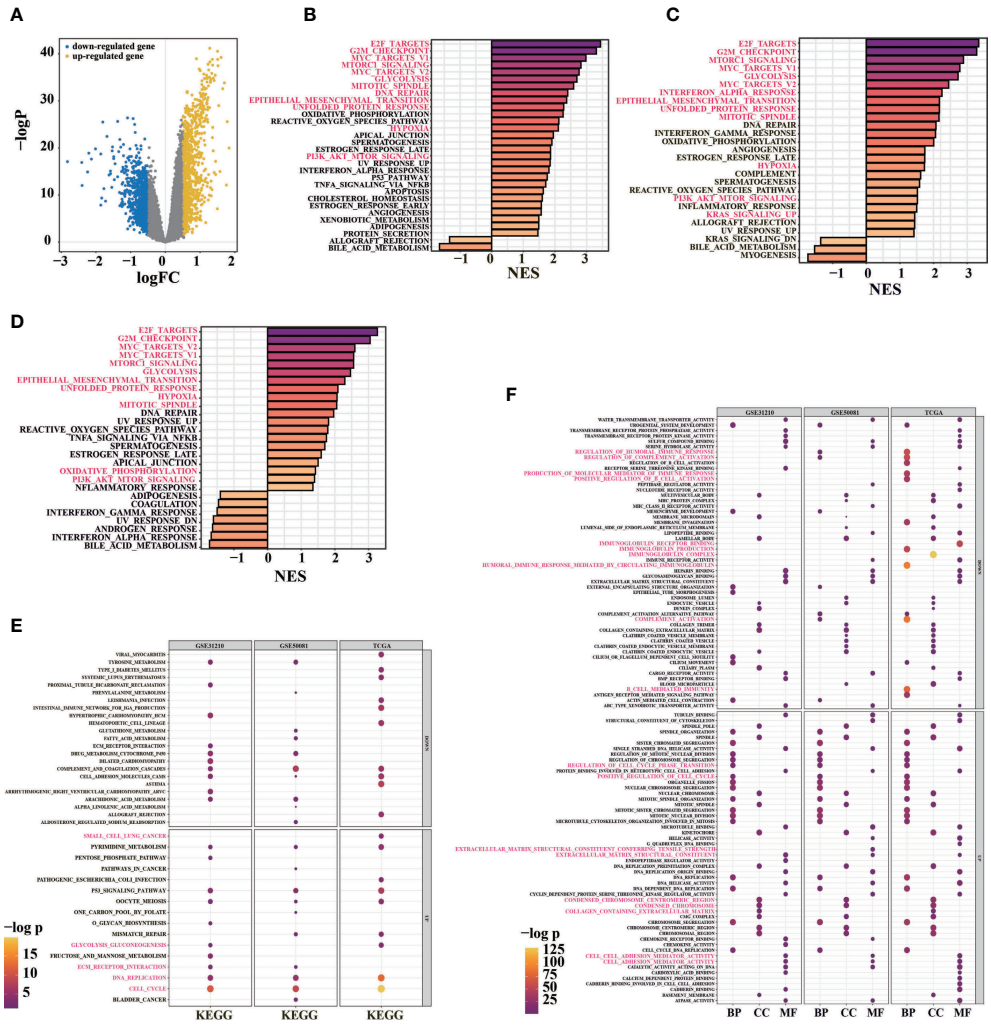


FIGURE 4
Functional enrichment analysis for prognosis signature. **(A)** volcano diagram show DEGs between high-risk and low-risk groups in TCGA-LUAD primary cancer samples. **(B)** GSEA enrichment analysis was performed using DEGs between high-risk and low-risk groups in TCGA cohort (30/50 genes were significant). **(C)** GSEA enrichment analysis was performed on the DEGs between high-risk and low-risk groups in the GSE31210 cohort (26/50 gene sets were significant). **(D)** 29/50 significant gene sets were obtained by GSEA enrichment analysis using DEGs between high-risk and low-risk groups in GSE50081 cohort (adj. pvalue< 0.05, color indicates the size of standardized enrichment score). **(E)** Based on LUAD samples in TCGA, GSE31210 and GSE50081 cohorts, the DEGs between high-risk and low-risk groups were analyzed for enrichment. Gene sets in KEGG database**(E)** and GO database**(F)** were used for enrichment analysis (FDR< 0.01; Color and dot size indicate enrichment significance).

with risk score (Figures 5F, S3B) (52). In addition, we found that *ALDOA*, *ENO1*, *YWHAG* and *SLC3A2* were positively correlated with risk scores (Figures 5F, S3C–F). Previous studies have shown that *SLC3A2* has a higher level in lung cancer (53). Besides being the marker of exosomes, *ALDOA*, ***ENO1*** and *YWHAG* can promote metastasis, invasion, activation and proliferation of lung cancer (54–56).

In addition, low-risk patients were correlated with significantly higher immune scores in TCGA cohort (Figure 5G). Immune cells, as shown in Figure 5H, were found to be significantly difference between high-risk and low-risk patients in TCGA database. By immune signatures, we discovered that low-risk patients had higher immune infiltration than high-risk patients. Assessment of immune-related gene expression profiles, the characteristic immune-related genes such as FADD were selected based on their expression patterns between LUAD and the adjacent non-tumorous samples in TCGA database. Then, we compared the expressions of those selected immune-related genes between LUAD patients with high and low risk score, it suggested that immune infiltration plays a vital role in prognosis in LUAD patients (Figure 5I).

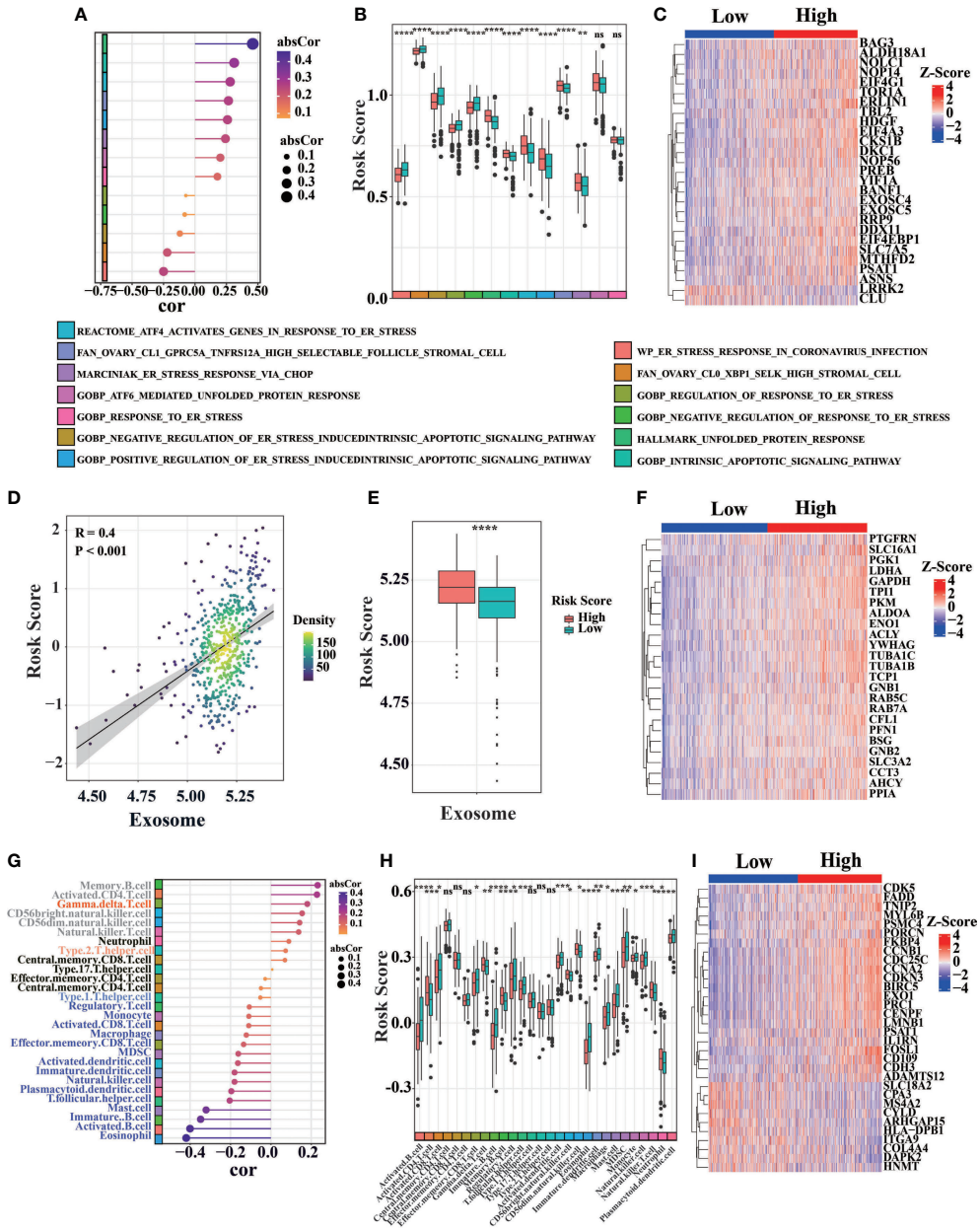


FIGURE 5
Bioprocess analysis base on prognosis signature. **(A)** Correlation diagram between risk score and ER stress-related gene set score (ssGSEA) in primary LUAD sample from TCGA cohort (n = 520). **(B)** Analyze ER stress-related gene set scores (ssGSEA) differences between high-risk and low-risk of primary LUAD samples in TCGA cohort. **(C)** Heat map of ER stress-related genes expression in LUAD patients from TCGA database. **(D–F)** The correlation between risk score and exosome associated score were analyzed similarly to the above ER stress pathway in TCGA cohort. **(G–I)** The correlation between risk score and different immune cell infiltration were analyzed similarly to the above ER stress pathway in TCGA cohort (the color block on the left represents different types of immune cells, the position indicates the magnitude of correlation coefficient, "ns" was non-significant, *P< 0.05, **P< 0.01, ***P< 0.001, ****P< 0.0001).

The prognostic signature can identify the degree of malignancy

Moreover, we found that mutation frequencies of genes with cancer development, such as TP53, significantly differed between high-risk and low-risk groups (Figure S5A). Compared to TP53, where mutations are randomly distributed, the distribution of KRAS is relatively focused on KRAS-G12 locus mutations (called hotspot, Figure S5B). The combined distribution of TP53 and KRAS-G12 mutation showed that most significant

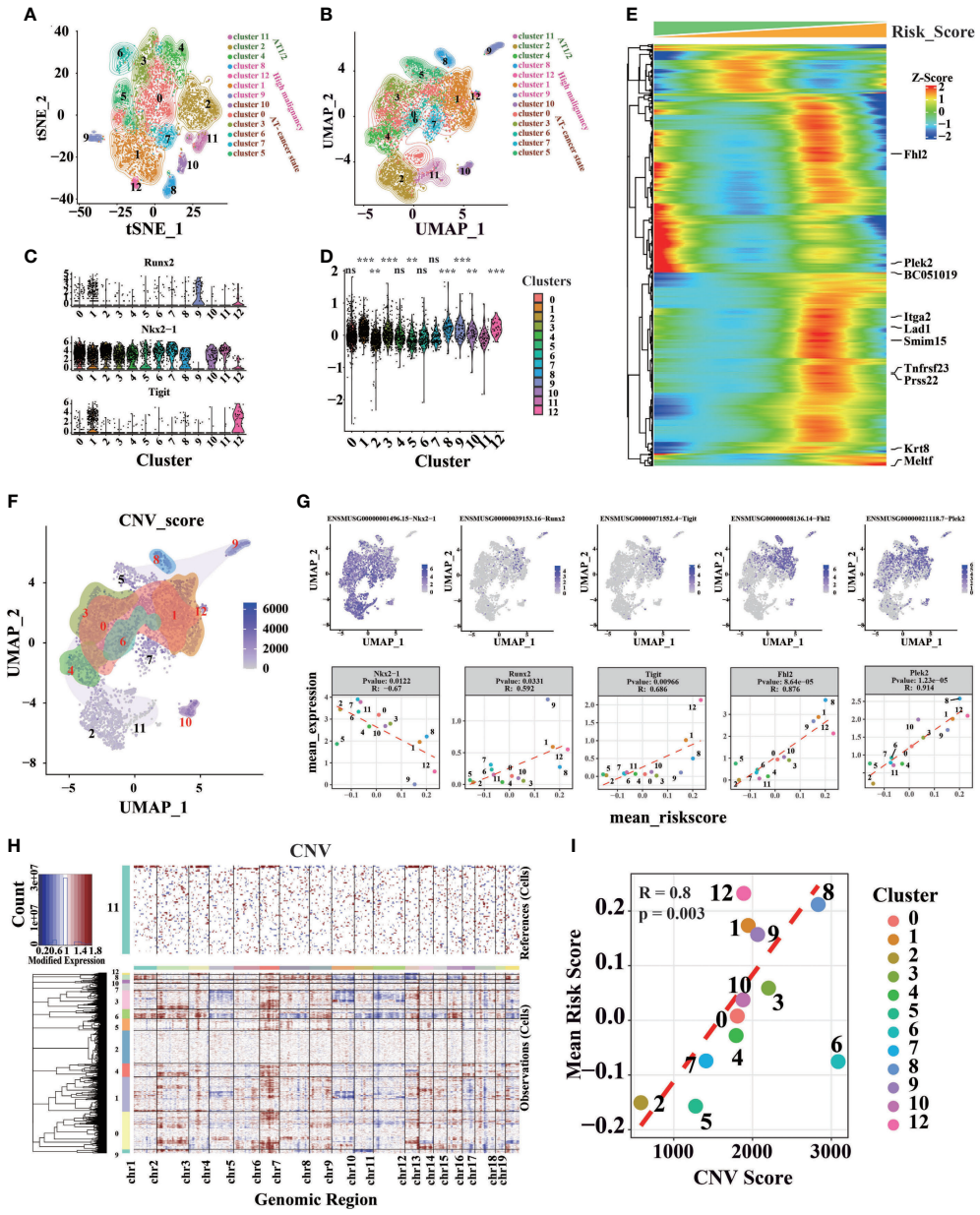


FIGURE 6
Different cell clusters identification and annotation in LUAD scRNA-seq data. **(A)** Seurat V4 was used to cluster and classify single cells, and t-SNE dimensional-reduction scatter diagram of sequencing data. **(B)** Sequencing data UMAP dimensional-reduction point scatter diagram. **(C)** Expression levels of genes associated with lung cancer development in different clusters. **(D)** The distribution of risk scores calculated by the prognostic signature in different cell clusters ("ns" is non-significant compared with cluster 11, *P< 0.05, **P< 0.01, ***P< 0.001, ****P< 0.0001). **(E)** genes associated with risk score (P.value< 10^{-25}, the genes marked on the right are the 10 genes with the strongest correlation). **(F)** Distribution of CNV in UMAP dimensionality reduction map. **(G)** Genes associated with risk scores in scRNA-seq data. **(H)** InferCNV was used to predict the CNV of cells, and heat maps of the distribution of CNV of different chromatin of different cells. **(I)** Average CNV score and risk score points from different clusters are scattered figure.

difference between high-risk and low-risk groups (Figure S5C). In addition, 10 genes of the prognostic signature may have TP53 binding sites in their untranslated regions (UTR, upstream and downstream 1 kb of a gene), indicating that TP53 mutations might directly affect transcription of these genes (Figure S5D). Therefore, GEM with Kras-G12D with or without TP53 deletion were utilized to evaluate the LUAD prognostic signature.

Three thousand eight hundred ninety-one high-quality, full-length single-cell transcriptomes from 39 KrasG12D/+Trp53-mutation mice, at 8 distinct LUAD evolution stages starting with

normal alveolar type 2 cells (AT2) and ending with fully formed LUAD, were downloaded from GSE154989. Additionally, Seurat V4 was used for clustering cells (57) (Figures 6A, B). Stemness-related genes (*Tight*, *Runx2*, and *NKX2-1*) revealed that cells in cluster 1, 9, and 12 have high stemness (Figures 6C, S6A–C). After calculate risk score of different clusters (Figure 6D), we found cluster 1, 9, and 12 had higher risk scores, which was consistent with stemness-related genes in different clusters. Furthermore, the correlation between risk score and tumor progression is positive (Figures S6C–F).

However, we found that cluster 3, 8, and 10 also had higher risk scores than others (Figures 6D, S6G–J). Therefore, feature genes were identified to evaluate the model. Cluster 3 has a higher *Vcan* and *Dcpp1*, which involve cell adhesion, proliferation, migration, and angiogenesis, and plays a central role in tissue morphogenesis and maintenance (58, 59). Cluster 10 has a higher level of *Ccnb1* and *Top2a*, which related to cell cycle and cell proliferation (60, 61). Cluster 8 has a higher level of *Rn7sk*, which was associated with Gastric Cancer (62) (Figure S6A). *Fhl2*, *Plek2*, and other cancer-related genes have a higher correlation coefficient with risk score (63–65) (Figure 6E). CNV was estimated using scRNA-seq data (Figure 6F) in contrast to cluster 11, as cluster 11 only appeared in the early stage of tumor progression (Figures S6B, C). Each cluster's average risk score is associated with the average CNV score; for instance, malignant cells usually have serious CNV (Figures 6G–I). On the other hand, we found that CNV was higher in KP mice after 12 weeks (Figures S6K–M). It shows that the intracellular CNV accumulates with tumor progression, and tumor with higher CNV indicates a higher degree of malignancy, consistent with these results described through the risk score. Therefore, cells with higher risk scores have higher CNV and a higher degree of malignancy.

Validated the biological process of prognostic model by scRNA-seq

To proving the details of risk score in tumor microenvironment, AUCell calculated the score in ER stress and exosome-related gene sets by scRNA-seq data. As expected, ER stress scores (Figure 7A) and exosome scores (Figure 7B) were higher in cluster 1, 9, and 12, which have higher risk scores. There is also a high consistency between exosome score and risk score (Figure 7C), correlating with our earlier findings.

Functional enrichment was performed in different clusters with hallmark gene sets to evaluate each cell's biological process (Figure 7D). Cluster 10 has a higher activation of E2F targets and G2M checkpoint, leading to higher risk scores. Cluster 1, 8, 9, 10, and 12 have a higher level of PI3K-AKT-mTOR, which could lead to increased risk scores. Interestingly, cluster 11, which only appeared in the early stage of tumor progression, had higher risk scores than cluster 2, 5, and 6. It suggests that immune response in the early stage of LUAD could promote tumor progression (66) (Figure 7D). Cluster 8 and cluster 12, which have high-risk scores and mostly appeared in the middle stages of LUAD (12w and 18w, Figures S6C–F), showed more cellular response (such as IL2, IL6), and the abilities were lost with tumor progression, which could be related to the tumor progression (Figure 7D).

Tumor progression can be described by risk score

To explore the relationship between risk score and tumor progression, LUAD marker genes and lung epithelial marker genes were calculated (as "bioscore") through UMAP to visualize tumor progression and compared with risk scores (Figure S7A). We found that LUAD-related markers were concentrated in high-risk regions, while more lung epithelial marker genes were concentrated in low-risk regions. The scRNA-seq data of different genotypes and experimental time were calculated using dimensionality reduction by UMAP. It was found that high-risk cells were concentrated in the region of KRAS and Trp53 mutations and mainly distributed in the late stage of the experiment. Meanwhile, cluster 1, 8, 9, and 12 were primarily distributed in the high-risk region. According to the pseudotime of Monocle V3, the region is also the accumulation of advanced LUAD cells (Figure S7B).

Furthermore, we found that risk score was positively correlated with most LUAD marker genes and negatively correlated with lung epithelial marker genes (Figure S7C). These results indicate that risk score can describe the development process of lung epithelial to LUAD cells.

For more details of tumor progression, high variable genes characterized by monocle V2 were used for estimating pseudotime. Cluster 11, which appeared in the early stage of LUAD, was set to the initial point of pseudotime (Figure 8A1). The minimum spanning tree, "Stemness" related genes downloaded from 'Msigdb', can also estimate pseudotime correctly (Figure 8A2). With tumor progression, malignant cells with strong stemness appear in LUAD, which have the potential for multi-directional differentiation, leading to intratumoral heterogeneity (67, 68). Therefore, LUAD progression should be considered in stemness levels by "Stemness" related genes (Figures S9, S10B). *Igfbp5*, *Ros1* (54–56), and other genes related to tumor progression profoundly correlate with pseudotime (Figure 8B–D), meaning the minimum spanning tree and pseudotime can describe tumor progression in more detail. With the increase of pseudotime, the malignant degree of tumor and the proportion of high-risk cells also increased, indicating that the risk score model can reflect more advanced tumor progression (Figure 8A3). Notably, seven states have been classified by a minimum spanning tree (Figure 8A4). Previously, we found that cluster 1, cluster 8, cluster 9, and cluster 12 have higher risk scores. According to the clusters, states 6 and 7, located under branch 1, also have high-risk scores (Figures 8E–G and S8A).

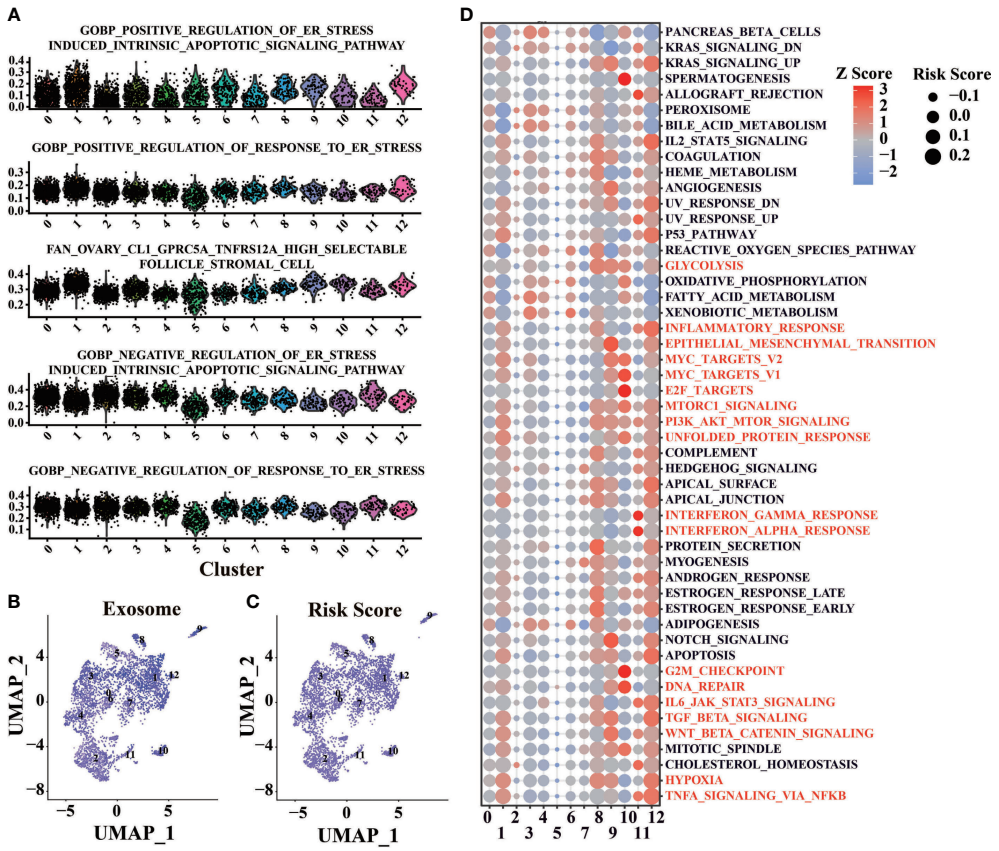


FIGURE 7
Functional enrichment of scRNA-seq base on the prognosis signature. **(A)** Enrichment analysis of ER stress-related gene sets was conducted by AUCell software. **(B)** Enrichment analysis of exosome related gene sets was conducted by AUCell software. **(C)** Use the prognostic signature to calculate the risk score of cells. **(D)** Enrichment analysis of tumor-associated Helmark gene sets was performed using AUCell.

More pseudotime-related genes and branch-related genes were recognized (Figures 8B, S8C–E). Pseudotime related genes *Msln* and *Slc4a11* can promote tumor progression (69, 70). Additionally, branch-related genes *Sftpc* and *Lyz2*, known as AT2 markers (19, 57), are shown as typical examples (Figures 8H–K, S10A). The genes can reflect tumor malignant degree and further describe the similarity and branching relationship in tumor clusters. So, we can infer that the AT2 cell features were missing with the cancer progress, accompanied by increased LUAD features (Figure S7). As we expect, the mean pseudotime and risk score of different states have a severe positive correlation. Interestingly, the mean expression of pseudotime-related genes and branch-related genes is used to clearly different states more clearly, showing some key points during tumor progression. Furthermore, a risk score can help us estimate the malignant of each state.

Discussion

LUAD is a complex disease in which multiple pathways are involved in pathogenesis. Exploring a novel accurately prognostic biomarkers would help select patients for adjuvant chemotherapy and improve prognosis in early-stage lung cancer. We build a prognostic model in TCGA cohort by LASSO regression, which has excellent specificity and sensitivity for predicting OS. To identify whether risk score as an independent predictor of survival time, univariate and multivariable cox proportional hazards regression was analyzed in LUAD patients. After controlling for confounding variables (including age, gender, invasion depth, distant metastasis, lymph node metastasis, and TNM stage), the model remained statistically significant for OS. The model can be an independent factor with better predictive potential than the pathological stage alone.

Several factors have been proved to be related to tumor progression; however, only hypoxia and glycolysis were highly associated with LUAD prognosis (71, 72). Here, we found risk score was related to several biological processes like epithelial-mesenchymal transition, ECM, glycolysis, and proliferation, which were involved in tumor progression. The prognostic model could reveal the critical elements involved in tumor microenvironments. More details about ER stress, exosome, and immune response,

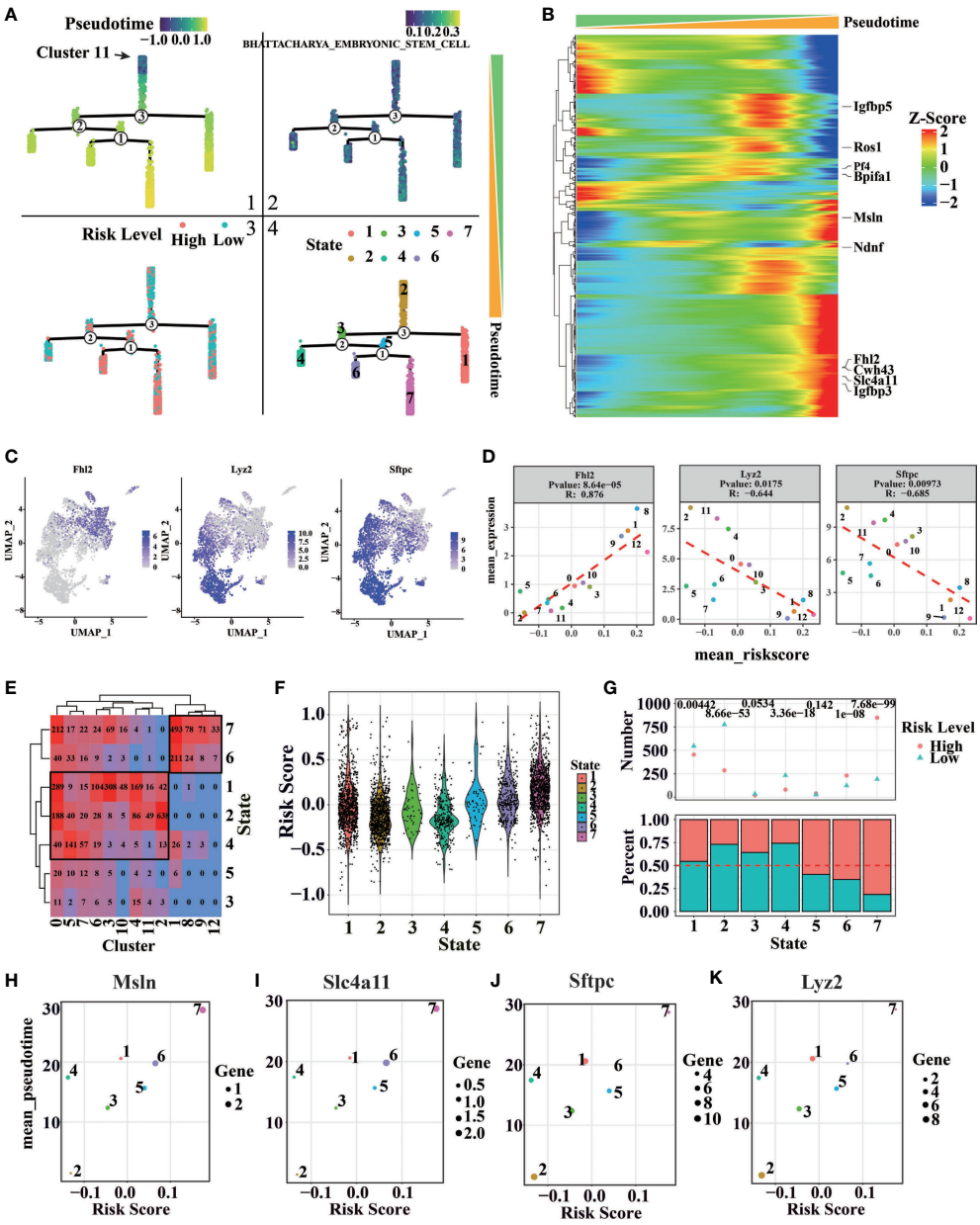


FIGURE 8
Cell trajectory and pseudotime analysis of LUAD by scRNA-seq. (A1) Distribution of pseudo-time values of different cells in the minimum spanning tree. (A2) The distribution of BHATTACHARYA_EMBRYONIC_STEM_CELL score in the minimum spanning tree. (A3) Distribution of risk rating in the minimum spanning tree. (A4) Distribution of different cell states in the minimum spanning tree. **(B)** Genes related to pseudo-time series, the genes marked on the right are the 10 genes with the strongest correlation. **(C, D)** Distribution and correlation of pseudo-time related genes and branching related genes. **(E)** The relationship between cell states and clusters in different pseudo-time series, the number represents the number of cells in the corresponding state. **(F)** Cell risk values corresponding to different quasi-sequential states. **(G)** Ratio of cell risk states corresponding to different pseudo temporal states (above: the top number represents the significance of binomial distribution test; below: the ratio of high-risk cells to low risk cells in each state). **(H)** Scatter plot of *Msln* expression mean and risk mean of cells in different states. **(I)** Scatter plot of *Slc4a11* expression and risk score of cells in different states. **(J)** Scatter plot of *Sftpc* expression and risk score of cells in different states. **(K)** Scatter plot of *Lyz2* expression average and risk average of cells in different states.

which play essential roles in tumor progression, were researched, and we found that risk score relates to these biological phenomena. The prognostic differences between high and low-risk groups could be explained by the critical elements involved in tumor microenvironments (71, 72). Recently, cancer-released exosomes could modify the distant microenvironment to a pre-metastatic niche to facilitate the formation of metastatic lesions, suggesting that cancer exosomes could result in both local and distant effects (73, 74), including malignant transformation, angiogenesis, immunosuppression, invasion, and treatment resistance (51, 75). The unfavorable intratumoral microenvironment, such as nutrient deficiency, hypoxia, high metabolic demand, oxidative stress, and unfolded protein response, can also ultimately induce ER stress, enhancing tumorigenicity, metastasis, and tumor drug resistance. It is reported that ER stress regulates proliferation, migration, and invasion through active c-Myc signaling and PI3K/AKT/mTOR signaling pathways, and mediated anti-tumor immune responses by inducing immunosuppressive microenvironment (76–78). We found that low-risk patients correlated with a significantly higher immune score in TCGA cohort, and immune cell populations are different between high-risk and low-risk patients in TCGA database. Therefore, risk score could be used to evaluate immune infiltration in LUAD patients, which may be useful for immune-targeted tumor therapy. It is well known that tumor microenvironment can be classified into two immunophenotypes based on their degree of immune infiltrations, hot tumors with high immune infiltration and cold tumors with low immune cell infiltration (79, 80). Low-risk patients with a high immune score suggest the presence of a hot tumor microenviroment, those patients could benefit more from immune-targeted therapy than high-risk patients with cold tumor microenviroment. However, further studies are still needed to confirm the prognostic value of risk score in determining hot/cold tumors.

By previous studies, we found some prognostic signatures may have poor repeatability due to insufficient sample size (81, 82). One limitation of this study is the relatively small sample and the poor quality of the GEO cohort, which was used to verify the prognostic signature. To address this issue, we have used multiple GEO cohorts to further verify the prediction performance of the prognostic signature. Further, tumor progression can also be influenced by complex and dynamic features in tumor surroundings, which means a model based on several gene sets may lead to bias. To develop a good prediction model for OS, GEM were used to test the model in different dimension through scRNA-seq data. By single-cell clustering, stemness-related genes (*Tight*, *Runx2*, and *NKX2-1*) revealed that cells in cluster 1, 9, and 12 have high stemness. Tight, a marker of high-plasticity cell, shows high proliferative potential and can be induced chemoresistance (57). *Runx2* can drive the metastatic phenotype in the primary tumors, and *NKX2-1* also shows the same consequence (83–85). Furthermore, the correlation between risk score and tumor progression is positive. According to tumor-related feature genes, we found that clusters 3, 8, and 10 also have higher risk scores than other clusters. Cluster 3 was involved in cell adhesion, proliferation, migration, and angiogenesis and played a central role in tissue morphogenesis and maintenance (58, 59). Cluster 10 has a higher level of cell cycle and cell proliferation (60, 61), and cluster 8 was related to Gastric Cancer (62). In conclusion, we found that the higher risk score, the higher degree of malignancy. In addition, risk score in each cluster is associated with average CNV score, such as malignant clusters usually have serious CNV. However, cluster 6 has a lower risk score and higher CNV level, which might be induced by intratumoral heterogeneity.

Furthermore, we found that risk score was related to tumor progression calculated by Monocle. Minimum spanning tree has been applied to describe LUAD progression, which was confirmed by enrichment analysis on 'BHATTACHARYA EMBRYONIC STEM CELL' collected from Msigdb (86). With the advancement of LUAD, risk score and the proportion of high-risk cells were increasing. We also notice that cells under branch2 (cells in stat6 and stat7) show higher risk scores. Moreover, we found several AT2 marker genes (*Sftpc* and *Lyz2*) at the branch2, which means that branch2 is a crucial point for LUAD progression. As we expect, the pseudotime and risk score of different states have a severe positive correlation. Interestingly, pseudotime-related genes and branch-related genes is used to clearly different states more clearly, showing some key points during tumor progression. Overall, the risk score was in accordance with the grade malignancy in each cluster, which was annotated by feature genes, pathways related to tumor progress, CNVs and genotype, and growth time of GEM. Based on this, our follow-up research will focus on clinical application and molecular mechanisms.

In conclusion, the study established a risk scoring model, which can be used as an independent prognostic signature to accurately evaluate the prognosis of LUAD patients. Compared with traditional clinical indicators, the model has higher accuracy and stability, and can provide guidance for follow-up treatment. The prognostic signature related to several biological processes, which may reveal the key molecular mechanisms in tumor development.

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

ZL, TZ and WY conceived and designed the study. ZL drafted the entire manuscript. TZ is responsible for LUAD

patient data acquisition and bioinformatics support. ZL and CZ participated in figures and table preparation. WY and YC reviewed and revised the manuscript. TZ and CZ guided the selection of statistical methods. WY and YC supervised the project. All authors contributed to the article and approved the submitted version.

Funding

This project was sponsored by the Natural Science Fund of China (81673462, 82073916, 81874452, and 91540119) and the Key development project of Jiangsu Province (BE2017712).

Acknowledgments

The authors thank AiMi Academic Services (www.aimieditor.com) for the English language editing and review services. In addition, ZL would like to thank Lan Zhang for the ongoing and unwavering support, patience, and love.

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.2022.1019442/full#supplementary-material

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin* (2022) 72(1):7–33. doi: 10.3322/caac.21708

2. Shi J, Hua X, Zhu B, Ravichandran S, Wang M, Nguyen C, et al. Somatic genomics and clinical features of lung adenocarcinoma: A retrospective study. *PloS Med* (2016) 13(12):e1002162. doi: 10.1371/journal.pmed.1002162

3. Goldstraw P, Crowley J, Chansky K, Giroux DJ, Groome PA, Rami-Porta R, et al. The iaslc lung cancer staging project: Proposals for the revision of the tnm stage groupings in the forthcoming (Seventh) edition of the tnm classification of malignant tumours. *J Thorac Oncol* (2007) 2(8):706–14. doi: 10.1097/JTO.0b013e31812f3c1a

4. Guerrera F, Errico L, Evangelista A, Filosso PL, Ruffini E, Lisi E, et al. Exploring stage I non-Small-Cell lung cancer: Development of a prognostic model predicting 5-year survival after surgical resection†. *Eur J Cardiothorac Surg* (2015) 47(6):1037–43. doi: 10.1093/ejcts/ezu410

5. Zhu CQ, Pintilie M, John T, Strumpf D, Shepherd FA, Der SD, et al. Understanding prognostic gene expression signatures in lung cancer. *Clin Lung Cancer* (2009) 10(5):331–40. doi: 10.3816/CLC.2009.n.045

6. Subramanian J, Simon R. Gene expression-based prognostic signatures in lung cancer: Ready for clinical use? *J Natl Cancer Inst* (2010) 102(7):464–74. doi: 10.1093/jnci/djq025

7. Li Y, Sun N, Lu Z, Sun S, Huang J, Chen Z, et al. Prognostic alternative mrna splicing signature in non-small cell lung cancer. *Cancer Lett* (2017) 393:40–51. doi: 10.1016/j.canlet.2017.02.016

8. Kadara H, Behrens C, Yuan P, Solis L, Liu D, Gu X, et al. A five-gene and corresponding protein signature for stage-I lung adenocarcinoma prognosis. *Clin Cancer Res* (2011) 17(6):1490–501. doi: 10.1158/1078-0432.Ccr-10-2703

9. Lu Y, Lemon W, Liu PY, Yi Y, Morrison C, Yang P, et al. A gene expression signature predicts survival of patients with stage I non-small cell lung cancer. *PloS Med* (2006) 3(12):e467. doi: 10.1371/journal.pmed.0030467

10. Genome Atlas Research Network. Comprehensive Molecular Profiling of Lung Adenocarcinoma. *Nature* (2014) 511(7511):543–50. doi: 10.1038/nature13385

11. Larsen JE, Minna JD. Molecular biology of lung cancer: Clinical implications. *Clin Chest Med* (2011) 32(4):703–40. doi: 10.1016/j.ccm.2011.08.003

12. Popper HH. Progression and metastasis of lung cancer. *Cancer Metastasis Rev* (2016) 35(1):75–91. doi: 10.1007/s10555-016-9618-0

13. Sutherland KD, Song JY, Kwon MC, Proost N, Zevenhoven J, Berns A. Multiple cells-of-Origin of mutant K-Ras-Induced mouse lung adenocarcinoma. *Proc Natl Acad Sci U.S.A.* (2014) 111(13):4952–7. doi: 10.1073/pnas.1319963111

14. Desai TJ, Brownfield DG, Krasnow MA. Alveolar progenitor and stem cells in lung development, renewal and cancer. *Nature* (2014) 507(7491):190–4. doi: 10.1038/nature12930

15. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, et al. Single-cell rna-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* (2014) 344(6190):1396–401. doi: 10.1126/science.1254257

16. Jackson EL, Olive KP, Tuveson DA, Bronson R, Crowley D, Brown M, et al. The differential effects of mutant P53 alleles on advanced murine lung cancer. *Cancer Res* (2005) 65(22):10280–8. doi: 10.1158/0008-5472.Can-05-2193

17. Jackson EL, Willis N, Mercer K, Bronson RT, Crowley D, Montoya R, et al. Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev* (2001) 15(24):3243–8. doi: 10.1101/gad.943001

18. Tirosh I, Venteicher AS, Hebert C, Escalante LE, Patel AP, Yizhak K, et al. Single-cell rna-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* (2016) 539(7628):309–13. doi: 10.1038/nature20123

19. Nabhan AN, Brownfield DG, Harbury PB, Krasnow MA, Desai TJ. Single-cell wnt signaling niches maintain stemness of alveolar type 2 cells. *Science* (2018) 359(6380):1118–23. doi: 10.1126/science.aam6603

20. Tirosh I, Izar B, Prakadan SM, Wadsworth MH2nd, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell rna-seq. *Science* (2016) 352(6282):189–96. doi: 10.1126/science.aad0501

21. Mounir M, Lucchetta M, Silva TC, Olsen C, Bontempi G, Chen X, et al. New functionalities in the tcgabiolinks package for the study and integration of cancer data from gdc and gtex. *PloS Comput Biol* (2019) 15(3):e1006701. doi: 10.1371/journal.pcbi.1006701

22. Davis S, Meltzer PS. Geoquery: A bridge between the gene expression omnibus (Geo) and bioconductor. *Bioinformatics* (2007) 23(14):1846–7. doi: 10.1093/bioinformatics/btm254

23. Carlson M, Falcon S, Pages H, Li N. Hgu133plus2. db: Affymetrix human genome U133 plus 2.0 array annotation data (Chip Hgu133plus2). *R Package version 323* (2016) 3(3). Available at: https://bioconductor.org/packages/release/data/annotation/html/hgu133plus2.db.html

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

25. Lin H, Zelterman D. Modeling survival data: Extending the cox model. *Technometrics* (2000) 44(1):85–6. doi: 10.1198/tech.2002.s656

26. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization paths for cox's proportional hazards model *Via* coordinate descent. *J Stat Softw* (2011) 39(5):1. doi: 10.18637/jss.v039.i05

27. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med* (2013) 32(30):5381–97. doi: 10.1002/sim.5958

28. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (Msigdb) 3.0. *Bioinformatics* (2011) 27 (12):1739–40. doi: 10.1093/bioinformatics/btr260

29. Keerthikumar S, Chisanga D, Ariyaratne D, Al Saffar H, Anand S, Zhao K, et al. Exocarta: A web-based compendium of exosomal cargo. *J Mol Biol* (2016) 428 (4):688–92. doi: 10.1016/j.jmb.2015.09.019

30. Mathivanan S, Simpson RJ. Exocarta: A compendium of exosomal proteins and RNA. *Proteomics* (2009) 9(21):4997–5000. doi: 10.1002/pmic.200900351

31. Hänzelmann S, Castelo R, Guinney J. Gsva: Gene set variation analysis for microarray and rna-seq data. *BMC Bioinf* (2013) 14(1):1–15. doi: 10.1186/1471-2105-14-7

32. Yu G, Wang L-G, Han Y, He Q-Y. Clusterprofiler: An r package for comparing biological themes among gene clusters. *Omics* (2012) 16(5):284–7. doi: 10.1089/omi.2011.0118

33. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* (2015) 160(1-2):48–61. doi: 10.1016/j.cell.2014.12.033

34. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic kras-driven cancers require Tbk1. *Nature* (2009) 462(7269):108–12. doi: 10.1038/nature08460

35. Hao Y, Hao S, Andersen-Nissen E, Mauck Iii WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell* (2021) 184(13):3573–87. doi: 10.1016/j.cell.2021.04.048

36. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck Iii WM, et al. Comprehensive integration of single-cell data. *Cell* (2019) 177(7):1888–902. doi: 10.1016/j.cell.2019.05.031

37. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* (2015) 33(5):495–502. doi: 10.1038/nbt.3192

38. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* (2015) 161(5):1202–14. doi: 10.1016/j.cell.2015.05.002

39. Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, et al. Scenic: Single-cell regulatory network inference and clustering. *Nat Methods* (2017) 14(11):1083–6. doi: 10.1038/nmeth.4463

40. Tickle T, Tirosh I, Georgescu C, Brown M, Haas B. *Infercnv of the trinity ctat project*. Cambridge, Ma, USA: Klarman Cell Observatory, Broad Institute of Mit and Harvard (2019). Available at: https://github.com/broadinstitute/inferCNV.

41. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* (2014) 32(4):381–6. doi: 10.1038/nbt.2859

42. Kim EY, Kim A, Kim SK, Kim HJ, Chang J, Ahn CM, et al. Inhibition of Mtorc1 induces loss of e-cadherin through Akt/Gsk-3β signaling-mediated upregulation of e-cadherin repressor complexes in non-small cell lung cancer cells. *Resp Res* (2014) 15(1):26. doi: 10.1186/1465-9921-15-26

43. Tulchinsky E, Demidov O, Kriajevska M, Barlev NA, Imyanitov E. Emt: A mechanism for escape from egfr-targeted therapy in lung cancer. *BBA-Rev Cancer* (2019) 1871(1):29–39. doi: 10.1016/j.bbcan.2018.10.003

44. Baudino TA, McKay C, Pendeville-Samain H, Nilsson JA, Maclean KH, White EL, et al. C-myc is essential for vasculogenesis and angiogenesis during development and tumor progression. *Genes Dev* (2002) 16(19):2530–43. doi: 10.1101/gad.1024602

45. De Bock K, Georgiadou M, Schoors S, Kuchnio A, Wong Brian W, Cantelmo Anna R, et al. Role of Pfkfb3-driven glycolysis in vessel sprouting. *Cell* (2013) 154(3):651–63. doi: 10.1016/j.cell.2013.06.037

46. He G, Peng X, Wei S, Yang S, Li X, Huang M, et al. Exosomes in the hypoxic tme: From release, uptake and biofunctions to clinical applications. *Mol Cancer* (2022) 21(1):19. doi: 10.1186/s12943-021-01440-5

47. Tsukumo Y, Tsukahara S, Furuno A, Iemura S, Natsume T, Tomida A. Tbl2 is a novel perk-binding protein that modulates stress-signaling and cell survival during endoplasmic reticulum stress. *PloS One* (2014) 9(11):e112761. doi: 10.1371/journal.pone.0112761

48. Yamaguchi S, Ishihara H, Yamada T, Tamura A, Usui M, Tominaga R, et al. Atf4-mediated induction of 4e-Bp1 contributes to pancreatic beta cell survival under endoplasmic reticulum stress. *Cell Metab* (2008) 7(3):269–76. doi: 10.1016/j.cmet.2008.01.008

49. Martindale JL, Holbrook NJ. Cellular response to oxidative stress: Signaling for suicide and survival. *J Cell Physiol* (2002) 192(1):1–15. doi: 10.1002/jcp.10119

50. Kalluri R. The biology and function of exosomes in cancer. *J Clin Invest* (2016) 126(4):1208–15. doi: 10.1172/JCI81135

51. Muller L, Simms P, Hong C-S, Nishimura MI, Jackson EK, Watkins SC, et al. Human tumor-derived exosomes (Tex) regulate treg functions *Via* cell surface signaling rather than uptake mechanisms. *OncoImmunology* (2017) 6(8):e1261243. doi: 10.1080/2162402X.2016.1261243

52. Dooley K, McConnell RE, Xu K, Lewis ND, Haupt S, Youniss MR, et al. A versatile platform for generating engineered extracellular vesicles with defined therapeutic properties. *Mol Ther* (2021) 29(5):1729–43. doi: 10.1016/j.ymthe.2021.01.020

53. Pan D, Chen J, Feng C, Wu W, Wang Y, Tong J, et al. Preferential localization of Muc1 glycoprotein in exosomes secreted by non-small cell lung carcinoma cells. *Int J Mol Sci* (2019) 20(2):12. doi: 10.3390/ijms20020323

54. Wang C, Xu J, Yuan D, Bai Y, Pan Y, Zhang J, et al. Exosomes carrying aldoa and Aldh3a1 from irradiated lung cancer cells enhance migration and invasion of recipients by accelerating glycolysis. *Mol Cell Biochem* (2020) 469(1-2):77–87. doi: 10.1007/s11010-020-03729-3

55. Greening DW, Ji H, Chen M, Robinson BW, Dick IM, Creaney J, et al. Secreted primary human malignant mesothelioma exosome signature reflects oncogenic cargo. *Sci Rep* (2016) 6:32643. doi: 10.1038/srep32643

56. Wen J, Ma Z, Livingston MJ, Zhang W, Yuan Y, Guo C, et al. Decreased secretion and profibrotic activity of tubular exosomes in diabetic kidney disease. *Am J Physiol Renal Physiol* (2020) 319(4):F664–f73. doi: 10.1152/ajprenal.00292.2020

57. Marjanovic ND, Hofree M, Chan JE, Canner D, Wu K, Trakala M, et al. Emergence of a high-plasticity cell state during lung cancer evolution. *Cancer Cell* (2020) 38(2):229–46. doi: 10.1016/j.ccell.2020.06.012

58. Li W, Han F, Fu M, Wang Z. High expression of vcan is an independent predictor of poor prognosis in gastric cancer. *J Int Med Res* (2020) 48 (1):0300060519891271. doi: 10.1177/0300060519891271

59. Song H, Song J, Kim YJ, Jeong HH, Min HJ, Koh SS. Dcpp1 is the mouse ortholog of human pauf that possesses functional analogy in pancreatic cancer. *Biochem Bioph Res Co* (2017) 493(4):1498–503. doi: 10.1016/j.bbrc.2017.10.015

60. Wang T, Lu J, Wang R, Cao W, Xu J. Top2a promotes proliferation and metastasis of hepatocellular carcinoma regulated by mir-144-3p. *J Cancer* (2022) 13 (2):589. doi: 10.7150/jca.64017

61. Zhang H, Zhang X, Li X, Meng WB, Bai ZT, Rui SZ, et al. Effect of Ccnb1 silencing on cell cycle, senescence, and apoptosis through the P53 signaling pathway in pancreatic cancer. *J Cell Physiol* (2019) 234(1):619–31. doi: 10.1002/jcp.26816

62. Song H, Sun W, Ye G, Ding X, Liu Z, Zhang S, et al. Long non-coding rna expression profile in human gastric cancer and its clinical significances. *J Transl Med* (2013) 11(1):1–10. doi: 10.1186/1479-5876-11-225

63. Liu Y, Yang S, Wang F, Zhou Z, Xu W, Xie J, et al. Plek2 promotes osteosarcoma tumorigenesis and metastasis by activating the Pi3k/Akt signaling pathway. *Oncol Lett* (2021) 22(1):1–9. doi: 10.3892/ol.2021.12795

64. Algaber A, Madhi R, Hawez A, Rönnow C-F, Rahman M. Targeting Fhl2–E–Cadherin axis by Mir–340–5p attenuates colon cancer cell migration and invasion. *Oncol Lett* (2021) 22(2):1–11. doi: 10.3892/ol.2021.12898

65. Shen H, He M, Lin R, Zhan M, Xu S, Huang X, et al. Plek2 promotes gallbladder cancer invasion and metastasis through Egfr/Ccl2 pathway. *J Exp Clin Canc Res* (2019) 38(1):1–14. doi: 10.1186/s13046-019-1250-8

66. Casanova-Acebes M, Dalla E, Leader AM, LeBerichel J, Nikolic J, Morales BM, et al. Tissue-resident macrophages provide a pro-tumorigenic niche to early nsclc cells. *Nature* (2021) 595(7868):578–84. doi: 10.1038/s41586-021-03651-8

67. Jolly MK, Celià-Terrassa T. Dynamics of phenotypic heterogeneity associated with emt and stemness during cancer progression. *J Clin Med* (2019) 8(10):1542. doi: 10.3390/jcm8101542

68. Ho DW-H, Tsui Y-M, Sze KM-F, Chan L-K, Cheung T-T, Lee E, et al. Single-cell transcriptomics reveals the landscape of intra-tumoral heterogeneity and stemness-related subpopulations in liver cancer. *Cancer Lett* (2019) 459:176–85. doi: 10.1016/j.canlet.2019.06.002

69. Yang X, Huang M, Zhang Q, Chen J, Li J, Han Q, et al. Metformin antagonizes ovarian cancer cells malignancy through msln mediated il-6/Stat3 signaling. *Cell Transplant* (2021) 30:09636897211027819. doi: 10.1177/09636897211027819

70. Qin L, Li T, Liu Y. High Slc4a11 expression is an independent predictor for poor overall survival in grade 3/4 serous ovarian cancer. *PloS One* (2017) 12(11): e0187385. doi: 10.1371/journal.pone.0187385

71. Lambrechts D, Wauters E, Boeckx B, Aibar S, Nittner D, Burton O, et al. Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat Med* (2018) 24(8):1277–89. doi: 10.1038/s41591-018-0096-5

72. Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M, et al. Understanding the tumor immune microenvironment (Time) for effective therapy. *Nat Med* (2018) 24(5):541–50. doi: 10.1038/s41591-018-0014-x

73. Liu Y, Gu Y, Han Y, Zhang Q, Jiang Z, Zhang X, et al. Tumor exosomal rnas promote lung pre-metastatic niche formation by activating alveolar epithelial Tlr3 to recruit neutrophils. *Cancer Cell* (2016) 30(2):243–56. doi: 10.1016/j.ccell.2016.06.021

74. Boelens Mirjam C, Wu Tony J, Nabet Barzin Y, Xu B, Qiu Y, Yoon T, et al. Exosome transfer from stromal to breast cancer cells regulates therapy resistance pathways. *Cell* (2014) 159(3):499–513. doi: 10.1016/j.cell.2014.09.051

75. Becker A, Thakur BK, Weiss JM, Kim HS, Peinado H, Lyden D. Extracellular vesicles in cancer: Cell-to-Cell mediators of metastasis. *Cancer Cell* (2016) 30(6):836–48. doi: 10.1016/j.ccell.2016.10.009

76. Zhang F, Liu Y, Yang Y, Yang K. Development and validation of a fourteen-innate immunity-related gene pairs signature for predicting prognosis head and neck squamous cell carcinoma. *BMC Cancer* (2020) 20(1):1–15. doi: 10.1186/s12885-020-07489-7

77. Gu X, Wang L, Boldrup L, Coates PJ, Fahraeus R, Sgaramella N, et al. Ap001056. 1, a prognosis-related enhancer rna in squamous cell carcinoma of the head and neck. *Cancers* (2019) 11(3):347. doi: 10.3390/cancers11030347

78. She Y, Kong X, Ge Y, Yin P, Liu Z, Chen J, et al. Immune-related gene signature for predicting the prognosis of head and neck squamous cell carcinoma. *Cancer Cell Int* (2020) 20(1):1–10. doi: 10.1186/s12935-020-1104-7

79. Borcoman E, de la Rochere P, Richer W, Vacher S, Chemlali W, Krucker C, et al. Inhibition of Pi3k pathway increases immune infiltrate in muscle-invasive bladder cancer. *OncoImmunology* (2019) 8(5):e1581556. doi: 10.1080/2162402X.2019.1581556

80. Liu Y-T, Sun Z-J. Turning cold tumors into hot tumors by improving T-cell infiltration. *Theranostics* (2021) 11(11):5365. doi: 10.7150/thno.58390

81. Li Y, Gu J, Xu F, Zhu Q, Chen Y, Ge D, et al. Molecular characterization, biological function, tumor microenvironment association and clinical significance of M6a regulators in lung adenocarcinoma. *Brief Bioinform* (2021) 22(4):bbaa225. doi: 10.1093/bib/bbaa225

82. Cao B, Dai W, Ma S, Wang Q, Lan M, Luo H, et al. An ev-associated gene signature correlates with hypoxic microenvironment and predicts recurrence in lung adenocarcinoma. *Mol Ther-Nucl Acids* (2019) 17:879–90. doi: 10.1016/j.omtn.2019.07.021

83. Chuang C-H, Greenside PG, Rogers ZN, Brady JJ, Yang D, Ma RK, et al. Molecular definition of a metastatic lung cancer state reveals a targetable Cd109–janus kinase–stat axis. *Nat Med* (2017) 23(3):291–300. doi: 10.1038/nm.4285

84. Snyder EL, Watanabe H, Magendantz M, Hoersch S, Chen TA, Wang DG, et al. Nkx2-1 represses a latent gastric differentiation program in lung adenocarcinoma. *Mol Cell* (2013) 50(2):185–99. doi: 10.1016/j.molcel.2013.02.018

85. Winslow MM, Dayton TL, Verhaak RGW, Kim-Kiselak C, Snyder EL, Feldser DM, et al. Suppression of lung adenocarcinoma progression by Nkx2-1. *Nature* (2011) 473(7345):101–4. doi: 10.1038/nature09881

86. Bhattacharya B, Miura T, Brandenberger R, Mejido J, Luo Y, Yang AX, et al. Gene expression in human embryonic stem cell lines: Unique molecular signature. *Blood* (2004) 103(8):2956–64. doi: 10.1182/blood-2003-09-3314

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 01 November 2022
DOI 10.3389/fonc.2022.986367

Check for updates

OPEN ACCESS

EDITED BY
Ping Zhan,
Nanjing University School of Medicine, China

REVIEWED BY
Greg L. Stewart,
University of British Columbia, Canada
Wuju Li,
Beijing Institute of Basic Medical Sciences, China
Wei-Qi Nian,
Chongqing Hospital of Traditional Chinese Medicine, China

*CORRESPONDENCE
Lianhua Ye
Lhye1204@aliyun.com
Yunchao Huang
huangych2001@aliyun.com

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

SPECIALTY SECTION
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

RECEIVED 05 July 2022
ACCEPTED 17 October 2022
PUBLISHED 01 November 2022

CITATION
Zhao J, Li G, Zhao G, Wang W, Shen Z, Yang Y, Huang Y and Ye L (2022) Prognostic signature of lipid metabolism associated LncRNAs predict prognosis and treatment of lung adenocarcinoma.
Front. Oncol. 12:986367.
doi: 10.3389/fonc.2022.986367

COPYRIGHT
© 2022 Zhao, Li, Zhao, Wang, Shen, Yang, Huang and Ye. 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.

Prognostic signature of lipid metabolism associated LncRNAs predict prognosis and treatment of lung adenocarcinoma

Jie Zhao[1†], Guangjian Li[1†], Guangqiang Zhao[1], Wei Wang[2], Zhenghai Shen[1], Yantao Yang[1], Yunchao Huang[1*] and Lianhua Ye[1*]

[1]Department of Thoracic Surgery, The Third Affiliated Hospital of Kunming Medical University (Yunnan Cancer Hospital), Kunming, China, [2]Department of Thoracic Surgery, Taihe Hospital (Hubei University of Medicine), Shiyan, China

Background: Lung adenocarcinoma (LUAD) is the most predominant histological subtype of lung cancer. Abnormal lipid metabolism is closely related to the development of LUAD. LncRNAs are involved in the regulation of various lipid metabolism-related genes in various cancer cells including LUAD. Here, we aimed to identify lipid metabolism-related lncRNAs associated with LUAD prognosis and to propose a new prognostic signature.

Methods: First, differentially expressed lncRNAs (DE-lncRNAs) from the TCGA-LUAD and the GSE31210 dataset were identified. Then the correlation analysis between DE-lncRNAs and lipid metabolism genes was performed to screen lipid metabolism-related lncRNAs. Cox regression analyses were performed in the training set to establish a prognostic model and the model was validated in the testing set and the validation set. Moreover, The role of this model in the underlying molecular mechanisms, immunotherapy, and chemotherapeutic drug sensitivity analysis was predicted by methods such as Gene Set Enrichment Analysis, immune infiltration, tumor mutational burden (TMB), neoantigen, Tumor Immune Dysfunction and Exclusion, chemosensitivity analysis between the high- and low-risk groups. The diagnostic ability of prognostic lncRNAs has also been validated. Finally, we validated the expression levels of selected prognostic lncRNAs by quantitative real-time polymerase chain reaction (qRT-PCR).

Results: The prognostic model was constructed based on four prognostic lncRNAs (LINC00857, EP300-AS1, TBX5-AS1, SNHG3) related to lipid metabolism. The receiver operating characteristic curve (ROC) and Kaplan Meier (KM) curves of the risk model showed their validity. The results of Gene Set Enrichment Analysis suggested that differentially expressed genes in high- and low-risk groups were mainly enriched in immune response and cell cycle. There statistical differences in TMB and neoantigen between high- and low-risk groups. Drug sensitivity analysis suggested that patients with low risk scores may have better chemotherapy outcomes. The results of qRT-PCR were

suggesting that compared with the normal group, the expressions of EP300-AS1 and TBX5-AS1 were down-regulated in the tumor group, while the expressions of LINC00857 and SNHG3 were up-regulated. The four prognostic lncRNAs had good diagnostic capabilities, and the overall diagnostic model of the four prognostic lncRNAs was more effective.

Conclusion: A total of 4 prognostic lncRNAs related to lipid metabolism were obtained and an effective risk model was constructed.

KEYWORDS

lung adenocarcinoma, lncRNA, prognostic, lipid metabolism genes, signature

Introduction

Lung cancer is the malignant tumor with the highest morbidity and mortality in the world. There are about 22 million new cases and 17.9 million deaths per year (1, 2). Among them, 85% are non-small cell lung cancer (NSCLC). Lung adenocarcinoma (LUAD) is the most common subtype of NSCLC, accounting for about 40% of all lung cancer subtypes, with characteristics of rapid progression, poor prognosis, and easy recurrence. In recent years, with the continuous application of molecular targeted therapy and immunotherapy, the overall survival rate (OS) has been improved to a certain extent (3). There is an urgent need to find new biomarkers that can effectively predict LUAD.

Long non-coding RNAs (lncRNAs) regulate gene expression through a variety of mechanisms, including transcriptional regulation, translation, protein modification, and activity regulation (4–8). As an important component of cellular biofilms and components, lipids are also involved in energy storage, metabolism, and cell activity signaling molecule transmission. The regulation of cellular processes such as cell growth, differentiation, inflammation, apoptosis, and drug resistance is inseparable from the extensive participation of lipid metabolism (9–11). Therefore, lipid metabolism regulation is crucial for maintaining cellular homeostasis. LncRNAs play a role in lipid metabolism through their effects on SREBP transcription factors, apolipoproteins, triglyceride metabolism, and macrophage cholesterol uptake and efflux (10, 12–15). It has been found that the lncRNA NEAT1 disrupted hepatocellular carcinoma lipolysis by regulating adipose triglyceride lipase, thereby driving hepatocellular carcinoma proliferation (16).In conclusion, LncRNAs are involved in the regulation of various lipid metabolism-related genes in cancer cells (17–21).

Studies have revealed that lncRNA MUC5B-AS1 is up-regulated in lung adenocarcinoma tissues, promotes cell migration and invasion by forming RNA-RNA duplexes with MUC5B (22). The plasma lncRNA H19 level was upregulated in LUAD patients which was correlated with clinicopathological characteristics and had a certain value in lung cancer diagnosis and could assist traditional tumor markers in lung cancer diagnosis and disease evaluation. Wang G et al. performed scRNA-seq detection on early-stage NSCLC and found that there were overall abnormalities in lipid metabolism in different cell types, of which glycerophospholipid metabolism was the most severely altered in lipid metabolism-related pathways (23). The relationship between abnormal lipid metabolism and LUAD has been confirmed by numerous studies. Phosphatidylcholine (PC) and phosphatidylethanolamine (PE) levels were significantly higher in LUAD patients than in healthy individuals (24). High density lipoprotein cholesterol (HDL-C), low density lipoprotein (LDL) and low density lipoprotein receptor (LDLR), sphinolipin, phos phatidylinositol, phosphatidylserine, phosphatidylethanolamine, phospholipid, and phosphatidylcholine are all abnormally expressed in LUAD (25–27).

At present, it is worthwhile to further study whether lipid-metabolism-related lncRNAs may be biomarkers for LUAD. To explore more effective biomarkers in LUAD and explore the potential molecular mechanism of novel lncRNAs in LUAD. In this study, based on The Cancer Genome Atlas (TCGA) database and Gene Expression Omnibus (GEO) datasets, lipid metabolism-related lncRNAs associated with LUAD prognosis were searched. A risk model was established and validated to explore the impact of the risk model on immunotherapy and chemotherapy in patients with LUAD.

Materials and methods

Data source

LUAD-related datasets were downloaded from the Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/) and the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/gds). The TCGA-LUAD dataset contains 59 normal samples and 514 cancer samples. Among these 514

cancer samples, 500 cancer samples have complete survival data and they were divided into a training set (350 samples) and a testing set (150 samples) randomly according to a ratio of 7:3. Moreover, two LUAD-related datasets (GSE31210, GSE50081) were downloaded from the GEO database. Among them, the GSE31210 dataset (containing 20 normal samples and 226 cancer samples) was used for differential analysis. The GSE50081 dataset (containing 181 cancer samples with complete survival data) was used as the validation set. The clinicopathology Characteristics of the TCGA-LUAD cohort, GSE50081 dataset and GSE31210 dataset were shown in Table 1 Lipid-specific keywords (fatty acyl, glycerolipid, glycerophospholipid, sphingolipid, sterol lipid, prenol) were searched on the Kyoto Encyclopedia of Genes and Genomes (KEGG)website(http://www.kegg.jp/blastkoala/) and the Molecular Signatures Database (MisDB) website (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp) (28).

Differential expression analysis and correlation analysis

Differential expression analysis was performed with the limma package (29). The DE-lncRNAs between LUAD samples and normal samples in the TCGA-LUAD dataset and the DE-lncRNAs between LUAD samples and normal samples in the GSE312110 dataset were analyzed. FDR< 0.05 and $|\log_2^{FC}| > 0.5$ were considered as a significant difference. The DE-lncRNAs of the two datasets were intersected to screen for common DE-lncRNAs. The correlation between lipid metabolism-related genes and common DE-lncRNAs in the TCGA dataset was calculated. Similarly, the correlation of lipid metabolism-related genes and common DE-lncRNAs in the GSE31210 dataset was also analyzed. $|cor| > 0.4$ and $P< 0.05$ was considered to have relevance.

Construction and validation of the prognostic signature

The independent prognostic Lipid metabolism-related DE-lncRNAs were screened by Cox regression analysis (univariate and multivariate) to construct the prognostic features (30). Based on the median risk score (calculated by the expression level of prognostic genes), All patients were divided into two groups(high- and low-risk). The Kaplan-Meier (KM) survival curve was plotted, and the area under curve (AUC) of the receiver operating characteristic (ROC) curve was applied to

TABLE 1 The clinicopathology Characteristics of the TCGA-LUAD cohort, GSE50081 dataset, and 110 GSE31210 dataset.

Characteristics	TCGA-LUAD cohort (N=328)	GSE50081 dataset (N=181)	GSE31210 dataset (N=226)
gender			
female	169 (51.52%)	83 (45.9%)	50 (44.2%)
male	159 (48.48%)	98 (54.1%)	105 (46.5%)
age (years)			
>=60	235 (71.65%)	154 (85.1%)	130 (57.5%)
<60	93 (28.35%)	27 (14.9%)	96 (42.5%)
M			
M0	307 (93.60%)	181 (100%)	–
M1	21 (6.40%)	0 (0.0%)	–
N-			
N0	205 (62.50%)	129 (71.3%)	–
N1	71 (21.65%)	52 (28.7%)	–
N2	51 (15.55%)	0 (0.0%)	–
N3	1 (0.30%)	0 (0.0%)	–
T			
T1	100 (30.49%)	57 (31.5%)	–
T2	184 (56.10%)	122 (67.4%)	–
T3	27 (8.23%)	2 (1.1%)	–
T4	17 (5.18%)	0 (0.0%)	–
STAGE			
Stage I	166 (50.61%)	127 (70.2%)	69 (61.1%)
Stage II	81 (24.70%)	54 (29.8%)	58 (25.7%)
Stage III	60 (18.29%)	0 (0.0%)	0 (0.0%)
Stage IV	21 (6.40%)	0 (0.0%)	0 (0.0%)

verify the predictive accuracy. Moreover, the clinical value of the signature was analyzed and the assessment and validation of the risk model were performed in the testing set and validation set.

Development and assessment of a nomogram

Cox regression analyses(univariate and multivariate) were implemented on the signature and clinical data involving age, gender, and stage (29). We constructed a prediction nomogram based on all independent predictors. And the predictivity of the nomogram was validated using ROC and calibration curves.

Analysis of potential regulatory mechanisms of prognostic lncRNAs

We calculated the correlation between prognostic lncRNA and mRNA in the TCGA dataset to explore the relevant molecular mechanisms (29). mRNAs with $|cor| > 0.7$ and $P< 0.05$ were considered as prognostic lncRNA-related mRNAs. Moreover, the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were done to reveal the potential functions of the prognostic genes. Furthermore, a protein-protein interaction network (PPI) (Confidence=0.4) was constructed using the Search Tool for Recurring instances of Neighbouring Genes (STRING) website (https://string-db.org). The survival analysis of the hub gene was carried out. Finally, the DEGs between the two groups were further functionally annotated using Gene Set Enrichment Analysis (GSEA) by Cluster Profiles package in R language.

Immune infiltration analysis and differences in response to immunotherapy

The linear support vector regression method in CIBERSORT was used to deconvolve the tumor tissue expression matrix to analyze the content of various types of immune cells in the tissue. The rank-sum test was used to analyze the differences in various immune cell contents between the two groups, and the Tumor Immune Dysfunction and Exclusion (TIDE) score, PD-L1, PD-1 distribution, T cell dysfunction score, and T cell exclusion score distribution of each sample (31).

Differences in TMB levels, neoantigen levels, and chemotherapy drug sensitivity

TMB-the total number of somatic mutations per DNA megabase (Mb) of tumor tissue. TMB per MB was calculated by dividing the total number of mutations by the size of the target coding region. The TMB values in the high- and low-risk groups were calculated separately and then the rank-sum test was carried out. The neoantigen indicators of the samples in the high- and low-groups were extracted from the TCGA database (https://gdc.cancer.gov/about-data/publications/panimmune) and performed tank-sum test. We used the pRRopheticPredict package to analyze the sensitivity of Commonly used drugs for the treatment of LUAD in the Genomics of Drug Sensitivity in Cancer (GDSC) database. Drug sensitivity is represented by the IC50 value.

Clinical tissue collection

We recruited ten LUAD patients at the Third Affiliated Hospital of Kunming Medical University and collected lung cancer tissue and paracancerous tissue samples from the patients. All participants were exempted from signing informed consent, and this study was reviewed by the Third Affiliated Hospital of Kunming Medical University ethics committee.

Diagnostic value analysis and expression validation of prognostic lncRNAs

Expression level validation of prognostic lncRNAs was performed in TCGA and GSE31210 datasets, respectively. Para-cancerous tissue and cancerous tissue samples from 10 different LUAD patients were collected, and qRT-PCR was used to verify the expression levels of prognostic genes (32). All tissue was lysed with TRIzol® reagent (Ambion by life technologies, USA, cat:356281), and total RNA was extracted following the manufacturer's instructions. The extracted RNA was reverse-transcribed to cDNA using the Script RT I First strand cDNA SynthesisAll-in-OneTM First-Strand cDNA Synthesis Kit (cat: G33330-50) before qRT-PCR. The qRT-PCR reaction consisted of 3 μl of reverse transcription product, 5 μl of 2xUniversal Blue SYBR Green qPCR Master Mix (cat: G3326-05), and 1 μl each of forward and reverse primer. PCR was performed in a BIO-RAD CFX96 Touch TM PCR detection system (Bio-Rad Laboratories, Inc., USA) under the following conditions: initial denaturation at 95°C for 1 min, followed by 40 cycles that each involved incubation at 95°C for 20 s, 55°C for 20 s, and 72°C for 30 s. The detailed forward and reverse primer is shown in Supplementary Table 9. All primers were synthesized by Servicebio (Servicebio, Wuhan, China). The GAPDH gene served as an internal control, and the relative expression of four lncRNAs was determined using the $2^{-\Delta\Delta Ct}$ method. The experiment was repeated in triplicate on independent occasions. Statistical differences in the four lncRNAs between Para-cancerous tissue and cancerous tissues samples were detected by paired t-test using

GraphPad Prism V6 (GraphPad Software, La Jolla, CA, USA), and the level of statistical significance was tested and expressed * as $P< 0.05$, ** means $P< 0.01$. Then, according to the expression levels of prognostic LncRNAs in each dataset, ROC curves of individual LncRNAs and all LncRNAs were drawn.

Statistical analysis

The statistical analyses in this study were all generated by R software. Wilcoxon test was used to perform a difference comparison between the two groups. Associations between risk scores and gene function or related pathways were calculated by Pearson correlation. Cox regression analysis was used to examine the prognostic power of prognostic features. KM survival analysis and the Cox proportional hazards model were used to analyze the association between the two risk stratifications with the R package Survival. *P*-values less than 0.05 were statistically significant.

Results

Differential expression analysis

The flowchart of the present study was displayed in Supplementary Figure 1. A total of 158 DE-lncRNAs(93 were upregulated and 65 were downregulated) were identified from the TCGA dataset(Figure 1A). The heat map of the top 100 DE-lncRNAs were shown in Supplementary Figure 2A. While a total of 206 DE-lncRNAs(97 were upregulated and 99 were downregulated) were identified from the GEO dataset (Figure 1B). The heat map of the top 100 DE-lncRNAs were shown in Supplementary Figure 2B. Finally, 50 overlapping lncRNAs were extracted (Figures 1C, D).

Identification of common differentially expressed lipid metabolism lncRNAs

A total of 1045 genes related to lipid metabolism were downloaded according to the literature (Supplementary Table 1). The expression levels of 50 common DE-lncRNAs and 1011 lipid metabolism-related genes in TCGA samples were extracted, the results of the correlation analysis are shown in Figure 2A and Supplementary Table 2. A total of 48 lipid metabolism-related DE-lncRNAs were obtained in the TCGA dataset. The correlation of 971 lipid metabolism-related genes and 50 common DE-lncRNAs in the GSE31210 dataset was also analyzed (Figure 2B and Supplementary Table 3). A total of 38 lipid metabolism-related DE-lncRNAs were obtained in the GSE31210 dataset. After taking the intersection, we obtained 38 common lipid metabolism-related DE-lncRNAs (Figure 2C).

Construction of risk feature

Univariate Cox regression analysis showed that 11 common DE-lncRNAs were significantly related to OS ($P< 0.05$, Table 2 and Figure 3A).

Subsequently, a new prognostic signature involving 4 lncRNAs (LINC00857, EP300-AS1, TBX5-AS1, SNHG3) was established after multivariate Cox regression analysis (Table 3 and Figure 3B). We calculated risk scores with the following equation: risk score = LINC00857*1.469417445 + EP300-AS1*0.757747569 + TBX5-AS1*0.660575856 + SNHG3*0.752442869. Then separated all samples into two groups depending on the median risk score, (n = 175 and 175, respectively). The patients with high-risk scores had a significantly shorter OS (Figure 3C). The AUCs were 0.639, 0.631, and 0.626 at 1, 3, and 5 years, respectively (Figure 3D), The distribution of risk scores, patient survival status, and survival time are shown in Figure 3E.

Testing and validation

The 150 LUAD samples in the testing set were separated into high- (n=75) and low-risk (n=75) groups too. The high-risk group has a worse prognosis than the low-risk group (Figure 4A). The AUCs for 1-year, 3-year, and 5-year were 0.741, 0.650, and 0.621, respectively (Figure 4B). The distribution of risk scores, patient survival status, and survival time is shown in Figures 4C. In addition, we also evaluated the correlation between the risk score and clinical traits (Supplementary Figure 2C), There is a significant difference in T stage between high- and low-risk groups (P=0.00371) (Table 4). The validation of the risk feature was done in the GSE50081 dataset. The 181 LUAD samples in the GSE50081dataset were divided into high-(n=90) and low-risk (n=91) groups. The high-risk group has a worse prognosis than the low-risk group (Figure 4D). The AUCs for 1-year, 3-year, and 5-year OS were 0.724, 0.666, and 0.651, respectively (Figure 4E). The distribution of risk scores, patient survival status, and survival time is shown in Figures 4F. There were significant differences in N, stage, and smoking status between high and low risk groups. (P=0.0291, P=0.0129, P=0.0055, Supplementary Figure 2D and Table 5). Overall, the 4 lncRNAs we detected were prognostic with both the TCGA-LUAD cohort and GSE50081 dataset.

Independent prognosis analysis

Incorporating clinicopathologic data from the TCGA-LUAD cohort into univariate Cox regression analysis, and detected that risk score, gender, M, N, T, and stage were

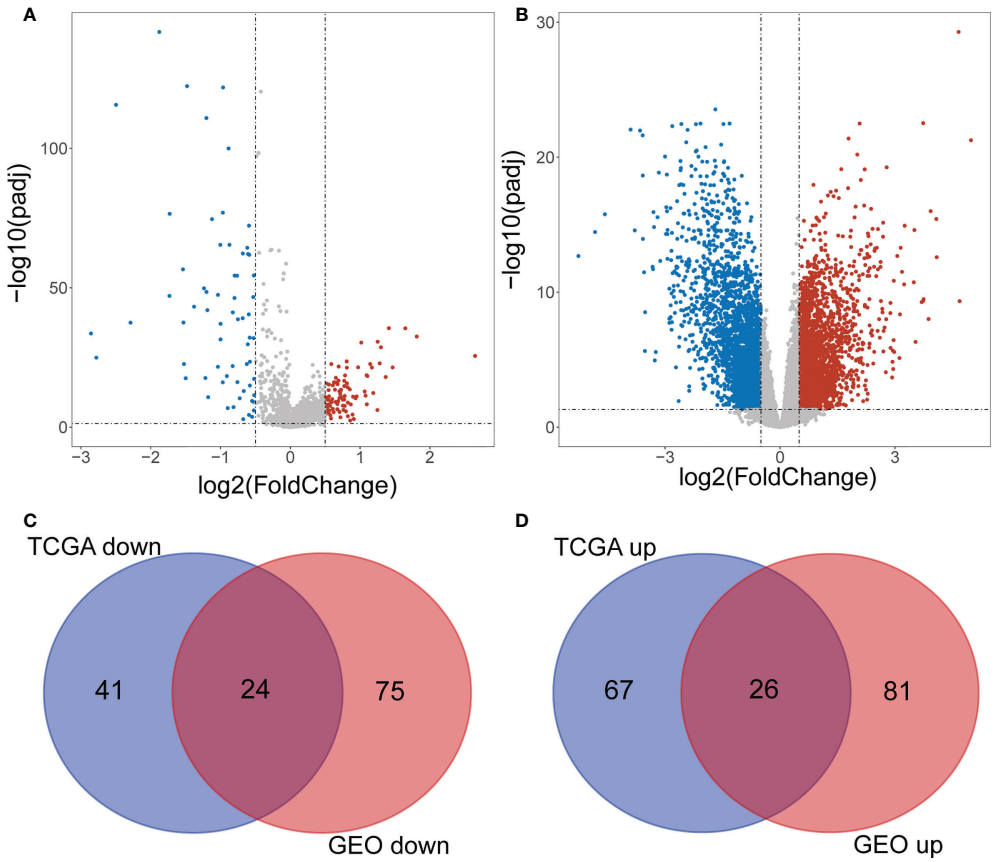


FIGURE 1
Screening differentially expressed lncRNAs. **(A)** Volcano plot of DE-lncRNAs from TCGA dataset. Red and green indicate up-regulated and down-regulated lncRNAs respectively. **(B)** Volcano plot of DE-lncRNAs from GEO dataset. Red and green indicate up-regulated and down-regulated lncRNAs respectively. **(C, D)** Venn diagram of the intersection of DE-lncRNAs.

significantly associated with LUAD prognosis (Table 6). Subsequently, enrolled these clinicopathological characteristics into multivariate Cox regression analysis and the results indicated that stage and risk score were independent prognostic indicators of LUAD (Table 7). Additionally, a nomogram was constructed to predict 1, 3, and 5 years OS of LUAD patients (Figure 5A), and the performance of the nomogram was evaluated by calibration curve and demonstrated that 1-, 3-, and 5-year predicted by this nomogram were close to the actual survival duration (Figure 5B). Moreover, univariate and multivariate Cox regression analysis were performed and noted that risk score and gender were independent prognostic factors for LUAD patients in the GSE50081 dataset (Tables 8, 9). A prediction nomogram was established and evaluated by calibration curve (Figures 5C, D), and showed good predictive accuracy. In summary, all the results indicated that the nomogram based on risk score exhibited a good predictive accuracy for the OS of LUAD patients.

Identification of mRNAs associated with prognostic lncRNAs

We also focused on the four prognostic lncRNA-associated mRNAs, but only two of them obtained associated mRNAs, a total of 171 mRNAs associated with the two prognostic lncRNAs were obtained (Tables 10, 11), and their relationships are shown in Figures 6A, B, with 9 mRNAs were both associated with prognostic lncRNAs and cis-regulated (Supplementary Figure 3). The GSEA enriched results showed that the prognostic-related lncRNA had a strong correlation with collagen-containing extracellular matrix, extracellular matrix organization, and extracellular matrix structural constituent conferring compression resistance (Figure 6C). The key enrichment pathways were “respiratory system development”, “lung development”, and lung alveolus development” (Figure 6D). Therefore, these prognostic genes may affect the prognosis of LUAD patients by regulating the structural changes of the extracellular matrix or affecting the development of the respiratory system. The detailed GSEA results are shown in

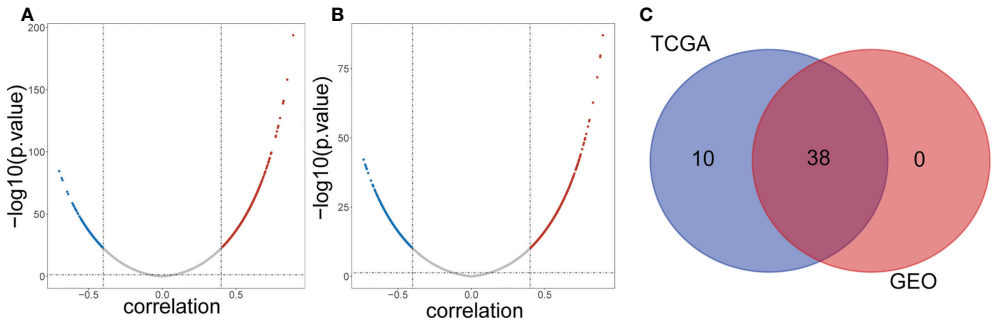


FIGURE 2
Screening lipid metabolism-related differential expressed lncRNAs. **(A)** Volcano plot of lipid metabolism-related DE-lncRNAs from TCGA dataset. Red and green indicate up-regulated and down-regulated lncRNAs respectively. **(B)** Volcano plot of lipid metabolism-related DE-lncRNAs from GEO dataset. Red and green indicate up-regulated and down-regulated lncRNAs respectively. **(C)** Venn diagram of the intersection of lipid metabolism-related DE-lncRNAs.

Supplementary Table 4 and Table 5. A 109-protein interaction network was constructed by the STRING (https://string-db.org) website (Table 12 and Figure 6E). After identification of hub genes by cytohubba function of Cytoscape arrayed the hub genes according to MCC (Supplementary Table 6). Finally, the top3 hub genes (MYH11, ELN, DCN) were obtained (Figure 6F). Moreover, the expression levels of MYH11, ELN, DCN and the corresponding survival information were extracted for survival analysis (Figures 6G–I). The results showed that patients with higher ENL gene expression had an optimistic prognosis (P=0.02295), while there was no significant difference in survival between the high and low expression groups of MYH11 and DCN.

GSEA

We performed GSEA to explore the potential mechanisms of the risk model. the enriched results showed that the prognostic differences between it may be related to the up-regulation of pathways such as DNA replication, oxidative-phosphorylation, and pyrimidine-metabolism (Figure 7A). The key enrichment pathways were “b cell-mediated immunity “,”cell cycle checkpoint”, “chromosome segregation” etc (Figure 7B). The detailed results are shown in Supplementary Tables 7 and 8. Notably, these pathways were significantly enriched in samples with high risk score.

Immune infiltration analysis and differences in response to immunotherapy

The RNA-seq data of 500 patients with LUAD from the TCGA database were analyzed to evaluate the immune landscape. Marker genes of 22 immune cell species were evaluated between the two groups (Figures 8A–C). Then, a rank-sum test was performed, and as shown in the boxplot, 10 of the 22 immune cells showed significant differences (Figure 8D). We also

TABLE 2 11 common DE-lncRNAS were identified by univariate Cox regression analysis.

id	z	HR	HR.95L	HR.95H	pvalue
LINC00857	3.03064	1.56222	1.170701	2.084675	0.00244
EP300-AS1	-2.80409	0.594574	0.413412	0.855123	0.005046
FENDRR	-2.71124	0.574403	0.384727	0.85759	0.006703
LINC00968	-2.68167	0.347052	0.160133	0.752157	0.007326
TBX5-AS1	-2.42441	0.667035	0.480824	0.925363	0.015333
LINC00092	-2.333	0.384128	0.171946	0.858142	0.019648
KIFC1	2.320891	1.200048	1.028765	1.399848	0.020293
PCAT19	-2.30929	0.677321	0.486614	0.942767	0.020927
LINC00460	2.250591	1.15256	1.018504	1.30426	0.024411
SNHG3	-2.10824	0.797356	0.645984	0.984199	0.03501
SFTA1P	-2.08702	0.904885	0.823814	0.993934	0.036887

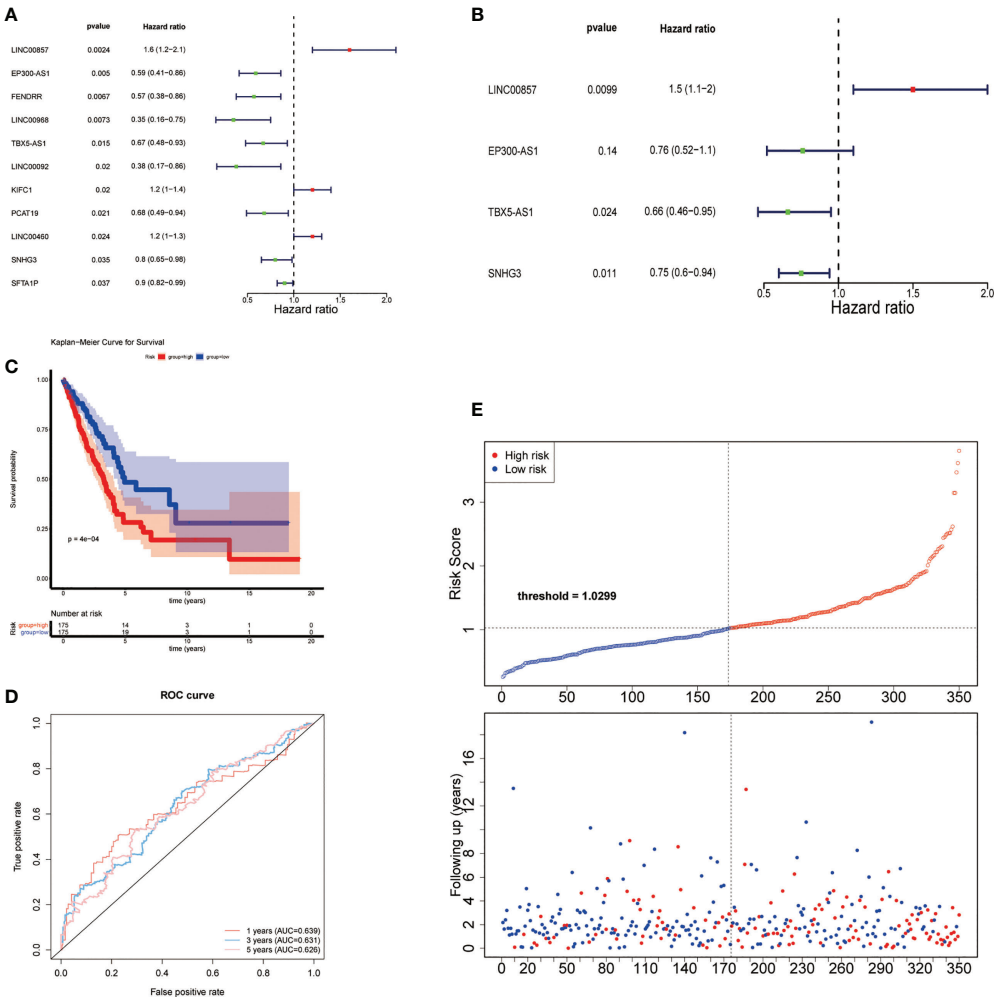


FIGURE 3
Construction of risk feature. **(A)** Forest plot of univariate Cox regression analysis results of lipid metabolism-related DE-lncRNAs. Red and green indicate risk and protective factors, respectively. **(B)** Forest plot of multivariate Cox regression analysis results of lipid metabolism-related DE-lncRNAs. **(C)** K–M survival curve of the risk score. **(D)** Time-ROC curve analysis of risk feature in 1, 3,5 years. **(E)** The distribution of risk scores, patient survival status, and survival time.

calculated the correlation between 10 significantly different immune cells and risk scores (Figure 8E). The results showed the expressions of resting memory CD4 T cells, resting dendritic cells, and resting mast cells were negatively correlated with the risk score, while activated memory CD4 T cells, M0 macrophages and M1 macrophages expressions were positively correlated with the risk score. The rank-sum test results showed that there were significant differences in T cell dysfunction score and T cell exclusion score between the two groups (Figures 8F, G, $P< 0.0001$ and $P< 0.05$). However, there was no significant differences in TIDE scores, PD-1, and PD-L1 between high- and low-risk groups (Figures 8H, I).

TABLE 3 4 prognostic signatures were identified by multivariate Cox regression analysis.

id	coef	HR	HR.95L	HR.95H	pvalue
LINC00857	0.384866	1.469417	1.096862	1.968513	0.00989
EP300-AS1	-0.2774	0.757748	0.523888	1.096001	0.140707
TBX5-AS1	-0.41464	0.660576	0.460979	0.946595	0.023885
SNHG3	-0.28443	0.752443	0.604604	0.936431	0.010821

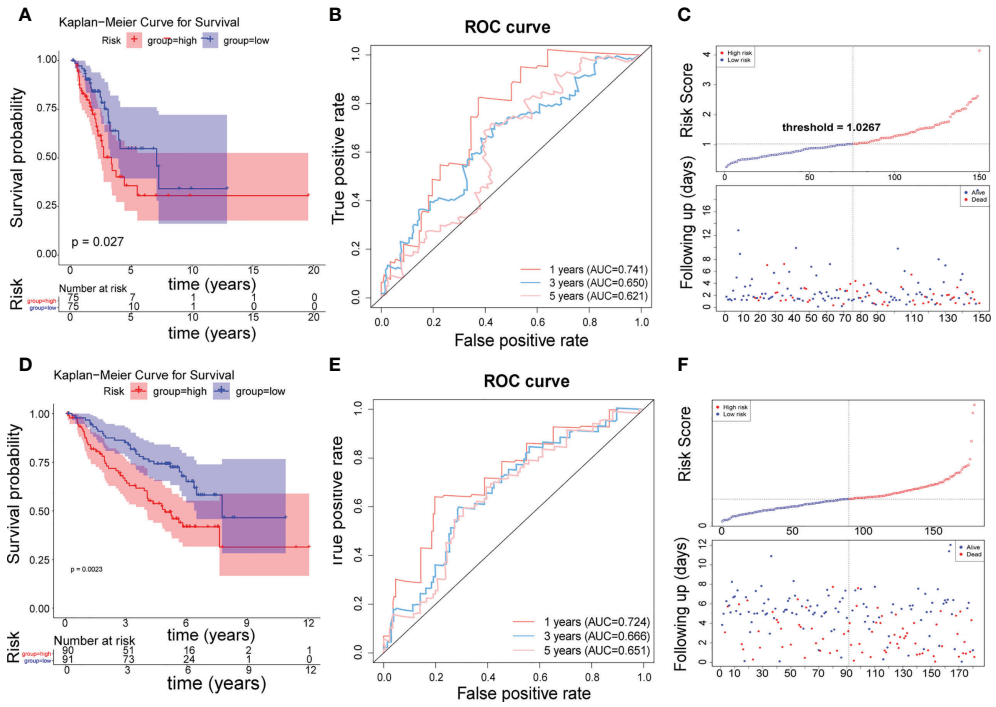


FIGURE 4
Testing and validation of risk features. **(A)** K–M survival curve of the risk score in the testing set. **(B)** Time-ROC curve analysis of risk feature in 1, 3,5 years in the testing set. **(C)** The distribution of risk scores, patient survival status, and survival time in the testing set. **(D)** K–M survival curve of the risk score in the validation set. **(E)** Time-ROC curve analysis of risk feature in 1, 3,5 years in the validation set. **(F)** The distribution of risk scores, patient survival status, and survival time in the validation set.

Differences in TMB levels, neoantigen levels and chemotherapy drug sensitivity

The rank-sum test of the TMB was performed, and the results showed that there was a significant difference between the two groups (Figure 9A, $P< 0.0001$). Moreover, the rank-sum test of the neoantigen values was performed, and the results suggested that the neoantigen values of the two groups were significantly different (Figure 9B, $P< 0.01$). A total of 33 drugs showed significant differences between the two groups (Figures 9C–F). Overall, the IC50 value of the low-risk group was lower, indicating that the chemotherapy effect of patients in the low-risk group may be more optimistic.

Expression validation and diagnostic value analysis of prognostic lncRNAs

The expression trends of the four prognostic factors in LUAD and normal samples from the TCGA-LUAD and the GSE31210 datasets were basically consistent. Compared with the normal group, the expressions of EP300-AS1 and TBX5-AS1 was down-regulated in the tumor group, while the expressions of LINC00857 and SNHG3 were up-regulated (Figures 10A, B). The qRT-PCR was used to verify the expression levels of prognostic genes and indicated that the expression levels of the four lncRNA were distinctly different between the Para-cancerous tissue and cancerous tissues samples (all P< 0.05; Figure 10C), with the expression of LINC00857 ($P = 0.0025$, $t = 2.388$) and SNHG3 ($P = 0.0361$, $t = 2.462$) being up-regulated in the cancerous tissues, the expression of TBX5-AS1 ($P = 0.0367$, $t = 2.451$) and EP300-AS1 ($P = 0.0407$, $t = 2.388$) being down-regulated in the cancerous tissues, consistent with the results of TCGA-LUAD cohort and GSE31210 dataset. Prognostic lncRNAs have a good diagnostic ability in TCGA and GSE31210 datasets, with AUC greater than 0.78 (Figures 10D, E). In addition, we also evaluated the overall diagnostic performance of 4 genes, which performed well in TCGA-LUAD cohort (AUC = 0.997) and GSE31210 dataset (AUC = 0.939), with AUC greater than 0.9 (Figures 10F, G). This indicates that the overall diagnostic model of the four prognostic lncRNAs is more effective than the single-gene diagnostic effect.

Discussion

Dysregulation of lipid metabolism is one of the most representative metabolic disorders in cancer. Lipid metabolism is used by cancer cells to enable cancer cells to proliferate,

TABLE 4 Characteristics of patients in low and high risk groups in TCGA-LUAD cohort.

Variable	Expression			
	Total (N=102)	high (N=56)	low (N=46)	P-value
gender				
female	53 (52.0%)	26 (46.4%)	27 (58.7%)	0.301
male	49 (48.0%)	30 (53.6%)	19 (41.3%)	
age (years)				
>=60	77 (75.5%)	40 (71.4%)	37 (80.4%)	0.412
<60	25 (24.5%)	16 (28.6%)	9 (19.6%)	
M				
M0	96 (94.1%)	53 (94.6%)	43 (93.5%)	1
M1	6 (5.9%)	3 (5.4%)	3 (6.5%)	
N				
N0	58 (56.9%)	27 (48.2%)	31 (67.4%)	0.15
N1	32 (31.4%)	21 (37.5%)	11 (23.9%)	
N2	12 (11.8%)	8 (14.3%)	4 (8.7%)	
T				
T1	35 (34.3%)	11 (19.6%)	24 (52.2%)	0.00371
T2	59 (57.8%)	41 (73.2%)	18 (39.1%)	
T3	5 (4.9%)	3 (5.4%)	2 (4.3%)	
T4	3 (2.9%)	1 (1.8%)	2 (4.3%)	
STAGE				
Stage I	52 (51.0%)	24 (42.9%)	28 (60.9%)	0.279
Stage II	30 (29.4%)	20 (35.7%)	10 (21.7%)	
Stage III	14 (13.7%)	9 (16.1%)	5 (10.9%)	
Stage IV	6 (5.9%)	3 (5.4%)	3 (6.5%)	

survive, invade, metastasize, and obtain the energy, biofilm components and signal molecules required for cancer microenvironment and cancer treatment response consists of various types of cells, cytokines, growth factors, and nutrients including lipids (33). LncRNAs have become key biomarkers for tumor diagnosis and treatment (34). Numerous studies have focused on the functions of genes involved in lipid metabolism (35–38). There were no studies have been seen on lipid metabolism-related lncRNA features to predict the prognosis and treatment of lung adenocarcinoma patients. Therefore, we conducted this study to establish a lipid metabolism-related lncRNA signature based on a large-scale database to predict the prognosis and treatment of LUAD patients.

In this study, we screened lipid metabolism-related lncRNAs for the first time by correlation analysis between lncRNAs and lipid metabolism-related genes using multiple datasets from the TCGA and GEO cohorts. Prognostic features based on four lipid metabolism-related lncRNAs (LINC00857, EP300-AS1, TBX5-AS1, SNHG3) were constructed using COX regression. And validated by ROC curve and KM curve in an independent cohort, this risk prediction model can be used as an independent prognostic factor for LUAD.

Our study found that four lncRNAs including LINC00857, EP300-AS1, TBX5-AS1, and SNHG3 were abnormally expressed in LUAD. Among them, LINC00857 is overexpressed in LUAD and can regulate the proliferation, apoptosis and glycolysis of LUAD cells by targeting the miR-1179/SPAG5 axis (39). TBX5-AS1 functions in LUAD, lung squamous cell carcinoma (LUSC), Adrenocortical carcinoma, and uterine corpus endometrial carcinoma. TBX5-AS1 belongs to a subclass of lncRNAs called enhancer RNAs. TBX5-AS1 is downregulated in LUAD. Inhibition of tumor progression through the PI3K/AKT pathway affects the prognosis of LUAD patients (40–43). SNHG3 is up-regulated in ovarian cancer, glioma, hepatocellular carcinoma, and osteosarcoma, all of which are associated with poor prognosis (44–47). SNHG3 is overexpressed in lung cancer tissues and cells, and a lot of studies have suggested that SNHG3 can affect the prognosis of LUAD through multiple pathways. For example, SNHG3 was activated by E2F1 and promoted not only proliferation but also the migration of LUAD cells through activating TGF-β pathway and IL-6/JAK2/STAT3 pathway (48). SNHG3 promotes the occurrence and progression of LUAD through regulating miR-515-5p/SUMO2 axis, miR-216a/ZEB1 axis, the miR-1343-3p/NFIX pathway or the expression of miR-890 (49–52).

In this study, the expression of EP300-AS1 was down-regulated in LUAD, which may play a tumor suppressor role in LUAD. At present, there is no relevant research report, and

10.3389/fonc.2022.986367

TABLE 5 Characteristics of patients in low and high risk groups GSE50081 dataset.

Variable	Expression			
	Total (N=181)	high (N=90)	low (N=91)	P-value
gender				
female	83 (45.9%)	38 (42.2%)	45 (49.5%)	0.408
male	98 (54.1%)	52 (57.8%)	46 (50.5%)	
age (years)				
>=60	154 (85.1%)	76 (84.4%)	78 (85.7%)	0.975
<60	27 (14.9%)	14 (15.6%)	13 (14.3%)	
M				
M0	181 (100%)	90 (100%)	91 (100%)	0.941
N				
N0	129 (71.3%)	57 (63.3%)	72 (79.1%)	0.0291
N1	52 (28.7%)	33 (36.7%)	19 (20.9%)	
T				
T1	57 (31.5%)	22 (24.4%)	35 (38.5%)	0.0556
T2	122 (67.4%)	66 (73.3%)	56 (61.5%)	
T3	2 (1.1%)	2 (2.2%)	0 (0%)	
STAGE				
STAGE I	127 (70.2%)	55 (61.1%)	72 (79.1%)	0.0129
STAGE II	54 (29.8%)	35 (38.9%)	19 (20.9%)	
smoking				
Current	57 (31.5%)	35 (38.9%)	22 (24.2%)	0.0055
Ex-smoker	79 (43.6%)	37 (41.1%)	42 (46.2%)	
Never	24 (13.3%)	5 (5.6%)	19 (20.9%)	
Unable to determine	21 (11.6%)	13 (14.4%)	8 (8.8%)	

the specific molecular mechanism and prognostic potential need to be further explored.

Since our model was confirmed to have good predictive accuracy, we performed GSEA enrichment analysis of prognostic lncRNA-related mRNAs, and the results suggested that they could influence the prognosis of LUAD patients by modulating structural changes in the extracellular matrix or affecting the development of the respiratory system. The GO and KEGG analysis showed that it was enriched in multiple biological processes, mainly involved in DNA replication, oxidative phosphorylation, and pyrimidine metabolism. The key enrichment pathways for prognostic features are B-cell mediated immunity, cell cycle checkpoint, chromosomal segregation, etc. We found that this feature is more involved in the biological processes and pathways of the cell cycle and immune response. Studies have shown that lnc00857 can regulate the cell cycle by regulating CCNE1 and CDK2 expression causing G1/S phase arrest (53). Down-regulation of TBX5-AS1 expression improves cell viability, migration, and invasion, while inhibiting apoptosis (40). SNHG3 regulates LUAD cell proliferation and cell cycle while inhibiting apoptosis (54). Based on the above results, the difference in prognosis between high and low risk groups may be related to cell cycle and immune pathways.

Based on enrichment analysis, prognostic features were associated with immune pathways. Finally, the immune

TABLE 6 Prognostic factors associated with LUAD were detected by univariate Cox regression analysis in the TCGA-LUAD cohort.

variable	coef	HR	HR.95L	HR.95H	pvalue
age	0.018574	1.018748	0.994246	1.043854	0.134812
gender	0.659332	1.933501	1.189575	3.142656	0.007804
T	0.743096	2.102435	1.22163	3.618308	0.007304
N	0.428074	1.534299	0.945599	2.489504	0.083015
STAGE	0.523916	1.688628	1.049138	2.717911	0.030967
riskscore	5.500862	244.9029	8.368246	7167.267	0.001407

TABLE 7 Independent prognostic indicators of LUAD were detected by multivariate Cox regression analysis in the TCGA-LUAD cohort.

variable	coef	HR	HR.95L	HR.95H	pvalue
gender	0.650719	1.916919	1.175847	3.12505	0.009065
T	0.516487	1.676129	0.974879	2.881801	0.061769
STAGE	0.418222	1.519259	0.934037	2.471152	0.091984
riskscore	4.243822	69.67362	1.631621	2975.209	0.026722

infiltration analysis revealed that the expressions of T cells CD4 memory rest, Dendritic cells resting and Mast cells resting were negatively correlated with risk score, while T cells CD4 memory activated, Macrophages M0 and Macrophages M1 expressions were positively correlated with risk score. The predicted prognosis of LUAD patients may be related to differences in immune cell composition.

We used three immunotherapy biomarkers to assess the predictive power of this signature. Notably, we predict the possibility of response to immunotherapy based on TIDE score and used the TIDE to predict immune checkpoint blockade response. It integrates the expression of two main mechanisms of tumor immune escape, T cell dysfunction and T cell rejection, to simulate tumor immune escape and can be used to predict the immunotherapy response of lung cancer (55). Interestingly, the low-risk group had higher T cell dysfunction score, while the high risk group had higher T cell exclusion score. This suggests that high-risk groups may benefit from the administration of checkpoint inhibitors (ICIs). Next, we also compared tumor mutational burden(TMB)and neoantigen between high and low-risk groups. Both TMB and neoantigens play an important role in tumor immune response. TMB can be used as a biomarker of response to checkpoint inhibitors (56), and neoantigens can be used as a biomarker to predict immune response to lung cancer (57). As predicted, the high-risk group exhibited higher mutational loads and neoantigens. Relatively

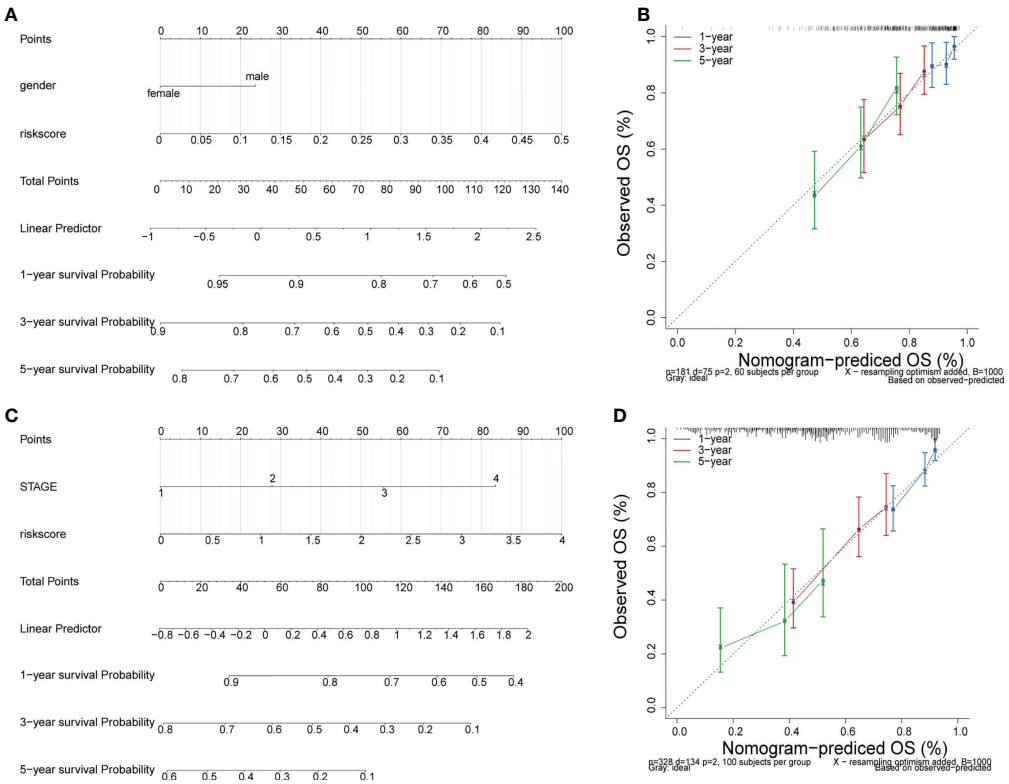


FIGURE 5
Independent prognosis analysis. **(A)** The nomogram of survival probability based on all independently predictive variables from GEO dataset. **(C)** The nomogram of survival probability based on all independently predictive variables from TCGA dataset. **(B, D)** Calibration curves for evaluating the agreement between the predicted and the actual survival rate for the prognosis model.

TABLE 8 Prognostic factors associated with LUAD were detected by univariate Cox regression analysis in the GSE50018 dataset.

variable	coef	HR	HR.95L	HR.95H	pvalue
STAGE	0.459406	1.583133	1.35183	1.854013	1.19E-08
N	0.535674	1.7086	1.405516	2.07704	7.59E-08
T	0.458022	1.580944	1.297362	1.926512	5.60E-06
riskscore	0.504493	1.656146	1.276325	2.148998	0.000147
M	0.608396	1.837482	1.034448	3.263905	0.03794
gender	0.053664	1.05513	0.751376	1.481682	0.756715
age	0.001727	1.001728	0.984147	1.019624	0.84841

speaking, the high-risk group had a higher objective response rate to immunotherapy and could benefit from the ICIs of LUAD (58). Studies have shown that the ability of TIDE to predict response to immunotherapy has been proven to be empirically superior to known immunotherapy biomarkers such as TMB and neoantigens (55). However, there was no significant difference in TIDE score, PD-1, and PD-L1 distribution between the two groups. Whether this feature can effectively predict the response to immunotherapy needs further research to verify.

We also focused on the differences in chemotherapeutic drug sensitivity. Interestingly, IC50 values in the low-risk group were significantly lower than those in the high-risk group, suggesting that the patients in the low-risk group may have better outcomes with chemotherapy. Based on this result, individualized treatment regimens can be developed according to the risk scores of different LUAD patients.

A large number of studies have shown that lncRNAs have been confirmed to be associated with poor prognosis of cancer, and the expression levels of lncRNAs can be used as diagnostic markers in addition to prognostic markers of intrinsic characteristics of cancer (59). A diagnostic biomarker detects or confirms the presence of a disease or condition of interest, or identifies an individual with a subtype of the disease. However, the diagnostic value of lncRNAs is rarely reported. The development of new diagnostic biomarkers is particularly important for early detection, early treatment and improved prognosis of LUAD patients (60). Therefore, to verify whether our model has diagnostic value, we performed expression validation of prognostic lncRNAs using an external independent dataset, and validated the expression levels of prognostic-related genes by performing qRT-PCR on paracancerous tissue and cancerous tissue samples from LUAD patients. Biomarkers need to ensure a low false positive rate. The use of ROC curve is conducive to the rational use of diagnostic biomarker evaluation. Decision thresholds and clinical utility are becoming important measures for assessing the value of biomarkers for clinical application (61). In this study, four lncRNAs (LINC00857, EP300-AS1, TBX5-AS1, and SNHG3) that are significantly related to the survival of LUAD patients ($p<0.05$) were used as diagnostic biomarkers, and their false positive concerns as LUAD biomarkers were ruled out by drawing the ROC curve of prediction model and single lncRNA. The results of clinical and independent prognostic analysis show that the risk score based on this model has independent prognostic value, and the corresponding nomogram and calibration curve also show that the model we built can be used in clinical diagnosis. On the other hand, the method of cross validation using multiple mutually exclusive "training" and "validation" samples is usually used for clinical validation of biomarkers (62). In this study, TCGA-LUAD (test set and validation set) and GSE50081 (external validation set) were jointly used to evaluate risk models, ensuring their effectiveness. As expected, the four prognostic lncRNAs signatures related to lipid metabolism showed good diagnostic ability and were able to distinguish between paracancerous and cancerous tissues. At the same time, the overall diagnostic model of the four prognostic lncRNAs is more robust than the individual lncRNAs. This study also suggests that this lncRNAs signature is a potential diagnostic tool for LUAD patients.

Based on survival analysis, prognostic features can effectively predict the total survival (OS) of early LUAD (63). By combining the prognostic features with classic clinical risk factors, it can

TABLE 9 Prognostic factors associated with LUAD were detected by multivariate Cox regression analysis in the GSE50018 dataset.

Variable	coef	HR	HR.95L	HR.95H	pvalue
N	0.22638	1.254052	0.977701	1.608513	0.074683
STAGE	0.250762	1.285005	1.023977	1.612572	0.030427
T	0.213124	1.237538	0.994005	1.540738	0.056626
riskscore	0.361873	1.436016	1.094731	1.883698	0.008957

Because of covering several pages, Tables 9–11 are provided in Supplementary Material.

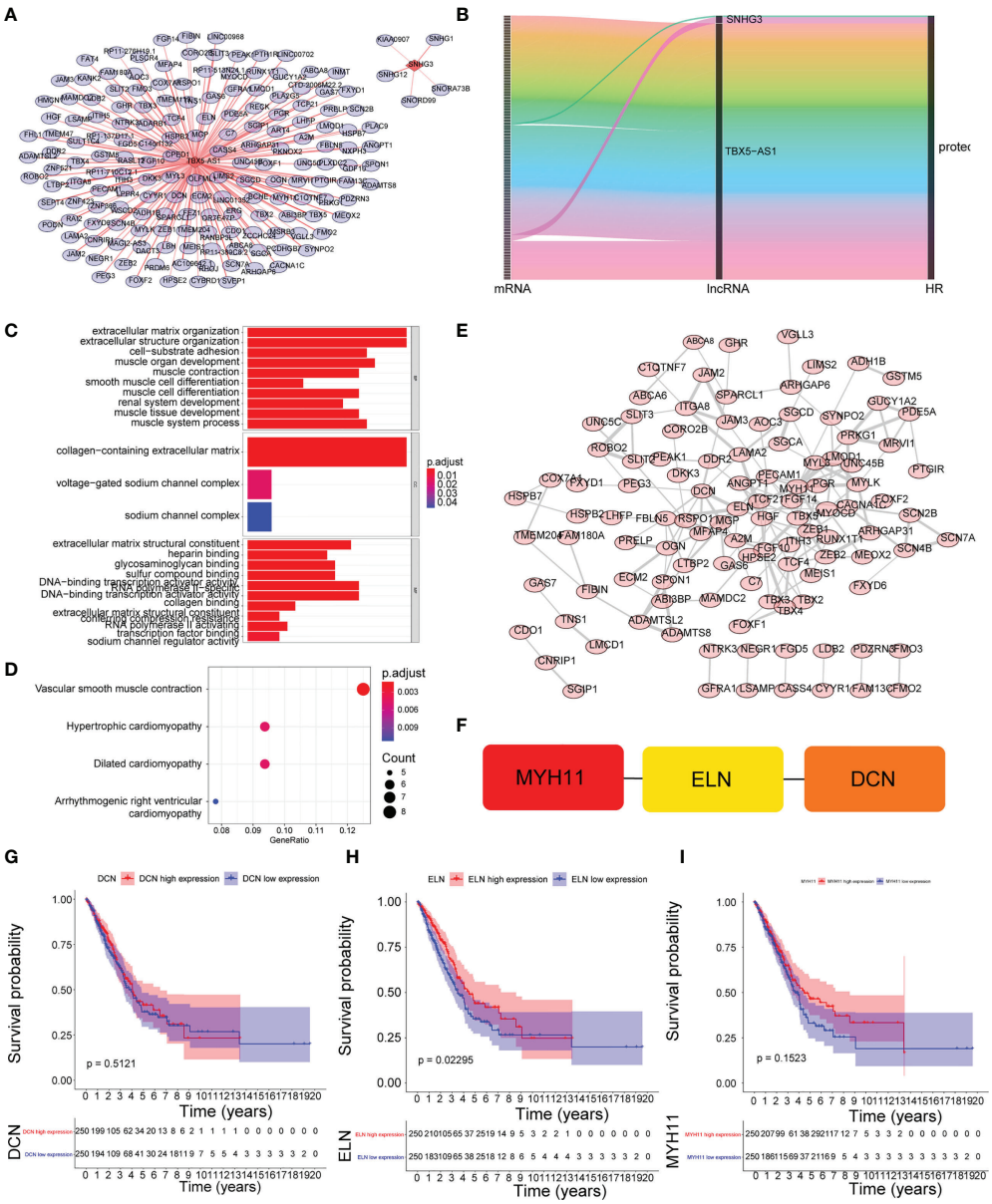


FIGURE 6
Identification of mRNAs associated with prognostic lncRNAs. **(A)** The coexpression network between prognostic lncRNAs and mRNA. Red diamond nodes represent prognostic lncRNAs, and the sky blue round nodes represent prognostic-related mRNAs. The coexpression network was visualized using Cytoscape 3.7.2 software. Green circles represent mRNAs that are both associated with prognostic lncRNAs and cis-regulated. **(B)** Sankey diagram showed the association between prognostic related lncRNAs and mRNAs.**(C)** GO enrichment analysis. **(D)** KEGG enrichment analysis. **(E)** Protein interaction network. **(F)** identification of the top3 hub genes. **(G–I)** K–M survival curve of the expression levels of MYH11, ELN, DCN and the corresponding survival information.

also be found that when patients are exposed to the same clinical risk factors, the prognosis of the high-risk group is significantly worse. At the same time, prognostic features can also be used to guide treatment (64). A large number of studies have shown that proteins and mRNAs have been validated as biomarkers of various cancers (38, 65). However, the stability of these biomarkers will be affected by the regulation and modification of proteins and mRNAs at the transcriptional level. LncRNAs are effectors whose function depends on their expression levels. Many lncRNAs have also been shown to be associated with poor cancers prognosis (59, 66, 67). Therefore, the expression level of lncRNAs can be used as a better biomarker to predict or diagnose the prognosis of tumors. In this study, this lncRNAs signature is not only related to lipid metabolism, but also related

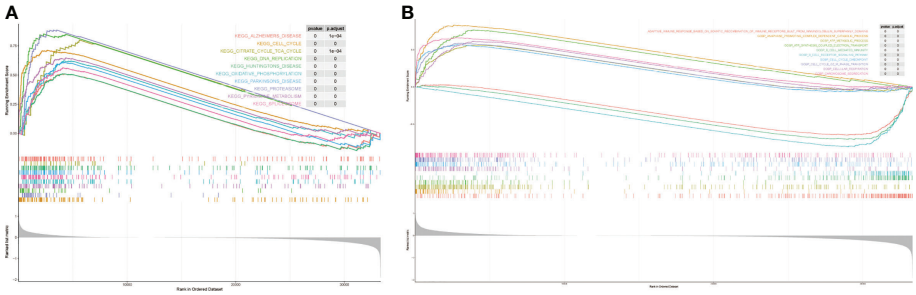


FIGURE 7
The results of gene set enrichment analysis. **(A)** KEEG enrichment analysis. **(B)** GO enrichment analysis.

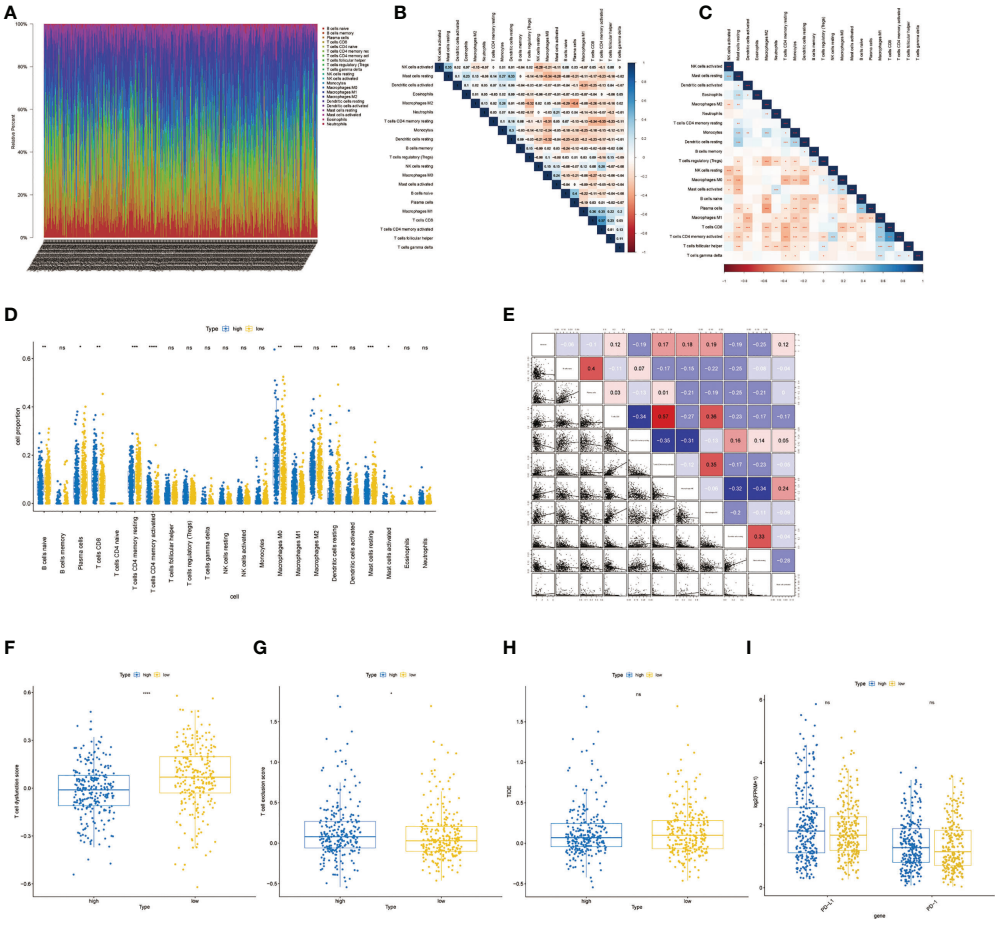


FIGURE 8
Immune infiltration analysis and differences in response to immunotherapy.**(A)** Proportion of immune cell infiltration.**(B)** Correlation heatmap of immune cell proportion. **(C)** Correlation p-value heatmap of immune cell proportion. **(D)** Boxplot of immune cell differences between the low- and high-risk group. **(E)** The correlation between significantly different immune cells and risk scores. **(F)** Boxplot of T cell dysfunction score between high- and low-risk groups. **(G)** Boxplot of T cell exclusion score between high- and low-risk groups.**(H)** Boxplot of TIDE score between high- and low-risk groups. *, **, ***, and **** represent P < 0.05, P<0.01, P < 0.001, and P < 0.0001, respectively. ns, not significant.

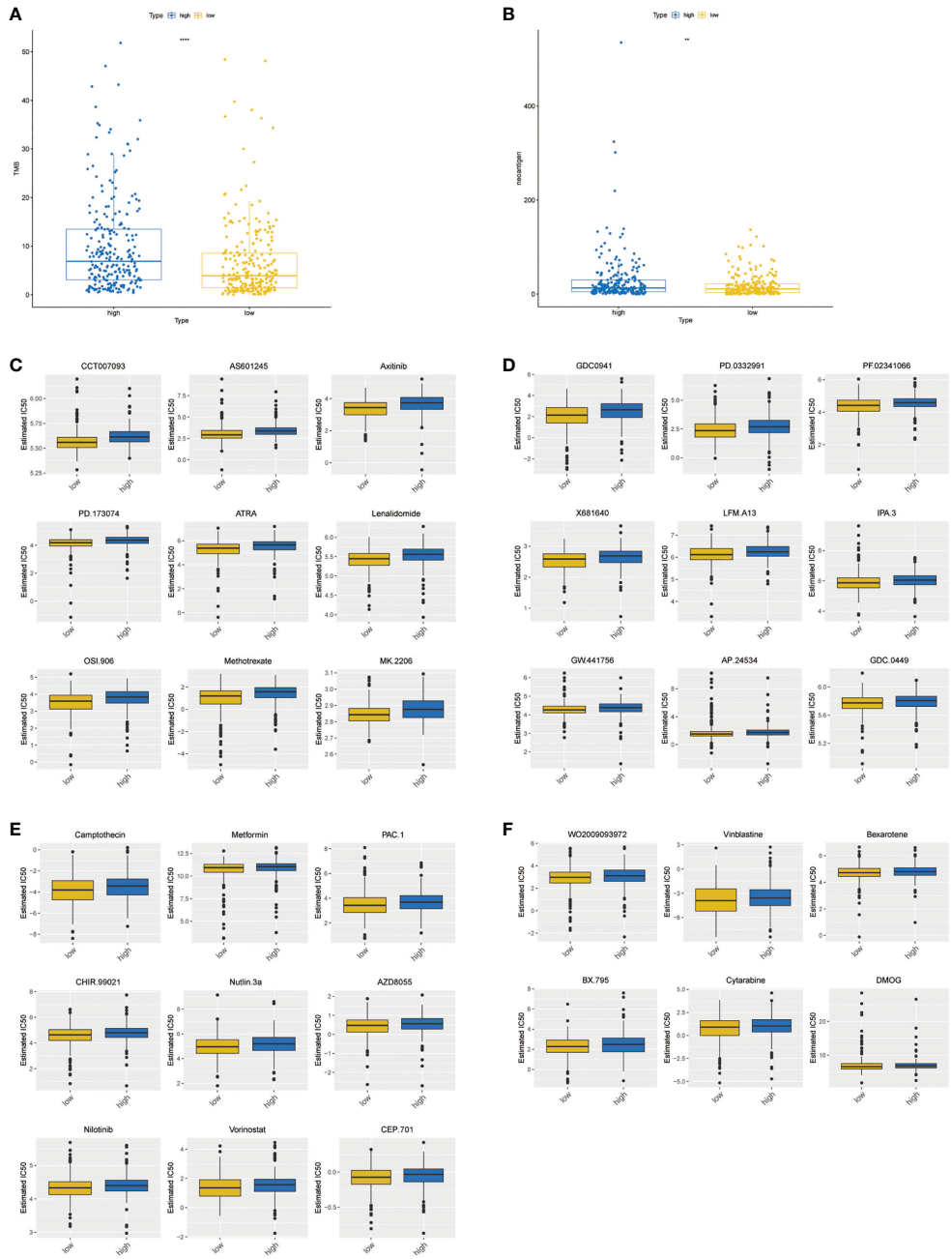


FIGURE 9
Differences in TMB levels, neoantigen levels, and chemotherapy drug sensitivity between high- and low-risk groups. **(A)** Boxplot of the TMB between high- and low-risk groups. **(B)** Boxplot of the neoantigen values between high- and low-risk groups. **(C–F)** Boxplot of 33 chemotherapy drug sensitivity between high- and low-risk groups. ** represents P < 0.01, **** represents P < 0.0001.

to specific biological mechanisms. In contrast, this is the advantage of our prognostic model.

Although this is the first time to construct the prognostic characteristics of lipid metabolism-related lncRNA of LUAD, and multi-dimensional verification has been carried out, our study still has certain limitations. First, the data analyzed in this study are all from online databases, and larger samples are needed to further study the clinical application of our findings in LUAD. Second, this paper is a retrospective study, needed to corroborate by corresponding prospective studies. Finally, functional experiments are needed to further elucidate the intrinsic molecular mechanisms of lipid metabolism-related lncRNAs.

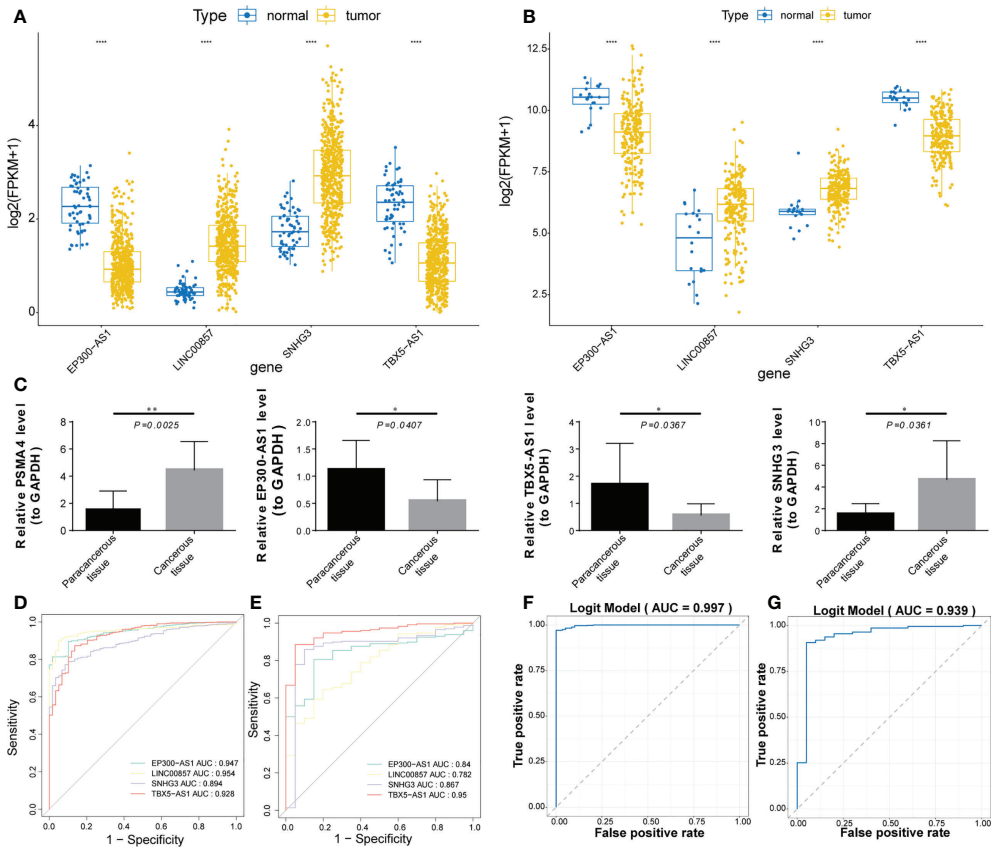


FIGURE 10
Expression validation and diagnostic value analysis of prognostic lncRNAs. **(A, B)** Boxplot of the expression trends of the four prognostic lncRNAs in 2 datasets. **(C)** Boxplots of the expression levels of the four lncRNAs between the paracancerous and cancerous tissue samples. **(D, E)** ROC curve analysis of prognostic lncRNAs diagnostic ability in 2 datasets. **(F, G)** ROC curve analysis of the overall diagnostic model in 2 datasets. *, **, and **** represent $P < 0.05$, $P < 0.01$, and $P < 0.0001$, respectively.

In conclusion, this study constructed the prognostic characteristics of 4 lncRNAs related to lipid metabolism for the first time by analyzing bioinformatics methods and based on multiple databases, which proved to have important prognostic and therapeutic value for LUAD patients, as well as good diagnostic ability.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by The Third Affiliated Hospital of Kunming Medical University ethics committee. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

JZ and GL contributed to conception and design of the study. JZ organized the database, performed the statistical analysis and wrote the first draft of the manuscript. GZ, WW, ZS, and YY wrote sections of the manuscript. LY and YH manuscript revision. All authors contributed to read, and approved the submitted version.

Funding

This work was supported by Yunnan Fundamental Research Projects (No. 202201AY070001-135); National Natural Science Foundation of China (No. 81960500); Young and Middle-aged

Academic and Technical Leaders Reserve Talents Project of Yunnan Province (No. C20048).

Acknowledgments

We appreciated TCGA and GEO databases for providing the original study data.

Conflict of interest

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

Publisher's note

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.2022.986367/full#supplementary-material

References

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

2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin* (2020) 70:7–30. doi: 10.3322/caac.21590

3. de Sousa VML, Carvalho L. Heterogeneity in lung cancer. *Pathobiology* (2018) 85:96–107. doi: 10.1159/000487440

4. Peng WX, Koirala P, Mo YY. LncRNA-mediated regulation of cell signaling in cancer. *Oncogene* (2017) 36:5661–7. doi: 10.1038/onc.2017.184

5. Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. *Cell* (2013) 154:26–46. doi: 10.1016/j.cell.2013.06.020

6. Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. *Mol Cell* (2011) 43:904–14. doi: 10.1016/j.molcel.2011.08.018

7. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. *Annu Rev Biochem* (2012) 81:145–66. doi: 10.1146/annurev-biochem-051410-092902

8. Gibb EA, Brown CJ, Lam WL. The functional role of long non-coding RNA in human carcinomas. *Mol Cancer* (2011) 10:38. doi: 10.1186/1476-4598-10-38

9. Rohrig F, Schulze A. The multifaceted roles of fatty acid synthesis in cancer. *Nat Rev Cancer* (2016) 16:732–49. doi: 10.1038/nrc.2016.89

10. Chen Z. Progress and prospects of long noncoding RNAs in lipid homeostasis. *Mol Metab* (2016) 5:164–70. doi: 10.1016/j.molmet.2015.12.003

11. Ackerman D, Tumanov S, Qiu B, Michalopoulou E, Spata M, Azzam A, et al. Triglycerides promote lipid homeostasis during hypoxic stress by balancing fatty acid saturation. *Cell Rep* (2018) 24:2596–2605 e5. doi: 10.1016/j.celrep.2018.08.015

12. van Solingen C, Scacalossi KR, Moore KJ. Long noncoding RNAs in lipid metabolism. *Curr Opin Lipidol* (2018) 29:224–32. doi: 10.1097/MOL.0000000000000503

13. Rayner KJ, Moore KJ. MicroRNA control of high-density lipoprotein metabolism and function. *Circ Res* (2014) 114:183–92. doi: 10.1161/CIRCRESAHA.114.300645

14. Schmitz SU, Grote P, Herrmann BG. Mechanisms of long noncoding RNA function in development and disease. *Cell Mol Life Sci* (2016) 73:2491–509. doi: 10.1007/s00018-016-2174-5

15. Eberle D, Hegarty B, Bossard P, Ferre P, Foufelle F. SREBP transcription factors: master regulators of lipid homeostasis. *Biochimie* (2004) 86:839–48. doi: 10.1016/j.biochi.2004.09.018

16. Liu X, Liang Y, Song R, Yang G, Han J, Lan Y, et al. Long non-coding RNA NEAT1-modulated abnormal lipolysis *via* ATGL drives hepatocellular carcinoma proliferation. *Mol Cancer* (2018) 17:90. doi: 10.1186/s12943-018-0838-5

17. Yan C, Chen J, Chen N. Long noncoding RNA MALAT1 promotes hepatic steatosis and insulin resistance by increasing nuclear SREBP-1c protein stability. *Sci Rep* (2016) 6:22640. doi: 10.1038/srep22640

18. Sallam T, Jones MC, Gilliland T, Zhang L, Wu X, Eskin A, et al. Feedback modulation of cholesterol metabolism by the lipid-responsive non-coding RNA LeXis. *Nature* (2016) 534:124–8. doi: 10.1038/nature17674

19. Halley P, Kadakkuzha BM, Faghihi MA, Magistri M, Zeier Z, Khorkova O, et al. Regulation of the apolipoprotein gene cluster by a long noncoding RNA. *Cell Rep* (2014) 6:222–30. doi: 10.1016/j.celrep.2013.12.015

20. Liu G, Zheng X, Xu Y, Lu J, Chen J, Huang X. Long non-coding RNAs expression profile in HepG2 cells reveals the potential role of long non-coding RNAs in the cholesterol metabolism. *Chin Med J (Engl)* (2015) 128:91–7. doi: 10.4103/0366-6999.147824

21. Koundouros N, Poulogiannis G. Reprogramming of fatty acid metabolism in cancer. *Br J Cancer* (2020) 122:4–22. doi: 10.1038/s41416-019-0650-z

22. Yuan S, Liu Q, Hu Z, Zhou Z, Wang G, Li C, et al. Long non-coding RNA MUC5B-AS1 promotes metastasis through mutually regulating MUC5B expression in lung adenocarcinoma. *Cell Death Dis* (2018) 9:450. doi: 10.1038/s41419-018-0472-6

23. Wang G, Qiu M, Xing X, Zhou J, Yao H, Li M, et al. Lung cancer scRNA-seq and lipidomics reveal aberrant lipid metabolism for early-stage diagnosis. *Sci Transl Med* (2022) 14:eabk2756. doi: 10.1126/scitranslmed.abk2756

24. Chen Y, Ma Z, Shen X, Li L, Zhong J, Min LS, et al. Serum lipidomics profiling to identify biomarkers for non-small cell lung cancer. *BioMed Res Int* (2018) 2018:5276240. doi: 10.1155/2018/5276240

25. Zhou T, Zhan J, Fang W, Zhao Y, Yang Y, Hou X, et al. Serum low-density lipoprotein and low-density lipoprotein expression level at diagnosis are favorable prognostic factors in patients with small-cell lung cancer (SCLC). *BMC Cancer* (2017) 17:269. doi: 10.1186/s12885-017-3239-z

26. Chi PD, Liu W, Chen H, Zhang JP, Lin Y, Zheng X, et al. High-density lipoprotein cholesterol is a favorable prognostic factor and negatively correlated with c-reactive protein level in non-small cell lung carcinoma. *PloS One* (2014) 9:e91080. doi: 10.1371/journal.pone.0091080

27. Marien E, Meister M, Muley T, Fieuws S, Bordel S, Derua R, et al. Non-small cell lung cancer is characterized by dramatic changes in phospholipid profiles. *Int J Cancer* (2015) 137:1539–48. doi: 10.1002/ijc.29517

28. Li J, Li Q, Su Z, Sun Q, Zhao Y, Feng T, et al. Lipid metabolism gene-wide profile and survival signature of lung adenocarcinoma. *Lipids Health Dis* (2020) 19:222. doi: 10.1186/s12944-020-01390-9

29. Zou W, Wang Z, Wang F, Li L, Liu R, Hu M. A metabolism-related 4-lncRNA prognostic signature and corresponding mechanisms in intrahepatic cholangiocarcinoma. *BMC Cancer* (2021) 21:608. doi: 10.1186/s12885-021-08322-5

30. Luo D, Yao W, Wang Q, Yang Q, Liu X, Yang Y, et al. The nomogram based on the 6-lncRNA model can promote the prognosis prediction of patients with breast invasive carcinoma. *Sci Rep* (2021) 11:20863. doi: 10.1038/s41598-021-00364-w

31. Zhang C, Zhang G, Sun N, Zhang Z, Zhang Z, Luo Y, et al. Comprehensive molecular analyses of a TNF family-based signature with regard to prognosis, immune features, and biomarkers for immunotherapy in lung adenocarcinoma. *eBioMedicine* (2020) 59:102959. doi: 10.1016/j.ebiom.2020.102959

32. Hu X, Wu L, Liu B, Chen K. Immune infiltration subtypes characterization and identification of prognosis-related lncRNAs in adenocarcinoma of the esophagogastric junction. *Front Immunol* (2021) 12:651056. doi: 10.3389/fimmu.2021.651056

33. Bian X, Liu R, Meng Y, Xing D, Xu D, Lu Z. Lipid metabolism and cancer. *J Exp Med* (2021) 218:1. doi: 10.1084/jem.20201606

34. Roth A, Diederichs S. Long noncoding RNAs in lung cancer. *Curr Top Microbiol Immunol* (2016) 394:57–110.

35. Luo D, Xiao H, Dong J, Li Y, Feng G, Cui M, et al. B7-H3 regulates lipid metabolism of lung cancer through SREBP1-mediated expression of FASN. *Biochem Biophys Res Commun* (2017) 482:1246–51. doi: 10.1016/j.bbrc.2016.12.021

36. Ni Y, Yang Y, Ran J, Zhang L, Yao M, Liu Z, et al. miR-15a-5p inhibits metastasis and lipid metabolism by suppressing histone acetylation in lung cancer. *Free Radic Biol Med* (2020) 161:150–62. doi: 10.1016/j.freeradbiomed.2020.10.009

37. Eltayeb K, La Monica S, Tiseo M, Alfieri R, Fumarola C. Reprogramming of lipid metabolism in lung cancer: An overview with focus on EGFR-mutated non-small cell lung cancer. *Cells* (2022) 11:413. doi: 10.3390/cells11030413

38. Zhu M, Zeng Q, Fan T, Lei Y, Wang F, Zheng S, et al. Clinical significance and immunometabolism landscapes of a novel recurrence-associated lipid metabolism signature in early-stage lung adenocarcinoma: A comprehensive analysis. *Front Immunol* (2022) 13:783495. doi: 10.3389/fimmu.2022.783495

39. Wang L, Cao L, Wen C, Li J, Yu G, Liu C. LncRNA LINC00857 regulates lung adenocarcinoma progression, apoptosis and glycolysis by targeting miR-1179/SPAG5 axis. *Hum Cell* (2020) 33:195–204. doi: 10.1007/s13577-019-00296-8

40. Qu QH, Jiang SZ, Li XY. LncRNA TBX5-AS1 regulates the tumor progression through the PI3K/AKT pathway in non-small cell lung cancer. *Onco Targets Ther* (2020) 13:7949–61. doi: 10.2147/OTT.S255195

41. Cheng L, Han T, Chen B, Nie K, Peng W. TBX5-AS1, an enhancer RNA, is a potential novel prognostic biomarker for lung adenocarcinoma. *BMC Cancer* (2021) 21:794. doi: 10.1186/s12885-021-08517-w

42. Shih JH, Chen HY, Lin SC, Yeh YC, Shen R, Lang YD, et al. Integrative analyses of noncoding RNAs reveal the potential mechanisms augmenting tumor malignancy in lung adenocarcinoma. *Nucleic Acids Res* (2020) 48:1175–91. doi: 10.1093/nar/gkz1149

43. Xiong Y, Zhang X, Lin Z, Xiong A, Xie S, Liang J, et al. SFTA1P, LINC00968, GATA6-AS1, TBX5-AS1, and FEZF1-AS1 are crucial long non-coding RNAs associated with the prognosis of lung squamous cell carcinoma. *Oncol Lett* (2019) 18:3985–93. doi: 10.3892/ol.2019.10744

44. Hong L, Chen W, Wu D, Wang Y. Upregulation of SNHG3 expression associated with poor prognosis and enhances malignant progression of ovarian cancer. *Cancer biomark* (2018) 22:367–74. doi: 10.3233/CBM-170710

45. Fei F, He Y, He S, He Z, Wang Y, Wu G, et al. LncRNA SNHG3 enhances the malignant progress of glioma through silencing KLF2 and p21. *Biosci Rep* (2018) 38:5. doi: 10.1042/BSR20180420

46. Zhang PF, Wang F, Wu J, Wu Y, Huang W, Liu D, et al. LncRNA SNHG3 induces EMT and sorafenib resistance by modulating the miR-128/CD151 pathway in hepatocellular carcinoma. *J Cell Physiol* (2019) 234:2788–94. doi: 10.1002/jcp.27095

47. Xu B, Mei J, Ji W, Bian Z, Jiao J, Sun J, et al. LncRNA SNHG3, a potential oncogene in human cancers. *Cancer Cell Int* (2020) 20:536. doi: 10.1186/s12935-020-01608-x

48. Shi J, Li J, Yang S, Hu X, Chen J, Feng J, et al. LncRNA SNHG3 is activated by E2F1 and promotes proliferation and migration of non-small-cell lung cancer cells through activating TGF-beta pathway and IL-6/JAK2/STAT3 pathway. *J Cell Physiol* (2020) 235:2891–900. doi: 10.1002/jcp.29194

49. Li Y, Gao L, Zhang C, Meng J. LncRNA SNHG3 promotes proliferation and metastasis of non-Small-Cell lung cancer cells through miR-515-5p/SUMO2 axis. *Technol Cancer Res Treat* (2021) 20:15330338211019376. doi: 10.1177/15330338211019376

50. Kang B, Qiu C, Zhang Y. The effect of lncRNA SNHG3 overexpression on lung adenocarcinoma by regulating the expression of miR-890. *J Healthc Eng* (2021) 2021:1643788. doi: 10.1155/2021/1643788

51. Zhao L, Song X, Guo Y, Ding N, Wang T, Huang L. Long noncoding RNA SNHG3 promotes the development of nonsmall cell lung cancer *via* the miR13433p/NFIX pathway. *Int J Mol Med* (2021) 48:147. doi: 10.3892/ijmm.2021.4980

52. Zhao S, Gao X, Zhong C, Li Y, Wang M, Zang S. SNHG3 knockdown suppresses proliferation, migration and invasion, and promotes apoptosis in non-small cell lung cancer through regulating miR-216a/ZEB1 axis. *Onco Targets Ther* (2020) 13:11327–36. doi: 10.2147/OTT.S263637

53. Wang L, He Y, Liu W, Bai S, Xiao L, Zhang J, et al. Non-coding RNA LINC00857 is predictive of poor patient survival and promotes tumor progression *via* cell cycle regulation in lung cancer. *Oncotarget* (2016) 7:11487–99. doi: 10.18632/oncotarget.7203

54. Liu L, Ni J, He X. Upregulation of the long noncoding RNA SNHG3 promotes lung adenocarcinoma proliferation. *Dis Markers* (2018) 2018:5736716. doi: 10.1155/2018/5736716

55. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med* (2018) 24:1550–8. doi: 10.1038/s41591-018-0136-1

56. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Cancer immunology. mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* (2015) 348:124–8. doi: 10.1126/science.aaa1348

57. Chae YK, Viveiros P, Lopes G, Sukhadia B, Sheikh MM, Saravia D, et al. Clinical and immunological implications of frameshift mutations in lung cancer. *J Thorac Oncol* (2019) 14:1807–17. doi: 10.1016/j.jtho.2019.06.016

58. Hellmann MD, Ciuleanu TE, Pluzanski A, Lee JS, Otterson GA, Audigier-Valette C, et al. Nivolumab plus ipilimumab in lung cancer with a high tumor mutational burden. *N Engl J Med* (2018) 378:2093–104. doi: 10.1056/NEJMoa1801946

59. Qi P, Du X. The long non-coding RNAs, a new cancer diagnostic and therapeutic gold mine. *Mod Pathol* (2013) 26:155–65. doi: 10.1038/modpathol.2012.160

60. Fan ZY, Liu W, Yan C, Zhu ZL, Xu W, Li JF, et al. Identification of a five-lncRNA signature for the diagnosis and prognosis of gastric cancer. *Tumour Biol* (2016) 37:13265–77. doi: 10.1007/s13277-016-5185-9

61. Califf RM. Biomarker definitions and their applications. *Exp Biol Med (Maywood)* (2018) 243:213–21. doi: 10.1177/1535370217750088

62. Dobbin KK, Cesano A, Alvarez J, Hawtin R, Janetzki S, Kirsch I, et al. Validation of biomarkers to predict response to immunotherapy in cancer: Volume II - clinical validation and regulatory considerations. *J Immunother Cancer* (2016) 4:77. doi: 10.1186/s40425-016-0179-0

63. Wu P, Zheng Y, Wang Y, Wang Y, Liang N. Development and validation of a robust immune-related prognostic signature in early-stage lung adenocarcinoma. *J Transl Med* (2020) 18:380. doi: 10.1186/s12967-020-02545-z

64. Wang Z, Liu Y, Zhan X, Wang X, Zhang C, Qin L, et al. A novel prognostic signature of metastasis-associated genes and personalized therapeutic strategy for lung adenocarcinoma patients. *Aging (Albany NY)* (2022) 14:5571–89. doi: 10.18632/aging.204169

65. Martinez-Terroba E, Behrens C, de Miguel FJ, Agorreta J, Monso E, Millares L, et al. A novel protein-based prognostic signature improves risk stratification to guide clinical management in early-stage lung adenocarcinoma patients. *J Pathol* (2018) 245:421–32. doi: 10.1002/path.5096

66. Li Y, Shen R, Wang A, Zhao J, Zhou J, Zhang W, et al. Construction of a prognostic immune-related LncRNA risk model for lung adenocarcinoma. *Front Cell Dev Biol* (2021) 9:648806. doi: 10.3389/fcell.2021.648806

67. Zhou W, Zhang S, Li HB, Cai Z, Tang S, Chen LX, et al. Development of prognostic indicator based on autophagy-related lncRNA analysis in colon adenocarcinoma. *BioMed Res Int* (2020) 2020:9807918. doi: 10.1155/2020/9807918

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 02 November 2022
DOI 10.3389/fonc.2022.988332

Check for updates

OPEN ACCESS

EDITED BY
Luciano Mutti,
Temple University, United States

REVIEWED BY
Ni Zeng,
Affiliated Hospital of Zunyi Medical University, China
Qingjia Chi,
Wuhan University of Technology, China

*CORRESPONDENCE
Shaohui Wang
winter9091@163.com
Li Xiang
xianglydr@cdutcm.edu.cn

SPECIALTY SECTION
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

RECEIVED 08 July 2022
ACCEPTED 21 October 2022
PUBLISHED 02 November 2022

CITATION
Li J, Du Q, Sun J, Xiang L and Wang S (2022) Identification and validation of a novel phagocytosis regulators-related signature with potential prognostic and immunotherapeutic value in patients with lung adenocarcinoma. *Front. Oncol.* 12:988332. doi: 10.3389/fonc.2022.988332

COPYRIGHT
© 2022 Li, Du, Sun, Xiang 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.

Identification and validation of a novel phagocytosis regulators-related signature with potential prognostic and immunotherapeutic value in patients with lung adenocarcinoma

Jingyang Li[1], Qinyun Du[1], Jiayi Sun[2], Li Xiang[2]* and Shaohui Wang[3]*

[1]State Key Laboratory of Southwestern Chinese Medicine Resources, School of Pharmacy, Chengdu University of Traditional Chinese Medicine, Chengdu, China, [2]State Key Laboratory of Southwestern Chinese Medicine Resources, Innovative Institute of Chinese Medicine and Pharmacy, Chengdu University of Traditional Chinese Medicine, Chengdu, China, [3]State Key Laboratory of Southwestern Chinese Medicine Resources, School of Ethnic Medicine, Chengdu University of Traditional Chinese Medicine, Chengdu, China

Background: Lung adenocarcinoma (LUAD) is a malignant tumor that seriously affects the prognosis of patients. Tumor-associated macrophages (TAMs) play a vital role in the tumor microenvironment and can be used as a potential target for tumor therapy, and phagocytosis regulators (PRs) are particularly important in this process. However, the PRs-related signature that can predict the potential prognostic and immunotherapeutic value in patients with LUAD has not been discovered.

Methods: In this study, we mainly analyzed the effect of phagocytosis regulators on the prognosis of LUAD, and based on multiple screening analyses including differential analysis, univariate Cox analysis, and Lasso analysis, we constructed a prognostic risk model consisting of five genes. To verify the stability of the model, survival analysis and ROC curve verification were carried out through multiple data sets. In addition, we also combined many factors, such as immune infiltrating cells, clinical grouping characteristics, immune examination sites, pro-inflammatory factors, and other factors as well as *in vitro* cell experiments and clinical tissue samples for further validation analysis.

Results: After identifying 29 differentially expressed PRs in LUAD samples, we further constructed a prognostic model consisting of five prognostic signatures (FURIN, KIF23, SASH3, GNPNAT1, and ITGAL). Further survival analysis tests, ROC verification, as well as univariate and multivariate Cox regression analysis showed that the risk score of the model could well predict the prognosis of

LUAD patients and could be used as an independent prognostic factor. In addition, we further found that these phagocytic regulators-related signatures were closely related to the immune microenvironment and immunotherapy in LUAD patients, and could well predict the efficacy of immunotherapy in patients. *In vitro* cell experiments and Immunohistochemistry of clinical tissues showed that the expressions of FURIN, KIF23, SASH3, GNPNAT1 and ITGAL in normal lung cells/tissues and LUAD cells/tissues were consistent with bioinformatics results, and 3 of them had significant differences.

Conclusion: Our study identified a novel PRs-related signature that has potential application value in predicting the prognosis of LUAD patients and predicting the efficacy of immunotherapy. This provides a new basis for the prognosis assessment of LUAD patients and provides a novel target for immunotherapy of LUAD patients.

KEYWORDS

lung adenocarcinoma, phagocytosis regulators, prognostic value, immunotherapeutic value, risk score

1 Introduction

Globally, lung cancer was the leading cause of cancer-related deaths in 2020 and accounted for approximately 11.4% of all cancers and 18% of deaths (1). Non-small cell lung cancer (NSCLC) is the most common subtype of lung cancer, which consists of two main histological types: lung adenocarcinoma (LUAD), and lung squamous cell carcinoma (LUSC) (2). Despite major advances in surgery, new chemotherapy, and radiation treatments, survival rates for lung cancer remain grim. The 5-year survival rate for advanced NSCLC is less than 3% (3). The low survival rate of lung cancer can be attributed to its high rate of metastasis and recurrence, as well as severe drug resistance (4). In particular, LUAD is more prone to distant hematogenous, hematogenous, and lymphatic metastasis than LUSC, and the prognosis is relatively poor. Therefore, the discovery of new potential prognostic biomarkers and therapies is imperative for patients with LUAD.

The tumor microenvironment has a profound influence on the development, therapeutic effect, and prognosis of a tumor (5, 6). Tumor-associated macrophages (TAMs) are the main immune cells in the tumor microenvironment (6). Macrophages are the main component of leukocyte infiltration, and there are different numbers of macrophages in all tumors, studies have shown that tumor-associated macrophages can be used as tumor therapeutic targets (7). Macrophages are a key driver of tumor inflammation, and TAMs contribute to tumor progression at various levels, including promoting genetic instability, culturing cancer stem cells, paving the way for metastasis, and taming protective adaptive immunity (8, 9). TAMs express triggers of the checkpoint that the regulation of T cell, and is a target of the checkpoint blocking immunotherapy. Macrophage-centered therapies include strategies to prevent tumor recruitment and survival, induce extracellular killing or phagocytosis of cancer cells, and so on (8, 10). Therefore, identification and characterization of phagocytosis regulators (PRs) are particularly important to elucidate the mechanisms of phagocytosis in LUAD.

The rapid development of bioinformatics provides us with an effective way to explore the key biomarkers and molecular mechanisms of the occurrence and development of diseases (such as cancers, diabetic nephropathy, etc.) (11–13). In this study, we aimed to characterize the influence of PRs on the occurrence and development of LUAD by using the PRs identified in two genome-wide CRISPR articles (14, 15), and constructed the PRs signature with potential prognostic value. Further, we explored the interaction between the signature and macrophages, and predicted its prognosis and immunotherapeutic effect in LUAD patients.

2 Materials and methods

2.1 Data collection

The genome sequencing data, clinical data, and survival information of TCGA-LUAD were downloaded from the XENA database (https://xenabrowser.net/datapages). Among them, there were 582 intersection samples of transcriptome

data (standardized count data were selected due to differences) and clinical data, 526 cancer samples, 59 control samples, and 572 intersection samples combined with survival information. The GSE68465, GSE31210, and GSE135222 expression profiling datasets of human LUAD were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). The basic clinical data of the above patients with LUAD were shown in Supplementary Figure S1.

2.2 Correlation analysis and group comparison analysis

The ssGSEA (single sample gene set enrichment analysis) analysis method was used to obtain the macrophage enrichment score, and further conducted a Pearson correlation analysis combined with the expression of PRs in samples. Then, according to the median expression value of phagocytic regulatory factors, we grouped them into high and low expression groups, and plotted the corresponding boxplot.

2.3 Enrichment analysis

ClusterProfiler package was used for GO and KEGG analysis, and its filtering parameter was pAdjustMethod ='none', pvalueCutoff = 0.05, qvalueCutoff = l.

2.4 Analysis of gene expression differences

Differences in gene expression analysis by using DEseq2 in the R package, |log2FC| ≥0.8 and padj<0.05 was used as the screening standard.

2.5 Screening of prognostic-related factors and construction and evaluation of prognostic risk scoring model

According to the influence of the expression level of differential expression factors on the survival time of patients, the univariate Cox regression model was used to identify the prognostic factors, and the threshold was selected as p<0.01. In addition, Lasso-logistic regression was used to remove redundant factors and further screen for prognostic factors. For the screened prognostic factors, the risk score was calculated according to the risk ratio regression coefficient in the multi-factor Cox regression model and its expression level, and finally, the prognostic risk Score model was constructed. Then, the samples were grouped according to the median risk score, and the correlation between risk score and patient survival time was analyzed by log-rank test. The ROC curve was drawn by R-package timeROC to evaluate the prognostic efficacy of the risk model. At the same time, GEO data were used for verification analysis.

In detail, for the Cox analysis, the coxph function of the survival package was used for the Cox analysis of samples and corresponding genes. Cox analysis can be divided into univariate and multivariate Cox regression analyses. In univariate Cox regression analysis, target genes were treated as independent factors affecting prognosis for regression analysis, and the risk score and significant degree of each gene were calculated. However, in multivariate Cox regression analysis, target genes are treated as cofactors associated with each other. By analyzing the multivariate Cox regression coefficient of each gene, the sum of the product of the multivariate Cox regression coefficient and the expression level of the corresponding gene was used as the risk value to measure the risk degree of the sample. Cox analysis was performed under the default parameters of the coxph function, and the significant degree was p<0.05 as the standard.

Risk score$_i$ =

$$\sum_{j=1}^{n} Cj* \exp ij \quad (1)$$

This formula calculated the Risk score value of the ith sample. Where C_j was the regression coefficient of the jth prognostic factor in the Cox regression model, exp_{ij} was the expression of the jth prognostic factor in the ith sample.For Lasso-logistic regression analysis, we used the glmnet function of the glmnet package to perform Lasso analysis on samples and corresponding genes. The parameters used in the lasso analysis were alpha=1, nlamba=100, and the significance degree was P<0.05 as the standard.

2.6 Immunoinfiltration analysis

Tumor immune invasion analysis was performed based on TCGA gene expression data by using the cibersort package to analyze the proportion of tumor immune cells in samples with cibersort default parameters. The results of immune infiltration calculated by the TIMER algorithm and XCELL algorithm were obtained by TIMER2.0 (http://timer.cistrome.org/) online analysis website.

2.7 Evaluation of immune score and gene score

The immunological scores and genetic scores were performed using the R package ESTIMATE under default parameters.

2.8 Cell culture

The human lung cancer cell lines (H1299, A549) were obtained from the Shanghai Zhong Qiao Xin Zhou Biotechnology Co., Ltd. The normal lung cell lines (BEAS-2B) were purchased from Hunan Fenghui Biotechnology Co., Ltd. H1299 and A549 were maintained in RIPM-1640 medium supplemented with 10% fetal bovine serum and antibiotics. BEAS-2B was maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum and antibiotics. All cells were incubated in 5% CO^2 at 37°C.

2.9 Quantitative real-time PCR and immunohistochemistry

The total RNA was extracted with RNA isolation Kit V2 (Vazyme, RC112) according to the product protocol. A reverse transcription reaction was carried out to acquire cDNA to prepare for the quantitative real-time PCR with the ABScript III RT Master Mix (ABclonal, RK20428). qPCR was cycled with the quantitative real-time gene amplification instrument (Jena qTower 3g) using 2X SYBR green Fast qPCR Mix (ABclonal, RK21205). Primers for GNPNAT1, SASH3, KIF23, FURIN, and ITGAL (Tsingke Biotechnology Co., Ltd) were listed in Supplementary Table S1. GADPH was used as endogenous control and further analyzed by the 2ΔddCt method. The amplification efficiency was assessed by the standard curve. The experiment was repeated three times. Further, we analyzed the immunohistochemical results of FURIN, KIF23, SASH3, GNPNAT1 and ITGAL by screening the HPA database (https://www.proteinatlas.org/).

3 Results

3.1 PRs could regulate macrophage phagocytosis and participate in the occurrence and development of lung adenocarcinoma

3.1.1 PRs could regulate the phagocytosis of macrophages

The analysis workflow is shown in Figure 1. We first obtained 183 PRs (Supplementary Table S2) from the two genome-wide CRISPR articles (14, 15), among which 178 genes existed in the TCGA dataset we used. Pearson correlation analysis showed that 118 genes had significant differences among these PRs, among which 69 were positively correlated and 49 were negatively correlated (Supplementary Figure S2). The group comparison result showed that a total of 92 PRs were significantly different, among which 43 PRs were significantly enriched in the high expression group. On the contrary, 49 PRs were significantly enriched in the low expression group (Supplementary Figure S3). Together, these results demonstrated that the most PRs could regulate the phagocytosis of macrophages and participate in the development of LUAD. Figure 2 showed the correlations and boxplots of partial PRs.

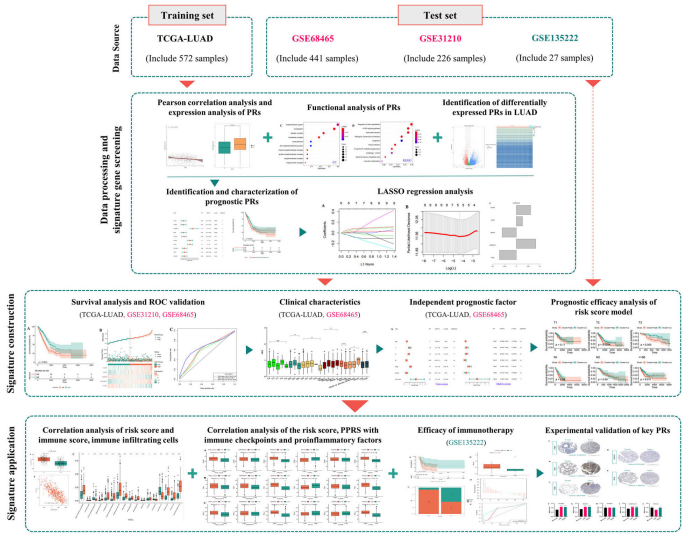


FIGURE 1
The analysis flowchart for identification and validation of a novel phagocytosis regulators-related signature with potential prognostic and immunotherapeutic value in LUAD.

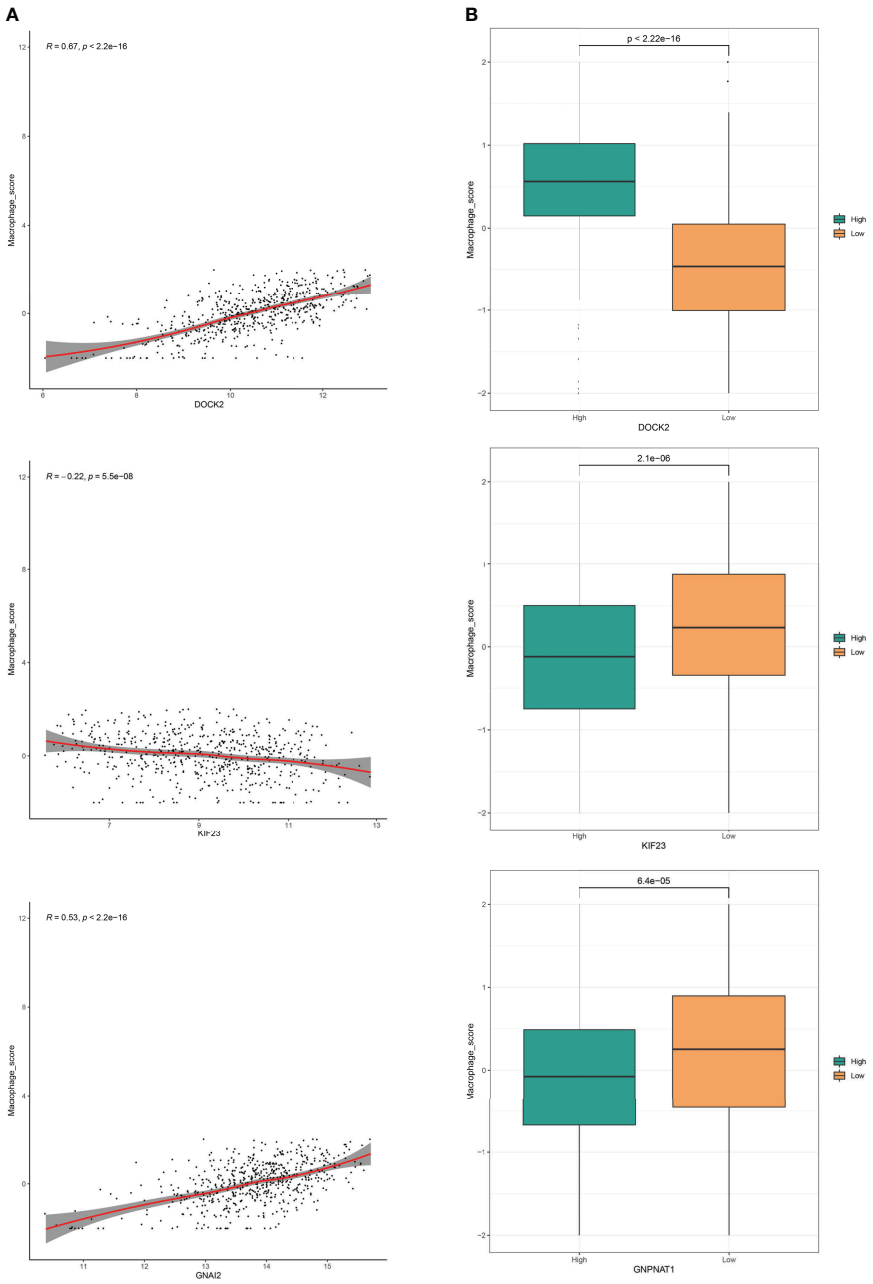


FIGURE 2
Phagocytosis regulators could regulate the phagocytosis of macrophages. **(A)** Correlation analysis of phagocytosis regulators expression level and macrophage enrichment score in LUAD samples. **(B)** Grouping comparison of phagocytosis regulators and macrophage enrichment scores.

3.1.2 Functional analysis of PRs

To fully explore the potential functions and pathways of PRs, we further performed GO and KEGG enrichment analyses. The ClusterProfiler package was used for GO and KEGG analysis of PRs, and the results were shown in Figure 3. The results showed that these phagocytic regulatory factors were significantly different in cellular protein modification process (Figure 3A), transferase activity (Figure 3B), enzyme binding (Figure 3B), endomembrane system (Figure 3C), regulation of actin cytoskeleton (Figure 3D) and mTOR (Figure 3D) related cell functions and signaling pathways.

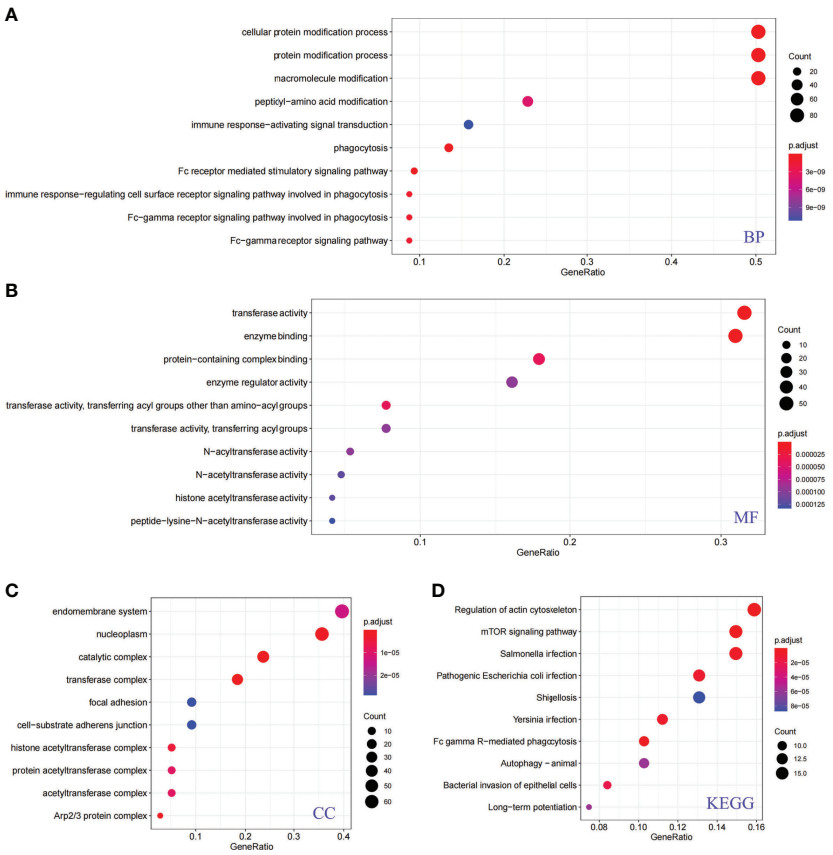


FIGURE 3
Functional analysis of phagocytosis regulators. **(A)** Biological process. **(B)** Molecular function. **(C)** Cellular component. **(D)** KEGG pathway.

3.1.3 Identification of differentially expressed PRs in LUAD

In order to further identify the differentially expressed PRs in LUAD, the DEseq2 package was used for differential expression analysis of the TCGA database, and the volcano maps and expression heat maps (R package pheatmap) were drawn. Through the analysis of the volcano map (Figure 4A) and heat map (Figure 4B), we found that among all differentially expressed genes in LUAD, a total of 29 PRs were present, including 13 up-regulated genes and 16 down-regulated genes.

3.2 Identification and characterization of prognostic PRs

To determine whether the differential expression of PRs is related to the prognosis of LUAD patients, based on the expression data and survival information of LUAD in TCGA, we further used differentially expressed PRs for univariate Cox screening. Finally, we obtained 10 prognostic PRs (Figure 5A), among which GMPNAT1, KIF23 and DOCK2 genes were independent prognostic genes (Figures 5B-D, Supplementary Figure S4).

3.3 Construction of prognostic PRs-related signature model

We selected 10 prognostic PRs as research objects, and further reduced the number of target gene sets by Lasso analysis. As a result, we constructed a signature model with 5 significant prognostic PRs through Lasso analysis (Figures 6A, B), and the formula of the optimal model is as follows: risk score = (0.06062) * FURIN + (0.02622) * KIF23 + (-0.05976) * SASH3 + (0.08207) * GNPNAT1 + (-0.07163) * ITGAL (Figure 6C).

3.4 Risk score was correlated with the prognosis and clinical features of LUAD patients

3.4.1 Risk score could predict the outcome of LUAD patients

To test whether the final model is stable, Further, we divided the samples into high-low risk groups based on the median of risk score, and conducted a survival analysis test and ROC

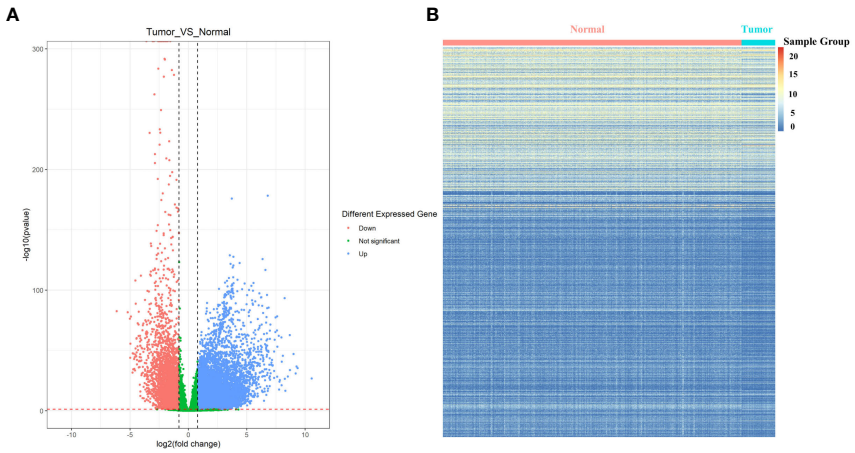


FIGURE 4
Identification of differentially expressed phagocytosis regulators in LUAD. **(A)** Volcano map of differentially expressed genes. **(B)** Expression heat map of differentially expressed genes. The normal sample shown here has more dark color changes than the tumor sample.

verification in the training set (TCGA) (Figures 7A-C) and two validation sets: GSE31210 (Figures 7D-F) and GSE68465 (Figures 7G-I) respectively to evaluate the prognostic efficacy of risk score in LUAD patients. Surprisingly, our results showed that the survival analysis (Figures 7A, D, G) and ROC curve (Figures 7C, F, I) of the three data sets proved to be meaningful, and we preliminarily believed that the risk prediction score had a certain ability to predict risk.

3.4.2 Risk score was correlated with clinical characteristics of LUAD patients

Based on clinical information from the TCGA dataset and the validation set (GSE68465), we assessed the relationship between risk score and clinical characteristics (including age, sex, TNM stage, race, etc.). The results showed that in the TCGA data set, the clinical parameters M stage, N stage, sex, and race were significantly correlated with risk score (Figure 8A). While in the GEO data set, sex and T stage were significantly correlated with risk score (Figure 8B). By analyzing the correlation between risk score and clinical characteristics of LUAD patients, we found that the sex was most associated with risk score.

3.4.3 Risk score could be an independent prognostic factor of LUAD patients

According to the clinical information of TCGA (Figure 9A) and GSE68465 (Figure 9B) data sets, univariate Cox analysis, and multivariate Cox analysis were performed on the two data sets respectively. The results showed that the *P* values of risk scores of both univariate Cox analysis and multivariate Cox analysis were less than 0.05 in these two data sets (Figure 9). Therefore, we can consider the risk score of this model as an independent prognostic factor for LUAD patients.

3.4.4 Prognostic efficacy analysis of risk score model

To evaluate the effectiveness and stability of the risk scoring model, we further analyzed the TCGA dataset. By extracting the clinical characteristics of LUAD samples from this dataset, including age, sex, stage, and TNM stage. Based on the above clinical information, the above risk scoring model was used to group the risk prediction of LUAD samples and compare the prognosis between groups. It is not difficult to find that risk score model with various clinical information, such as age, gender, sample stage and sample TMN stage, have significant inter-group prognostic differences in these clinical characteristics (Figure 10). These clinical features are also indicative clinical features of LUAD. The above results show that the prediction efficiency and stability of this model are high.

3.5 PRs signature was related to the immune microenvironment and immunotherapy of LUAD patients

3.5.1 Correlation analysis of risk score and immune score, immune infiltrating cells

To determine the relationship between risk score and the tumor immune microenvironment, we first used the ESTIMATE method to calculate the ImmuneScore and EstimateScore scores of LUAD samples, and then we combined with the risk scores of LUAD samples to analyze the expression difference between the two types of scores (ImmuneScore and EstimateScore) in the case of high and low-risk groups (Figure 11A), as well as the correlation

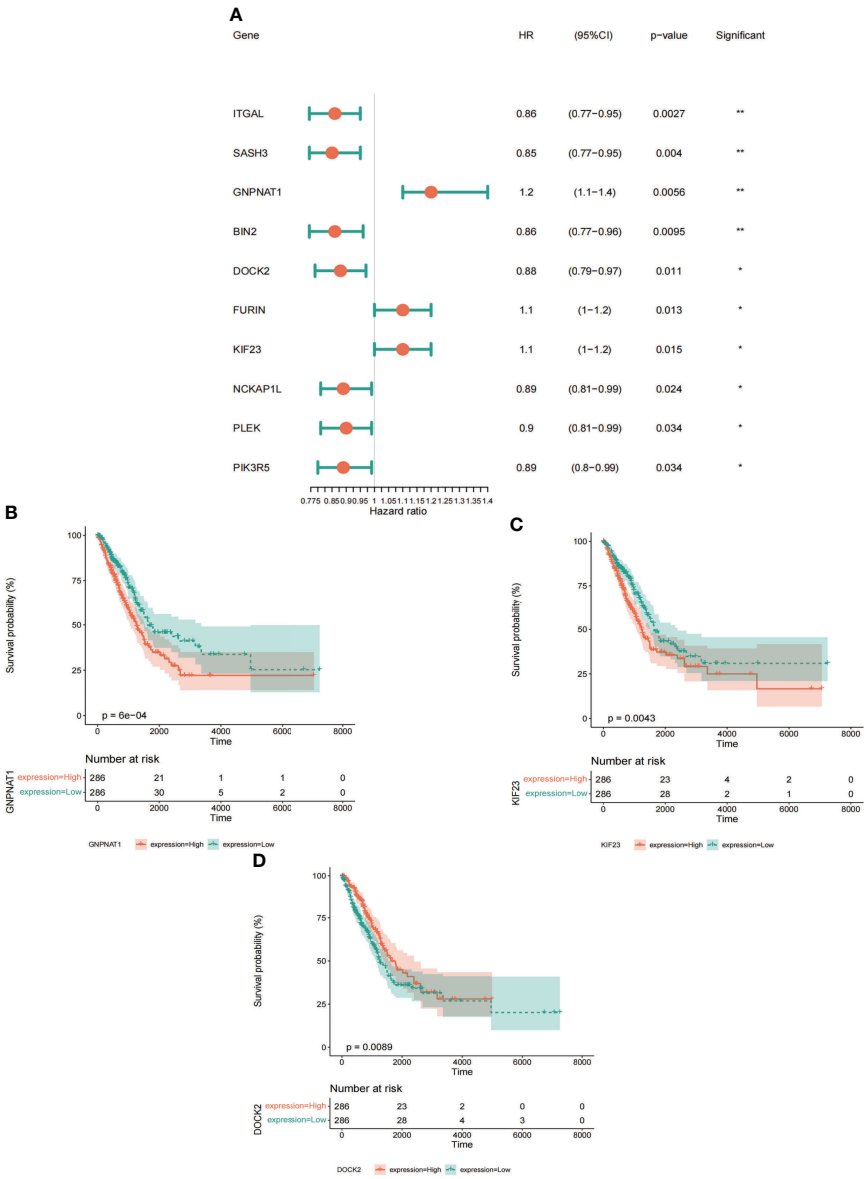


FIGURE 5
Identification and characterization of prognostic phagocytosis regulators. **(A)** Univariate Cox regression analysis of prognostic phagocytosis regulators in LUAD patients. **(B-D)** Kaplan-meier curves of phagocytosis regulators (Including GMPNAT1, KIF23 and DOCK2). *p < 0.05, **p < 0.01.

between the two types of scores and risk scores (Figure 11B). Excitingly, we found that the enrichment scores of these two types of immunity in the high-risk group were significantly lower than those in the low-risk group. moreover, correlation analysis showed that the two types of immunity scores were significantly negatively correlated with the risk score.

In addition, we also used TIMER2.0 (http://timer.cistrome.org/) online analysis website to obtain the immune infiltration of LUAD samples calculated by the TIMER algorithm and XCELL algorithm. Meanwhile, the Cibersort algorithm was also used to calculate the immune infiltration of LUAD samples. And the results showed that the immune infiltration scores obtained by the XECLL algorithm were different among groups under the high and low-risk group of LUAD samples, including macrophages, myeloid dendritic cell, B cell, T cell $CD4^+$, T cell $CD8^+$, etc. (Figure 11C). Similarly, the two other algorithms (TIMER algorithm and Cibersort algorithm) also showed similar results (Supplementary Figure S5, Supplementary Figure S6). In conclusion, our results showed that there was a strong correlation between these risk scores and immune scores and immune infiltrating cells, especially in macrophages.

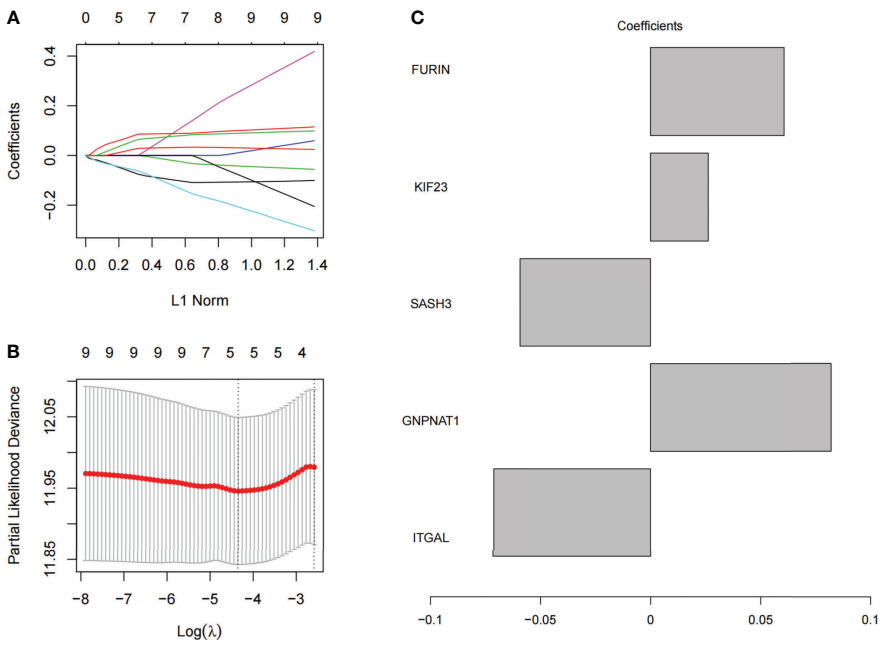


FIGURE 6
Model construction of phagocytosis regulators signature. **(A)** The coefficient of 10 prognostic phagocytosis regulators signature. **(B)** the partial likelihood deviance of prognostic phagocytosis regulators signature. **(C)** The coefficient display of the 5 phagocytosis regulators signature in the optimal combination model.

3.5.2 Correlation analysis of the risk score, PPRS with immune checkpoints and proinflammatory factors

According to the risk score of LUAD samples and the high and low expression of the five prognostic PRs that constitute the model, and LUAD patients were divided into two subgroups, the high expression group and the low expression group. Subsequently, the correlation analysis of the risk score and the five prognostic phagocytosis regulators signature (PPRS) with immune checkpoints (PD-1, PD-L1, CTLA4) and proinflammatory factors (IL-1α, IL-1β, IL-6, IL-8 and IL-18) was conducted. we found that CTLA4 was not significantly different in the high and low groups of FURIN, while there was a significant difference between the groups with high and low expression of other factors (Figure 12A). In addition, the PD-1 expression level was negatively correlated with a risk score, and positively correlated with GNPNAT1, SASH3, KIF23, FURIN, and ITGAL (Figure 12B). The PD-L1 expression level was negatively correlated with risk score and FURIN, and positively correlated with GNPNAT1, SASH3, KIF23 and ITGAL (Figure 12C). By analyzing the correlation between major pro-inflammatory factors and risk score and the five PPRS, we found that the most pro-inflammatory factors (IL-1α, IL-1β, IL-6, IL-8 and IL-18) were significantly correlated with the risk score and the five PPRS (Supplementary Figure S7).

3.5.3 The model could predict the efficacy of immunotherapy in LUAD patients

Model prediction score was performed for GSE135222 dataset containing immunotherapy response, and the survival analysis curve and ROC verification curve were drawn according to the PFS information of the dataset. In addition, the correlation bars were plotted based on immunotherapy responses (DCB: durable clinical benefit, including LUAD patients with CR, PR and SD>6 months; NDB: non-durable clinical benefit, including LUAD patients with SD ≤ 6 months and PD) in the data set, and the difference of risk score under different immunotherapy response subgroups were analyzed. We found that in the GSE135222 data set, the high-risk group had a worse prognosis than the low-risk group (Figure 13A), which was consistent with the previous results. The risk score was higher in the NDB immunotherapy response than in the DCB immunotherapy response (Figure 13B), and NDB was the largest proportion in the high-risk group (Figure 13C). Further ROC curve verification showed that the areas under the ROC curve at 120, and 300 days were 0.67 and 0.8, respectively (Figures 13D, E).

3.6 Experimental validation of key PRs

To confirm that PRs play an important role in LUAD, we further compared their expression in normal lung epithelial cells

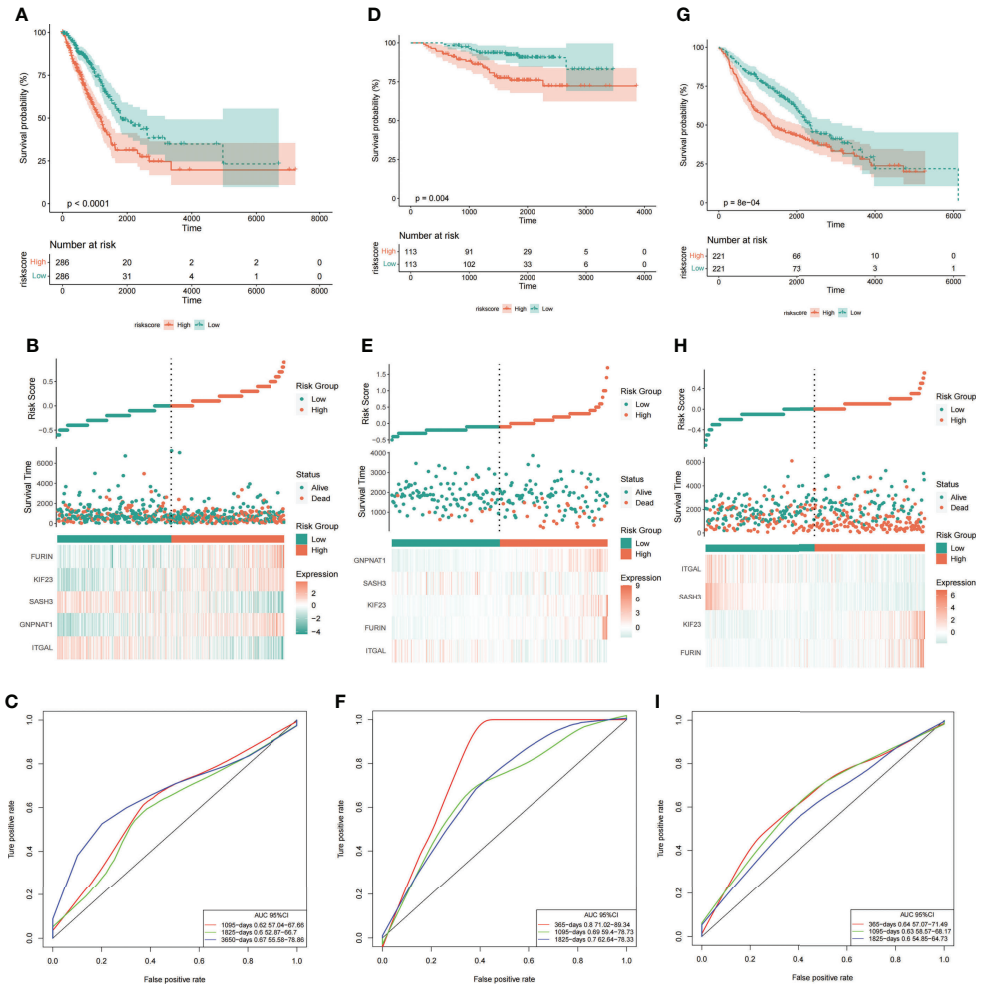


FIGURE 7
Risk score could predict the outcome of LUAD patients. Survival analysis **(A, D, G)**, risk score and grouping of samples **(B, E, H)**, and ROC curve **(C, F, I)** in the training set (TCGA) and two validation sets (GSE31210 and GSE68465). From left to right are the TCGA, GSE31210 and GSE68465 datasets respectively.

(BEAS-2B) and different lung adenocarcinoma cell lines (A549 and H1299). The results showed that the expression of these 5 genes in the above cell lines was consistent with the expression trend in LUAD patients, and 3 of them had significant differences (Figures 14A-E). Further, by analyzing the results of immunohistochemistry, we could qualitatively observe significant differences in the protein expression levels of the three PRs (FURIN, KIF23 and GNPNAT1) in normal lung tissues and LUAD samples (Figures 14F-J).

4 Discussion

LUAD is one of the most serious malignant tumors threatening human life, health and quality of life in the world. At present, immunotherapy is the main frontier in the treatment of LUAD patients (16). Macrophages are key drivers of tumor inflammation, and TAM promotes tumor progression at different levels (8, 9). Phagocytosis is required for a variety of physiological functions, from pathogen defense to tissue homeostasis (14). Therefore, it is of great significance to identify and characterize phagocytosis regulatory factors and clarify their roles in LUAD prognosis and immunotherapy response for the prognosis and treatment of LUAD patients. Gene signature is a biological function model constructed from the expression data of multiple genes, which can be used to predict the prognosis and progression of many types of malignant tumors (17). Yi et al. constructed a prognostic model of 17- immune-related genes signature to predict survival and response to ICI (immune checkpoint inhibitors) in LUAD patients and the result showed that patients with a low-risk score had a better prognosis and predicted benefit from ICI

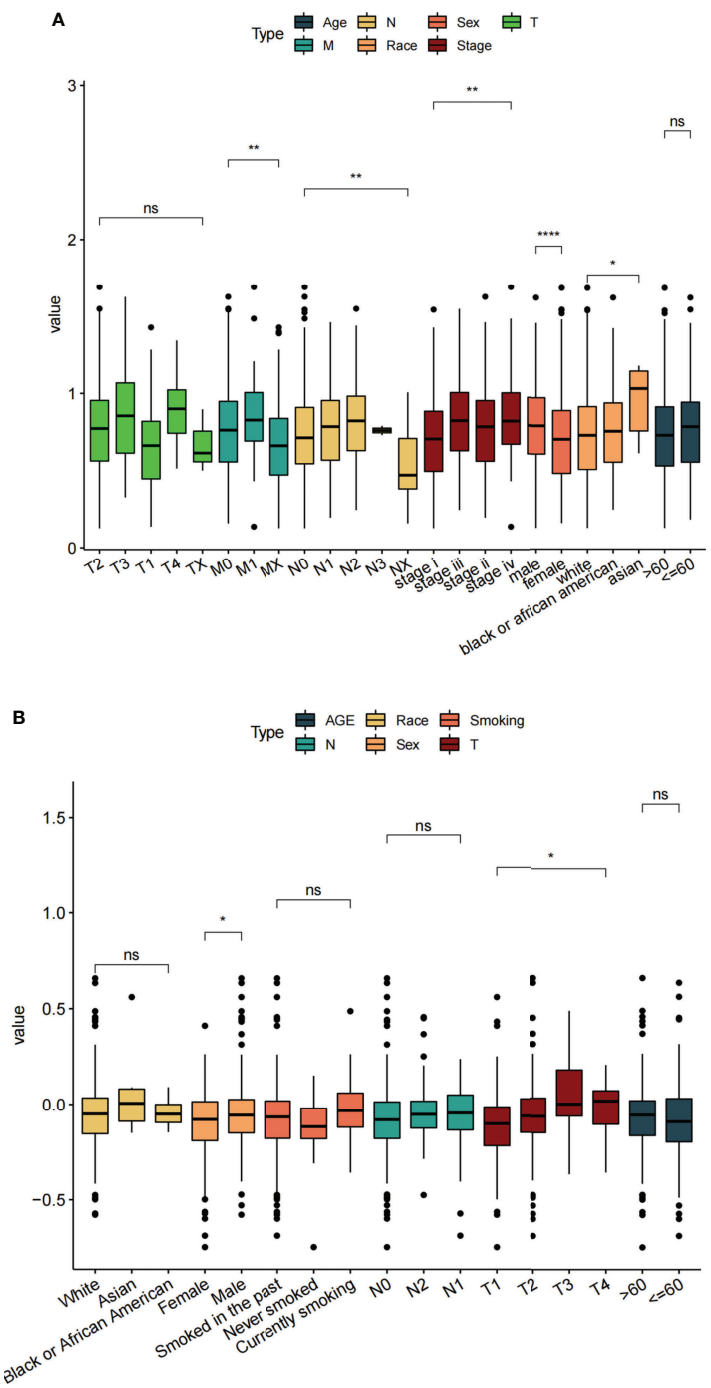


FIGURE 8
Correlation between different clinical features and risk score. **(A)** TCGA datasets. **(B)** GSE68465 datasets. ns: $p > 0.05$, *$p < 0.05$, **$p < 0.01$, ****$p < 0.0001$.

treatment (18). Sun et al. established a four-gene signature named IPSLUAD (ARNTL2, ECT2, PPIA, and TUBA4A) using stability selection and Lasso COX regression. It has been proved that it has good performance in multiple LUAD queues (19). The construction of this gene signature models provides a certain reference for the prognosis and immunotherapy of LUAD patients. However, the interaction between PRs and macrophages from the perspective of phagocytosis regulators has not been explored to predict the prognosis and therapeutic effect of lung adenocarcinoma patients. In our study, we

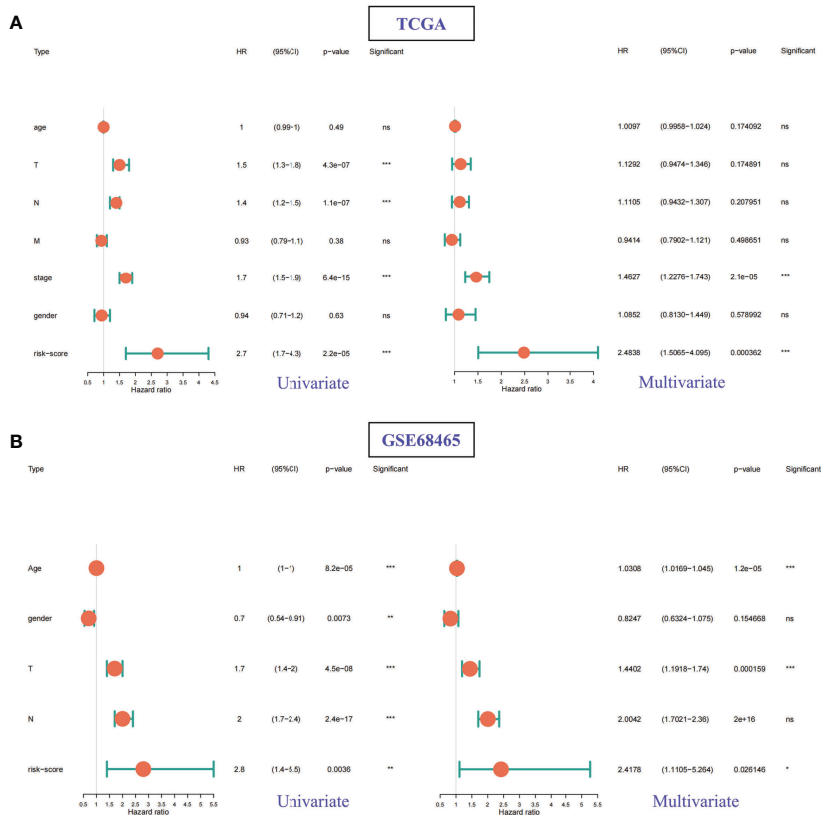


FIGURE 9
Univariate Cox and multivariate Cox regression analysis for different clinical features and risk scores. **(A)** TCGA datasets. **(B)** GSE68465 datasets. ns: $p > 0.05$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

constructed a novel PRs-related signature model containing five genes. Further, we explored the interactions of these signatures with macrophages and verified their accuracy in predicting patient outcomes and therapeutic effects.

We first demonstrated that most of the 183 PRs that we identified in the literature could modulate macrophage phagocytosis in LUAD samples. Then, we analyzed the functional enrichment of these PRs and found that these phagocytic regulatory

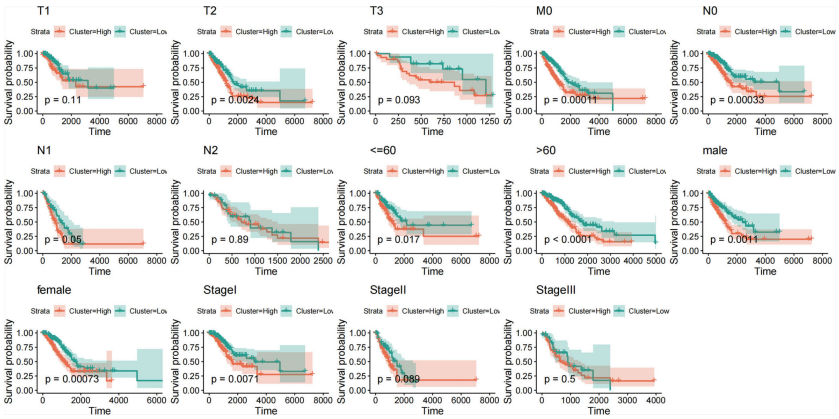


FIGURE 10
Prognostic efficacy and stability assessment of the risk score model.

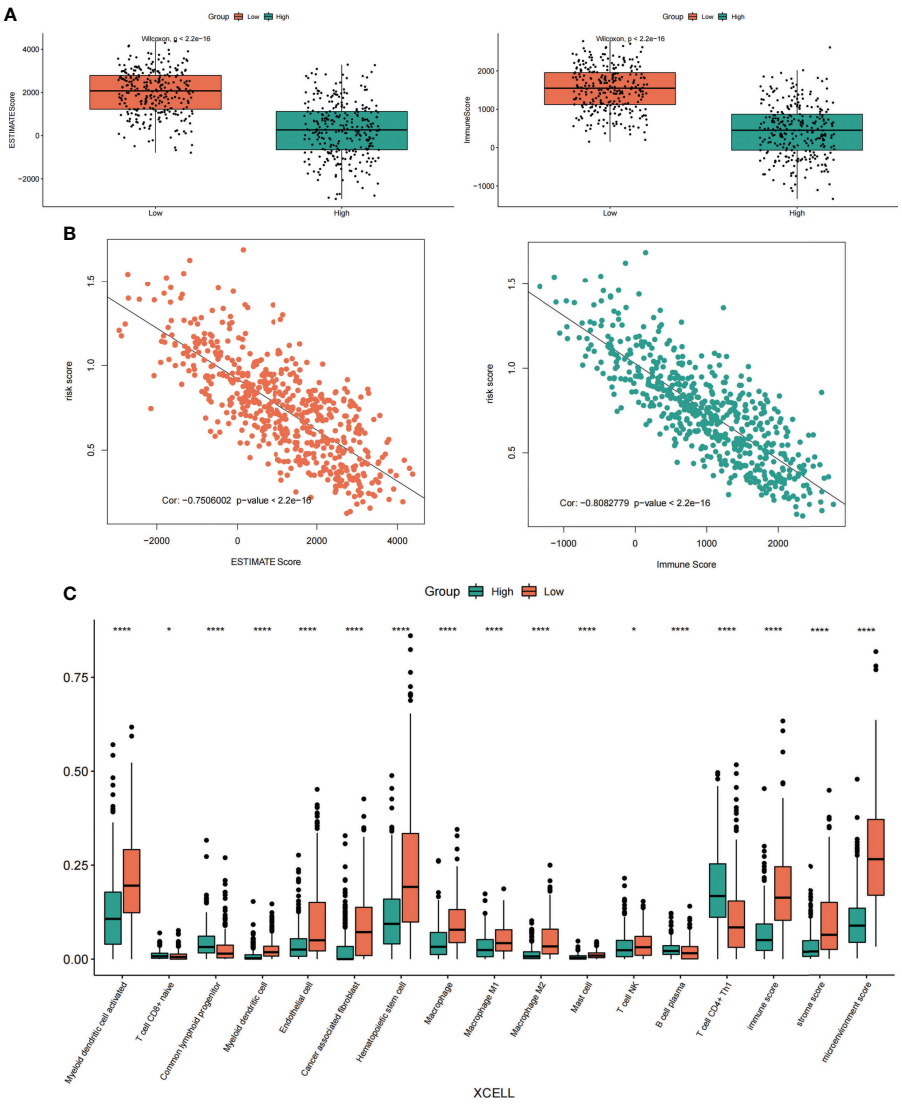


FIGURE 11
Correlation analysis of risk score, immune score and immune infiltrating cells. **(A)** Differential expression of ImmuneScore and EstimateScore scores in the high-low risk groups of LUAD samples. **(B)** Correlation analysis between ImmuneScore and EstimateScore scores and LUAD sample risk score. **(C)** The expression differences of different immune cell infiltrates between the high and low-risk groups were analyzed by the XCELL algorithm. *p < 0.05, ****p < 0.0001.

factors were mainly enriched in cellular protein modification process enzyme binding, regulation of actin cytoskeleton and mTOR, and other related functions and pathways. As pleiotropic cells, macrophages can undertake a variety of functions according to the tissue they inhabit and the state of the tissue. Studies have confirmed that mTOR plays a key role in the activation of macrophages, especially its ability to control the activation and metabolic process of macrophages (20). Further, we analyzed the differential expression of these phagocytic regulatory factors in LUAD samples, and finally, we identified a total of 29 differential expression phagocytic regulatory factors. Ten prognostic PRs were further screened by univariate Cox regression analysis, including ITGAL, SASH3, GNPNAT1, BIN2, DOCK2, FURIN, KIF23, NCKAP1L, PLEK and PIK3R5. Among them, GMPNAT1, KIF23 and DOCK2 could be used as independent prognostic genes.

Subsequently, we constructed a model with five prognostic signatures (FURIN, KIF23, SASH3, GNPNAT1, and ITGAL). Glucosamine 6-phosphateN-acetyltransferase 1 (GNPNAT1) is a key enzyme in the hexosamine biosynthetic pathway, which is abnormally expressed in tumor cells and promotes tumor progression and metastasis (21). Recently, studies have shown that abnormal expression of GNPNAT1 is related to the carcinogenesis of lung cancer. Wang et al. showed that the mRNA and protein expression levels of GNPNAT1 in lung

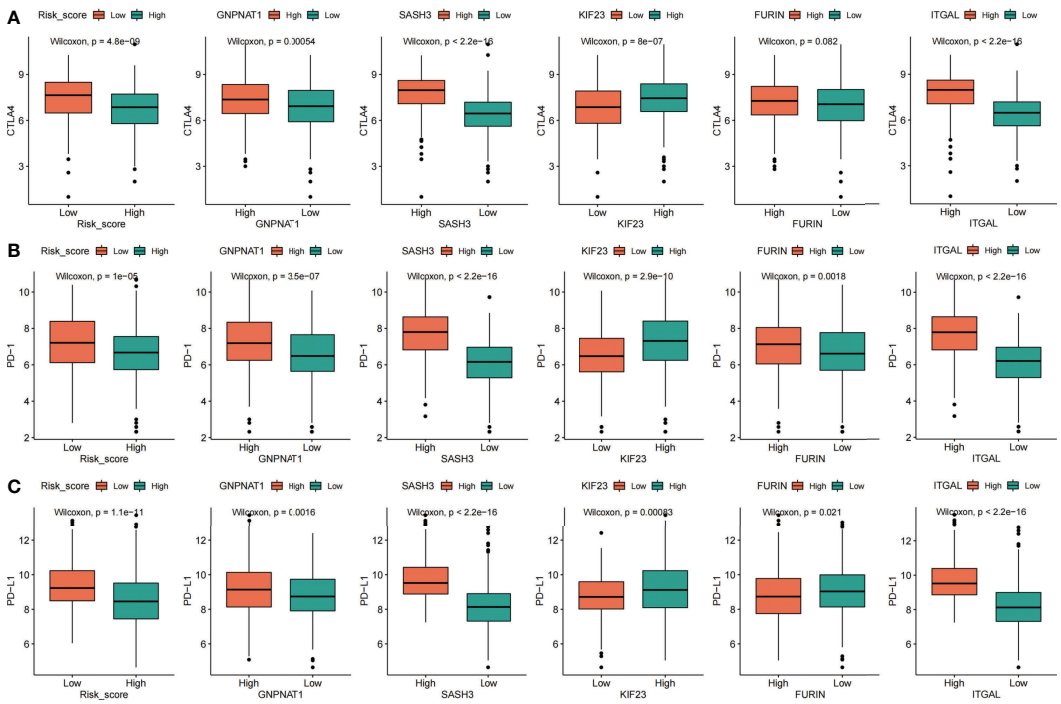


FIGURE 12
Correlation analysis between risk score, five prognostic phagocytosis regulators signatures and immune checkpoints. **(A)** CTLA4. **(B)** PD-1. **(C)** PD-L1.

cancer tissues were upregulated compared with normal tissues (22, 23). The expression level of GNPNAT1 is related to the clinical stage and prognosis of lung cancer. Patients with high expression of GNPNAT1 are more likely to develop advanced lung cancer, with a poor prognosis and low survival rate (24). In addition, GNPNAT1 is associated with immune infiltration in lung cancer, which has a converse correlation with infiltration of B cells, $CD4^+$ T cells, and dendritic cells, all of which have antitumor effects in NSCLC (25). In particular, B cell infiltration may be one of the key reasons that caused GNPNAT1 to become a prognostic factor (26). Meanwhile, in the study of the crosstalk between GNPNAT1 and immune genes, it was found that both immunosuppressive genes (LGALS9, TGFB1, CD160, CSF1R, and CD96) and immunostimulatory genes (CD40LG, CD48, IL6R, CD27, CD40, CXCR4, LTA, CXCL12, and CD28) were negatively correlated with GNPNAT1 (27). Moreover, GNPNAT1 is associated with immune signaling and is involved in lung cancer immune evasion. In conclusion, GNPNAT1 can be used as a potential prognostic biomarker and a new immunotherapy target for lung cancer. Kinesin family member 23 (KIF23), also known as MKLP1, is a key regulator of cytokinesis, transporting organelles within cells and moving chromosomes during cell division (28). It has been observed in a variety of human malignancies and is considered as a potential tumor marker. KIF23 overexpression was recently shown in lung cancer, and was associated with a low survival rate in lung cancer patients (29). Vikberg et al. found that the elevated level of KIF23 in lung cancer may be due to the extra copy of chromosome 15, and that KIF23 plays a crucial role in the last step of mitosis, so this gene is a potential molecular marker for lung cancer treatment (30). Based on this, some studies further confirmed that RNA interference-mediated KIF23 deletion can effectively inhibit lung cancer cell growth and lung tumor formation *in vivo*, and induce apoptosis of lung cancer cell lines (31, 32). In addition, KIF23 is a potential key gene regulating hypoxia-induced tumor cell stemness in the immune microenvironment of lung tumors (33). Meanwhile, the expression of KIF23 is significantly correlated with B cell infiltration, and the interaction between KIF23 expression and B cell infiltration plays an important role in the immune response and prognosis of lung cancer, which affects the tumor microenvironment and the tumor immune effect of patients (34). Furin (PCSKC3), as an important member of proprotein processing enzyme, is overexpressed in numerous human malignancies. Furin can cut many proteins closely related to tumor development, such as Notch, Wnt, VEGF, etc., so the expression of Furin can be used as a marker of tumor progression (35). Brant et al. showed that the increased expression of Furin in lung cancer was significantly associated with a low overall survival rate (36). In view of this, several studies have shown that Furin inhibitors, such as a1-PDX, can inhibit the growth and migration of lung cancer cells, and have inhibitory effects on the growth and proliferation of xenograft

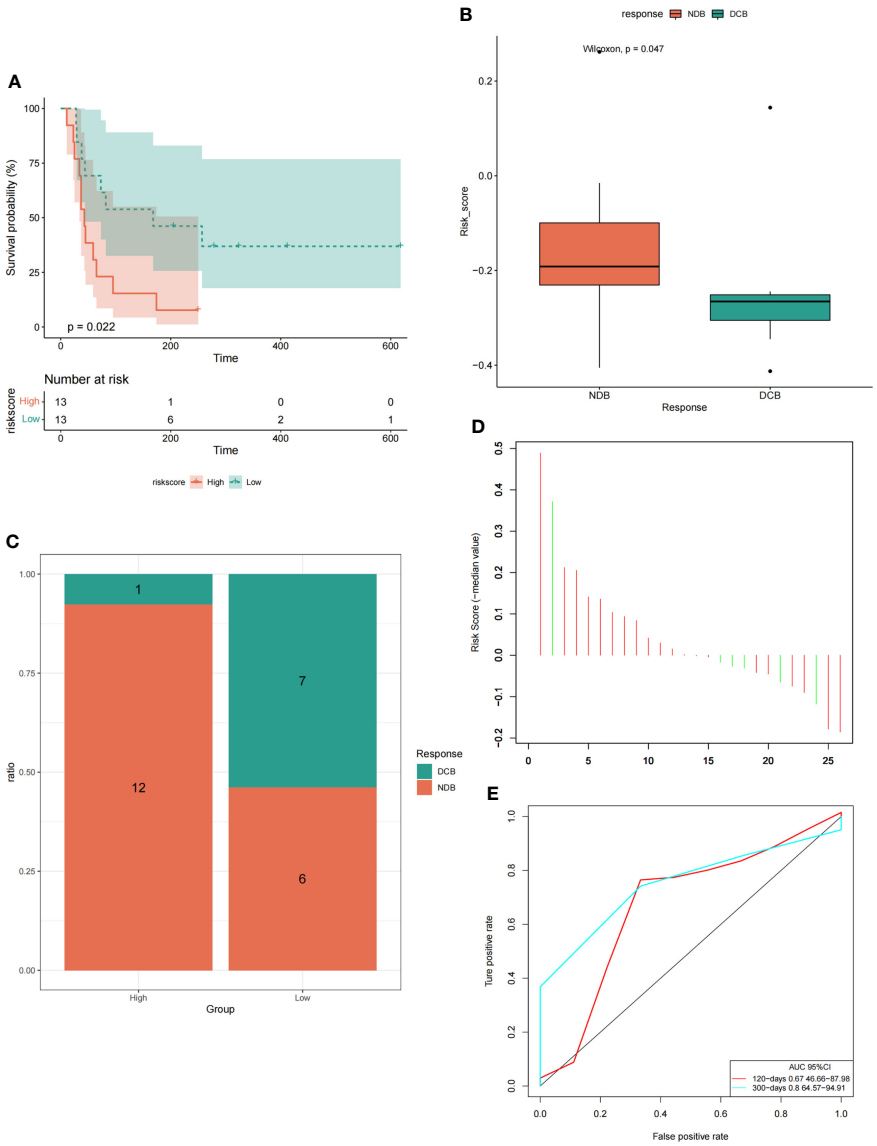


FIGURE 13
The model could predict the efficacy of immunotherapy in LUAD patients. **(A)** Survival curve analysis under high and low-risk groups. **(B)** Analysis of intergroup differences in risk scores under two immunotherapy responses. **(C)** The proportion of two immunotherapy responses in the high-low risk group. **(D)** The risk score of the LUAD samples. **(E)** ROC verification curve of the model.

human lung cancer cells (37, 38). In addition, Luo et al. found that the expression of Furin in lung cancer was significantly correlated with the expression of immunomarker genes in CD8 T cells, T cells, monocytes, TAM and dendritic cells, which confirmed the relationship between Furin expression and immune infiltration in lung cancer (39). In conclusion, Furin can be used as a promising biomarker for the judgment of immune invasion and treatment of lung cancer. SASH3 encodes a signal adapter protein that plays a role in many developmental processes, including cell growth and migration (40). Li et al. revealed that SASH3 was significantly correlated with the survival rate of lung cancer patients. Moreover, SASH3 is associated with gene regulatory sites (such as WAS and CD53) in lung cancer, which has diagnostic value for lung cancer metastasis (41). In addition, SASH3 was one of the co-expressed factors related to tumor purity, which was significantly negatively correlated with tumor purity, but positively correlated with $CD8^+$ T lymphocytes and immune score (42). Therefore, SASH3 has important clinical and biological significance in the microenvironment of lung cancer. ITG subunit alpha L (ITGAL) encodes the LFA-1 (aLb2) subunit of integrin, which is highly expressed in most immune cell populations (43). The results of current research indicate that ITGAL can be identified as a

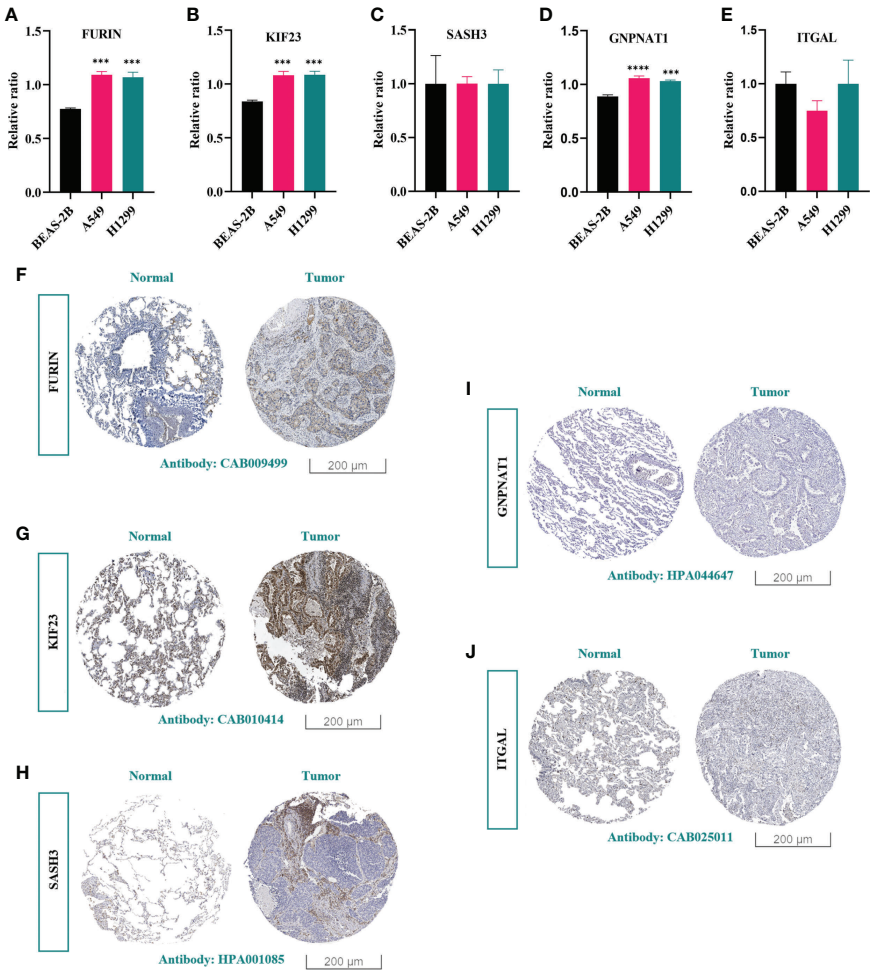


FIGURE 14
Experimental validation of key PRs. **(A-E)** The expression levels of 5 prognostic phagocytosis regulators in the model in BEAS-2B, A549 and H1299 cell lines were detected by RT-qPCR. ***: $p < 0.001$, ****: $p < 0.0001$. **(F-J)** Immunohistochemical results of FURIN, KIF23, SASH3, GNPNAT1 and ITGAL.

prognostic indicator of lung cancer, and patients with high expression of ITGAL have a better prognosis. Pathways enriched by high expression of ITGAL are mainly related to immune cell recognition and killing of lung cancer cells (44). In particular, ITGAL is strongly positively correlated with genes related to immune surveillance and recognition, such as CD3E and CD2 of $CD4^+$ T cells. The increased activity of immune cells may be the reason for the better prognosis of patients with high expression of ITGAL (45). In summary, it can be seen that these five prognostic signature genes may play an important role in LUAD. Therefore, we then carried out survival analysis and ROC curve verification through three data sets (TCGA, GSE31210, GSE68465). In addition, we also carried out clinical characteristics analysis, univariate and multivariate Cox analysis, and analyzed the prognostic efficacy of the model. The results fully prove that the model has high predictive efficiency and stability.

More and more tumor treatment strategies focus on reversing the immunosuppressive state of the tumor microenvironment, and macrophages are the key effector cells of innate immunity, whose main functions are phagocytosis and antigen presentation. Targeting TAMs can enhance tumor immune response. The development of effective phagocyte targets and the search for new innate immune checkpoints are important strategies to improve the response rate of tumor immunotherapy (46, 47). Therefore, we further analyzed the correlation between these five phagocytic factor signatures and the immune microenvironment and immunotherapy in LUAD patients. Through the above analysis, we further proved that the model composed of five PPRS could well predict the immunotherapy effect of LUAD patients. Our RT-qPCR results showed that there were significant differences in the expression of three genes (FURIN, KIF23, GNPNAT1) *in vitro*

cell experiments, and this was also confirmed by the results of immunohistochemistry.

5 Conclusions

Taken together, our study explored and constructed a new phagocytic regulatory signature-based model from the perspective of the combination of phagocytic regulatory factors and immunity. The model has also been verified by multiple analyses of training sets and validation sets, and its stability has been confirmed by analyses of immune infiltration, immune checkpoints and pro-inflammatory factors. More importantly, the model has also been validated by immunotherapy responses. In addition, we used cell experiments and clinical tissue samples to verify the gene and protein expression in the model. In brief, the model we constructed can well predict the prognosis and immunotherapy efficacy of LUAD patients. However, our current study did not further explore the specific regulatory roles and mechanisms of these genes in LUAD. Furthermore, the stability of this model needs to be verified by more clinical samples and experiments. In the future, we can conduct further exploration from the above perspectives.

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

SW and LX designed the topic; SW carried out the bioinformatics analysis; JL carried out the cell experiments and wrote the manuscript. QD and JS studied the relevant literature and reviewed the manuscript. LX reviewed and directed the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the National Natural Science Foundation of China (82104534, 81903902), and the Sichuan Province Science Research Special Funding Project for Postdoctoral Fellows (2022BSH044).

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.2022.988332/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. Ramalingam SS, Owonikoko TK, Khuri FR. Lung cancer: New biological insights and recent therapeutic advances. *CA Cancer J Clin* (2011) 61:91–112. doi: 10.3322/caac.20102

3. Bodor JN, Boumber Y, Borghaei H. Biomarkers for immune checkpoint inhibition in non-small cell lung cancer (NSCLC). *Cancer* (2020) 126:260–70. doi: 10.1002/cncr.32468

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

5. Vitale I, Manic G, Coussens LM, Kroemer G, Galluzzi L. Macrophages and metabolism in the tumor microenvironment. *Cell Metab* (2019) 30:36–50. doi: 10.1016/j.cmet.2019.06.001

6. Zhai K, Huang Z, Huang Q, Tao W, Fang X, Zhang A, et al. Pharmacological inhibition of BACE1 suppresses glioblastoma growth by stimulating macrophage phagocytosis of tumor cells. *Nat Cancer* (2021) 2:1136–51. doi: 10.1038/s43018-021-00267-9

7. Noy R, Pollard JW. Tumor-associated macrophages: from mechanisms to therapy. *Immunity* (2014) 41:49–61. doi: 10.1016/j.immuni.2014.06.010

8. Mantovani A, Marchesi F, Malesci A, Laghi L, Allavena P. Tumour-associated macrophages as treatment targets in oncology. *Nat Rev Clin Oncol* (2017) 14:399–416. doi: 10.1038/nrclinonc.2016.217

9. Pan Y, Yu Y, Wang X, Zhang T. Tumor-associated macrophages in tumor immunity. *Front Immunol* (2020) 11:583084. doi: 10.3389/fimmu.2020.583084

10. Brown BN, Sicari BM, Badylak SF. Rethinking regenerative medicine: A macrophage-centered approach. *Front Immunol* (2014) 5:510. doi: 10.3389/fimmu.2014.00510

11. Liang W, Luo Q, Zhang Z, Yang K, Yang A, Chi Q, et al. An integrated bioinformatics analysis and experimental study identified key biomarkers CD300A or CXCL1, pathways and immune infiltration in diabetic nephropathy mice. *BIOCELL* (2022) 46:1989–2002. doi: 10.32604/biocell.2022.019300

12. Wang S, Xing N, Meng X, Xiang L, Zhang Y. Comprehensive bioinformatics analysis to identify a novel cuproptosis-related prognostic signature and its ceRNA regulatory axis and candidate traditional Chinese medicine active ingredients in lung adenocarcinoma. *Front Pharmacol* (2022) 13:971867. doi: 10.3389/fphar.2022.971867

13. Xu R, Wu Q, Gong Y, Wu Y, Chi Q, Sun D. A novel prognostic target-gene signature and nomogram based on an integrated bioinformatics analysis in hepatocellular carcinoma. *BIOCELL* (2022) 46:1261–88. doi: 10.32604/biocell.2022.018427

14. Haney MS, Bohlen CJ, Morgens DW, Ousey JA, Barkal AA, Tsui CK, et al. Identification of phagocytosis regulators using magnetic genome-wide CRISPR screens. *Nat Genet* (2018) 50:1716–27. doi: 10.1038/s41588-018-0254-1

15. Kamber RA, Nishiga Y, Morton B, Banuelos AM, Barkal AA, Vences-Catalán F, et al. Inter-cellular CRISPR screens reveal regulators of cancer cell phagocytosis. *Nature* (2021) 597:549–54. doi: 10.1038/s41586-021-03879-4

16. Steven A, Fisher SA, Robinson BW. Immunotherapy for lung cancer. *Respirology* (2016) 21:821–33. doi: 10.1111/resp.12789

17. Li X, Dai Z, Wu X, Zhang N, Zhang H, Wang Z, et al. The comprehensive analysis identified an autophagy signature for the prognosis and the immunotherapy efficiency prediction in lung adenocarcinoma. *Front Immunol* (2022) 13:749241. doi: 10.3389/fimmu.2022.749241

18. Yi M, Li A, Zhou L, Chu Q, Luo S, Wu K. Immune signature-based risk stratification and prediction of immune checkpoint inhibitor's efficacy for lung adenocarcinoma. *Cancer Immunol Immunother* (2021) 70:1705–19. doi: 10.1007/s00262-020-02817-z

19. Sun S, Guo W, Wang Z, Wang X, Zhang G, Zhang H, et al. Development and validation of an immune-related prognostic signature in lung adenocarcinoma. *Cancer Med* (2020) 9:5960–75. doi: 10.1002/cam4.3240

20. Covarrubias AJ, Aksoylar HI, Horng T. Control of macrophage metabolism and activation by mTOR and akt signaling. *Semin Immunol* (2015) 27:286–96. doi: 10.1016/j.smim.2015.08.001

21. Zhao M, Li H, Ma Y, Gong H, Yang S, Fang Q, et al. Nanoparticle abraxane possesses impaired proliferation in A549 cells due to the underexpression of glucosamine 6-phosphate n-acetyltransferase 1 (GNPNAT1/GNA1). *Int J Nanomed* (2017) 12:1685–97. doi: 10.2147/ijn.s129976

22. Meng J, Cao L, Song H, Chen L, Qu Z. Integrated analysis of gene expression and DNA methylation datasets identified key genes and a 6-gene prognostic signature for primary lung adenocarcinoma. *Genet Mol Biol* (2021) 44:e20200465. doi: 10.1590/1678-4685-gmb-2020-0465

23. Wang Z, Embaye KS, Yang Q, Qin L, Zhang C, Liu L, et al. Establishment and validation of a prognostic signature for lung adenocarcinoma based on metabolism-related genes. *Cancer Cell Int* (2021) 21:219. doi: 10.1186/s12935-021-01915-x

24. Zhu P, Gu S, Huang H, Zhong C, Liu Z, Zhang X, et al. Upregulation of glucosamine-phosphate n-acetyltransferase 1 is a promising diagnostic and predictive indicator for poor survival in patients with lung adenocarcinoma. *Oncol Lett* (2021) 21:488. doi: 10.3892/ol.2021.12750

25. Liu W, Jiang K, Wang J, Mei T, Zhao M, Huang D. Upregulation of GNPNAT1 predicts poor prognosis and correlates with immune infiltration in lung adenocarcinoma. *Front Mol Biosci* (2021) 8:605754. doi: 10.3389/fmolb.2021.605754

26. Zhang J, Zhang J, Yuan C, Luo Y, Li Y, Dai P, et al. Establishment of the prognostic index reflecting tumor immune microenvironment of lung adenocarcinoma based on metabolism-related genes. *J Cancer* (2020) 11:7101–15. doi: 10.7150/jca.49266

27. Zheng X, Li Y, Ma C, Zhang J, Zhang Y, Fu Z, et al. Independent prognostic potential of GNPNAT1 in lung adenocarcinoma. *BioMed Res Int* (2020) 2020:8851437. doi: 10.1155/2020/8851437

28. Ye L, Li H, Zhang F, Lv T, Liu H, Song Y. [Expression of KIF23 and its prognostic role in non-small cell lung cancer: Analysis based on the data-mining of oncomine]. *Zhongguo Fei Ai Za Zhi* (2017) 20:822–6. doi: 10.3779/j.issn.1009-3419.2017.12.05

29. Wu X, Sui Z, Zhang H, Wang Y, Yu Z. Integrated analysis of lncRNA-mediated ceRNA network in lung adenocarcinoma. *Front Oncol* (2020) 10:554759. doi: 10.3389/fonc.2020.554759

30. Vikberg AL, Vooder T, Lokk K, Annilo T, Golovleva I. Mutation analysis and copy number alterations of KIF23 in non-small-cell lung cancer exhibiting KIF23 over-expression. *Onco Targets Ther* (2017) 10:4969–79. doi: 10.2147/ott.s138420

31. Kato T, Wada H, Patel P, Hu HP, Lee D, Ujiie H, et al. Overexpression of KIF23 predicts clinical outcome in primary lung cancer patients. *Lung Cancer* (2016) 92:53–61. doi: 10.1016/j.lungcan.2015.11.018

32. Iltzsche F, Simon K, Stopp S, Pattschull G, Francke S, Wolter P, et al. An important role for myb-MuvB and its target gene KIF23 in a mouse model of lung adenocarcinoma. *Oncogene* (2017) 36:110–21. doi: 10.1038/onc.2016.181

33. Guo L, Li H, Li W, Tang J. Construction and investigation of a combined hypoxia and stemness index lncRNA-associated ceRNA regulatory network in lung adenocarcinoma. *BMC Med Genomics* (2020) 13:166. doi: 10.1186/s12920-020-00816-8

34. Chen W, Li W, Liu Z, Ma G, Deng Y, Li X, et al. Comprehensive analysis of competitive endogenous RNA associated with immune infiltration in lung adenocarcinoma. *Sci Rep* (2021) 11:11056. doi: 10.1038/s41598-021-90755-w

35. Ma YC, Fan WJ, Rao SM, Gao L, Bei ZY, Xu ST. Effect of furin inhibitor on lung adenocarcinoma cell growth and metastasis. *Cancer Cell Int* (2014) 14:43. doi: 10.1186/1475-2867-14-43

36. Brant KA, Leikauf GD. Dysregulation of FURIN by prostaglandin-endoperoxide synthase 2 in lung epithelial NCI-H292 cells. *Mol Carcinog* (2014) 53:192–200. doi: 10.1002/mc.21963

37. Bassi DE, Zhang J, Renner C, Klein-Szanto AJ. Targeting proprotein convertases in furin-rich lung cancer cells results in decreased *in vitro* and *in vivo* growth. *Mol Carcinog* (2017) 56:1182–8. doi: 10.1002/mc.22550

38. Rosell R, Karachaliou N, Fancelli S, Arrieta O, Troncone G, Cao P. Proprotein convertase furin in SARS-CoV-2 and non-small cell lung cancer. *Transl Lung Cancer Res* (2020) 9:945–7. doi: 10.21037/tlcr-20-716

39. Luo L, Li M, Su J, Yao X, Luo H. FURIN correlated with immune infiltration serves as a potential biomarker in SARS-CoV-2 infection-related lung adenocarcinoma. *Clin Exp Med* (2021) 22:371–84. doi: 10.1007/s10238-021-00760-6

40. Beer S, Simins AB, Schuster A, Holzmann B. Molecular cloning and characterization of a novel SH3 protein (SLY) preferentially expressed in lymphoid cells. *Biochim Biophys Acta* (2001) 1520:89–93. doi: 10.1016/s0167-4781(01)00242-1

41. Li L, Peng M, Xue W, Fan Z, Wang T, Lian J, et al. Integrated analysis of dysregulated long non-coding RNAs/microRNAs/mRNAs in metastasis of lung adenocarcinoma. *J Transl Med* (2018) 16:372. doi: 10.1186/s12967-018-1732-z

42. Bai M, Pan Q, Sun C. Tumor purity coexpressed genes related to immune microenvironment and clinical outcomes of lung adenocarcinoma. *J Oncol* (2021) 2021:9548648. doi: 10.1155/2021/9548648

43. Liu L, Xu S, Huang L, He J, Liu G, Ma S, et al. Systemic immune microenvironment and regulatory network analysis in patients with lung adenocarcinoma. *Transl Cancer Res* (2021) 10:2859–72. doi: 10.21037/tcr-20-2275

44. Gao P, Ren G. Identification of potential target genes of non-small cell lung cancer in response to resveratrol treatment by bioinformatics analysis. *Aging (Albany NY)* (2021) 13:23245–61. doi: 10.18632/aging.203616

45. Wang Y, Hou K, Jin Y, Bao B, Tang S, Qi J, et al. Lung adenocarcinoma-specific three-integrin signature contributes to poor outcomes by metastasis and immune escape pathways. *J Transl Int Med* (2021) 9:249–63. doi: 10.2478/jtim-2021-0046

46. Klichinsky M, Ruella M, Shestova O, Lu XM, Best A, Zeeman M, et al. Human chimeric antigen receptor macrophages for cancer immunotherapy. *Nat Biotechnol* (2020) 38:947–53. doi: 10.1038/s41587-020-0462-y

47. Tian L, Lei A, Tan T, Zhu M, Zhang L, Mou H, et al. Macrophage-based combination therapies as a new strategy for cancer immunotherapy. *Kidney Dis (Basel)* (2022) 8:26–43. doi: 10.1159/000518664

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 09 November 2022
DOI 10.3389/fonc.2022.1059253

Check for updates

OPEN ACCESS

EDITED BY
Hongda Liu,
Nanjing Medical University, China

REVIEWED BY
Zhongshi Hong,
Fujian Medical University, China
Runze Zhang,
University at Buffalo, United States

*CORRESPONDENCE
Diang Chen
382455945@qq.com
Lin Zeng
381820093@qq.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

RECEIVED 01 October 2022
ACCEPTED 24 October 2022
PUBLISHED 09 November 2022

CITATION
Wang Q, Zhang X, Du K, Wu X, Zhou Y, Chen D and Zeng L (2022) Machine learning identifies characteristics molecules of cancer associated fibroblasts significantly correlated with the prognosis, immunotherapy response and immune microenvironment in lung adenocarcinoma.
Front. Oncol. 12:1059253.
doi: 10.3389/fonc.2022.1059253

COPYRIGHT
© 2022 Wang, Zhang, Du, Wu, Zhou, Chen and Zeng. 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.

Machine learning identifies characteristics molecules of cancer associated fibroblasts significantly correlated with the prognosis, immunotherapy response and immune microenvironment in lung adenocarcinoma

Qian Wang[1†], Xunlang Zhang[2†], Kangming Du[3†], Xinhui Wu[2], Yexin Zhou[4], Diang Chen[5*] and Lin Zeng[6*]

[1]Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, Sichuan, China, [2]Department of Geriatric, Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, Sichuan, China, [3]Department of Vascular Surgery, Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, Sichuan, China, [4]Guangxi University of Chinese Medicine, Nanning, China, [5]Department of Urology Surgery, Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, Sichuan, China, [6]Department of Neurosurgery, Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, Sichuan, China

Background: Lung adenocarcinoma (LUAD) is a highly lethal disease with a dramatic pro-fibrocytic response. Cancer-associated fibroblasts (CAFs) have been reported to play a key role in lung adenocarcinoma.

Methods: Marker genes of CAFs were obtained from the Cell Marker website. Single sample gene set enrichment analysis (ssGSEA) was used for CAFs quantification. R and GraphPad Prism software were utilized for all analysis. Quantitative real-time PCR (qRT-PCR) was utilized to detect the RNA level of specific molecules.

Results: Based on the ssGSEA algorithm and obtained CAFs markers, the LUAD patients with low- and high-CAFs infiltration were successfully identified, which had different response patterns to immunotherapy. Through the machine learning algorithm – LASSO logistic regression, we identified 44 characteristic molecules of CAFs. Furthermore, a prognosis signature consisting of seven characteristic genes was established, which showed great prognosis prediction ability. Additionally, we found that patients in the low-risk group might have better outcomes when receiving immunotherapy of PD-1, but not CTLA4. Also,

the biological enrichment analysis revealed that immune response-related pathways were significantly associated with CAFs infiltration. Meanwhile, we investigated the underlying biological and microenvironment difference in patients with high- and low-risk groups. Finally, we identified that AMPD1 might be a novel target for LUAD immunotherapy. Patients with a high level of AMPD1 were correlated with worse responses to immunotherapy. Moreover, immunohistochemistry showed that the protein level of AMPD1 was higher in lung cancer. Results of qRT-PCR demonstrated that AMPD1 was upregulated in A549 cells compared with BEAS-2B. Meanwhile, we found that the knockdown of AMPD4 can significantly reduce the expression of CTLA4 and PDCD1, but not CD274 and PDCD1LG2.

Conclusion: We comprehensively explored the role of CAFs and its characteristics molecules in LUAD immunotherapy and developed an effective signature to indicate patients prognosis and immunotherapy response. Moreover, AMPD1 was identified as a novel target for lung cancer immunotherapy.

KEYWORDS

cancer-associated fibroblasts, lung adenocarcinoma, immunotherapy, TIDE, immune microenvironment

Introduction

Lung cancer remains a threatening disease globally, despite declining incidences and mortality rates (1). The majority of lung cancer patients suffer from lung adenocarcinoma (LUAD). Generally, the development of LUAD is influenced by air pollution, smoking, and other factors (2). Most patients with LUAD remain incurable despite advances in treatment over the past decade (3). For this reason, the poor prognosis of LUAD patients demands a more mature early diagnosis and treatment strategy. Nowadays, combining chemoradiotherapy with immunotherapy has become the most common treatment for patients who cannot undergo surgery or resection of their cancers.

Cancer-associated fibroblasts (CAFs) is a type of activated fibroblast that is associated with cancerous cells. A significant role of CAFs is believed to be played in cancer progression (4). Multi-faceted responses by CAFs to stress in tumor environments allow them not only to adapt to tough conditions but also to enhance pro-tumorigenic behavior (5). CAFs are major tumor-promoting components in the tumor microenvironment (TME) of LUAD. According to one study, miR-210 was upregulated in CAFs-derived exosomes, enhancing cell migration, proliferation, invasiveness, and epithelial-mesenchymal transition (EMT) in lung cancer cells (6). To develop novel LUAD treatments and technologies targeted at CAFs, additional studies are needed to elucidate how tumor parenchyma and stroma interact during tumorigenesis (7).

The immune system monitors and destroys cancer cells, but tumor cells can circumvent this natural defense and develop tolerance (8). The immune system normally protects healthy tissue from cytotoxic immune reactions triggered by infection, but tumor cells can use the same immune checkpoints to evade immune destruction (9). By recognizing tumor antigens, T cells release interferon-x (INF-x), which attracts other cytotoxic immune cells and activates checkpoints. The use of immunotherapy has been beneficial to many LUAD patients, and the expression of PD-L1 helps identify those patients who are likely to respond (10). Consequently, exploring the biomarkers for predicting immune responses to lung cancer is important to maximize the effectiveness of new immunotherapy drugs (11).

High-throughput molecular technologies can help us identify genomic biomarkers that can be used to tailor precision medicine to individual patients using genomic biomarkers (12). In addition to being powerful diagnostic tools, these biomarkers also provide excellent prognostic information. The development of immunotherapy in LUAD treatment has shown exciting potential in recent years. Therefore, it is important to conduct prospective research on the biological molecules and mechanisms that affect LUAD immunotherapy.

Methods

Datasets downloaded

In addition to transcriptome expression data, The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases also provide clinical information of LUAD patients. For the TCGA database, the TCGA-LUAD project was selected and the initial expression profile file was "bcr-xml". Data in GSE68465 was downloaded from the GEO database in MINiML format, which contains the corresponding platform file GPL96 (13). Data were normalized by "preprocessCore" package in R software. Annotation information of the platform is used to convert probe IDs into gene symbols. For a gene with multiple probe IDs, the mean value of these probe values was identified as the gene expression. The baseline feature of patients in TCGA and GSE68465 were shown in Table 1 and Table 2.

TABLE 1 Baseline information of patients in TCGA.

Clinical features		Number	Percentage (%)
Age	<=65	241	46.2
	>65	262	50.2
	Unknown	19	3.6
Gender	Female	280	53.6
	Male	242	46.4
Stage	Stage I	279	53.4
	Stage II	124	23.8
	Stage III	85	16.3
	Stage IV	26	5.0
	Unknown	8	1.5
T-stage	T1	172	33.0
	T2	281	53.8
	T3	47	9.0
	T4	19	3.6
	Unknown	3	0.6
M-stage	M0	353	67.6
	M1	25	4.8
	Unknown	144	27.6
N-stage	N0	335	64.2
	N1	98	18.8
	N2	75	14.4
	N3	2	0.4
	Unknown	12	2.3

TABLE 2 Baseline information of patients in GSE68465.

Clinical features		Number	Percentage (%)
Age	<=65	231	52.1
	>65	212	47.9
Gender	Female	220	49.7
	Male	223	50.3
T-stage	T1	150	33.9
	T2	251	56.7
	T3	28	6.3
	T4	12	2.7
	Unknown	2	0.5
N-stage	N0	299	67.5
	N1	88	19.9
	N2	53	11.9
	Unknown	3	0.7

Single sample gene set enrichment analysis

The ssGSEA method was used to quantify the individual scores for each tumor sample (14). The ssGSEA algorithm computes overexpression measures for a list of genes of interest using a rank-based method. The list of CAFs markers genes obtained from the Cell Marker website (http://bio-bigdata.hrbmu.edu.cn/CellMarker/) was used to calculate ssGSEA scores.

Identification of optimal variables

The characteristic gene screening was based on LASSO logistic regression (15). LASSO logistic regression is a popular variable selection method that was implemented using the "glmnet" package.

Prognosis analysis

Based on the identified characteristics of molecules, a univariate Cox regression analysis was conducted to screen the prognosis-related genes. Next, the LASSO regression algorithm was utilized for dimensionality reduction. Multivariate Cox-regression analysis was responsible for model construction with the formula "Risk score = "Coeff A * Exp A + Coeff B * Exp B + … + Coeff N * Exp N". Kaplan-Meier (KM) survival and receiver operating characteristic (ROC) curve was utilized to evaluate the performance of prognosis model.

Immune-related analysis

Immune infiltration cells in LUAD patients were determined using the CIBERSORT, EPIC, MCPCOUNTER, QUANTISEQ, XCELL and TIMER algorithms (16–20). An evaluation of LUAD immunotherapy was conducted through the Tumor Immune Dysfunction and Exclusion (TIDE) website (21).

Enriched pathways

Gene set variation analysis (GSVA) and GSEA analysis were conducted based on the Hallmark gene set from MSIGBD project (22). For the GSEA analysis, the normalized enrichment score (NES) of the individual pathway was calculated. The reference gene sets were Hallmark and Gene Ontology (GO).

Single-cell analysis

A brief single-cell analysis was performed based on the Tumor Immune Single-cell Hub (TISCH) website (23).

Immunohistochemistry

The immunohistochemistry image of AMPD1 in lung cancer and normal tissue were downloaded from The Human Protein Atlas (HPA) project (https://www.proteinatlas.org/).

Cell culture

The normal lung cell BEAS-2B and lung cancer cell A549 were previously restored in our laboratory. These cell lines were grown with 1640 medium containing 10% fetal bovine serum.

Cell transfection

Control and knockdown plasmids were purchased from synthesized by Shanghai Jima Biotechnology. Cell transfection was conducted according to the standard process (24).

Quantitative-real time PCR

With the help of the Trizol reagent, RNA was isolated from BEAS-2B and A549 cells. Total RNA was extracted with an RNA ultrapure extraction kit, a UV-Vis spectrophotometer was used to determine total RNA concentration and purity (OD260/OD280). The PCR assay was performed with the protocol. The primer sequences are shown below: AMPD1, forward primer, 5′-TATAGTGTCAGTCAGTCACCCC-3′, reverse primer, 5′-GAGTTTGAACAGAGGCATTGTTG -3′; PDCD1, forward primer, 5’-CCAGGATGGTTCTTAGACTCCC-3’, reverse primer, 5’-TTTAGCACGAAGCTCTCCGAT-3’; CTLA4, forward primer, 5’-GCCCTGCACTCTCCTGTTTTT-3’, reverse primer, 5’-GGTTGCCGCACAGACTTCA-3’; PDCD1LG2, forward primer, 5’-ATTGCAGCTTCACCAGATAGC-3’, reverse primer, 5’-AAAGTTGCATTCCAGGGTCAC-3’; CD274, forward primer, 5’- TGGCATTTGCTGAACGCATTT-3’; reverse primer, 5’-TGGCATTTGCTGAACGCATTT-3’; GAPDH, forward primer, 5’-GGAGCGAGATCCCTCCAAAAT-3’, reverse primer, 5’-GGAGCGAGATCCCTCCAAAAT-3’.

Statistical analysis

R software was used to perform all statistical analyses. Statistical significance was determined by P values under 0.05.

Result

CAFs quantification and its role in LUAD

The flow chart of the whole study was shown in Figure S1. The single-cell analysis provides high-dimensional information that allows the identification of CAFs from heterogeneous pools of cells, as well as the ability to cluster transcriptome data into multiple CAFs subtypes based on their signature genes. According to their original annotations, we identified 11 markers of CAFs based on the Cell Marker website (Figure 1A). The single cell analysis indicated these markers ACTA2, CAV1, FAP, FN1, FOXF1, MMP2, PDGFRA, PDGFRB, PDPN, SPARC, and ZEB1 were highly expressed in the CAFs of the lung cancer microenvironment (Figure 1B). The expression patterns of these markers in TCGA and GSE68465 were presented in Figures 1C, D. Using these CAFs markers, the infiltration of CAFs in LUAD was quantified (Figures 1E, F). Based on the GSEA analysis, we explored the potential pathways in different CAFs infiltration groups. The results based on the GO gene set revealed that adaptive immune response, blood vessel morphogenesis and calcium ion transport are associated with high CAFs infiltration. In addition, results based on the Hallmark gene set showed that the EMT, inflammatory response, and IL6/JAK/STAT3 are activated in patients with high CAFs infiltration (Figures 2A, B). The TIDE score was used to evaluate tumor immune escape mechanisms. We found that CAFs infiltration was positively correlated with the TIDE score (Figure 2C). Also, the LUAD patients with high CAF infiltration had a higher TIDE score, suggesting a worse immunotherapy response (Figures 2D, E).

Identification of the CAFs characteristics molecules based on the machine learning algorithm

We then performed the differential expressed analysis between the patients with high- and low-CAFs infiltration (Figure 3A). Subsequently, the LASSO logistics regression was used to identify the characteristic molecules of CAFs with the

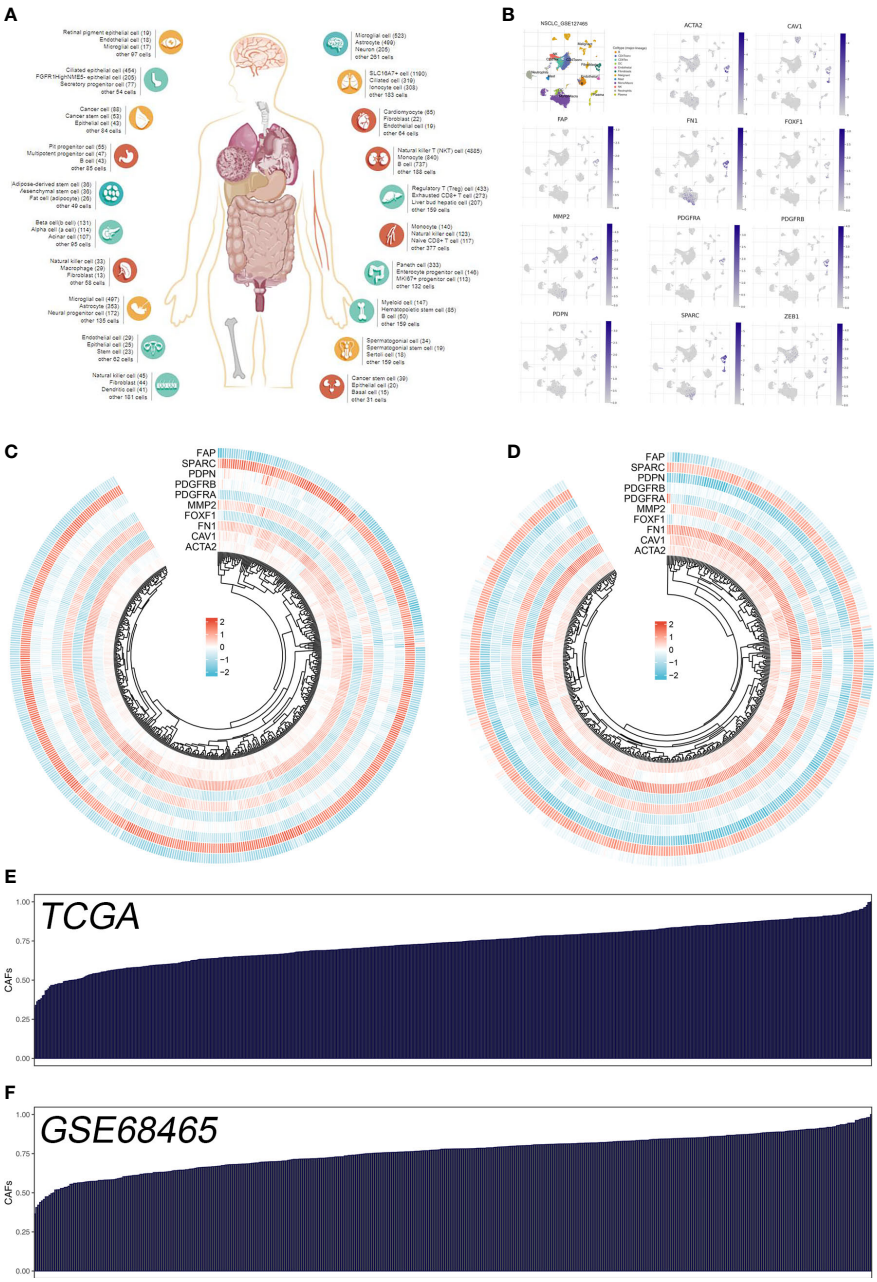


FIGURE 1
Quantification of the CAFs infiltration based on identified markers. **(A)** The human single-cell atlas from the online website Cell Marker, http://bio-bigdata.hrbmu.edu.cn/CellMarker/; **(B)** The expression map of marker genes in different subgroup cells; **(C)** The heatmap demonstrated the expression of CAFs marker genes in TCGA cohort; **(D)** The heatmap demonstrated the expression of CAFs signature genes in GSE68465 cohort; **(E)** CAFs score in TCGA cohort quantified by ssGSEA; **(F)** CAFs score in GSE68465 cohort quantified by ssGSEA.

best optimization (Figures 3B, C). A total of 44 characteristics molecules were identified when the misclassification error is the lowest. Additionally, we found that the 44 characteristic molecules were closely associated with the TIDE score, which suggests that 44 characteristics molecules may be essential in LUAD immunotherapy (Figures 3D, E). The polygenic correlation plot also revealed the strong correlations between 44 characteristic molecules (Figure 3F).

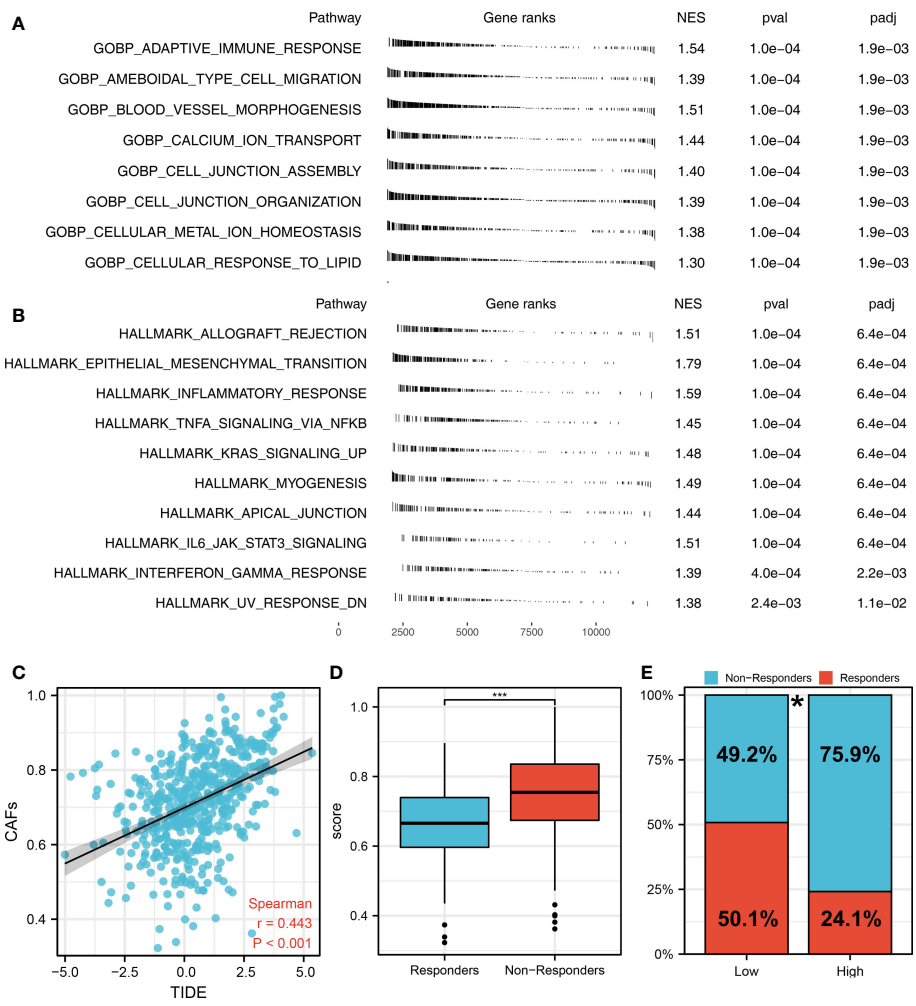


FIGURE 2
The role of CAFs in affecting lung cancer biological pathways and immunotherapy response. **(A)** The GSEA analysis between high- and low-CAFs groups based on the Hallmark gene set; **(B)** The GSEA analysis between high- and low-CAFs groups based on the GO gene set; **(C)** Correlation analysis between TIDE score and CAFs score; **(D)** CAFs level in immunotherapy responders and non-responders ***P < 0.001; **(E)** Percentage of immunotherapy responders in patients with high and low CAFs infiltration, *P < 0.05.

Establishment of a prognosis signature consisting of seven CAFs characteristics molecules

Based on 44 characteristic molecules identified by the machine learning algorithm, we performed univariate cox regression, LASSO regression, and multivariate cox regression for prognosis model establishment (Figures 4A-C). Ultimately, seven characteristic molecules were involved in the prognosis model, including ZG16, KRT85, AMPD1, L1CAM, HS3ST3A1, KRT14 and SSTR4. Each LUAD patient was assigned a risk score with the formula of "Risk score =ZG16 * 0.270 + KRT85 * 0.789 + AMPD1 * -0.366 + L1CAM * 0.099 + HS3ST3A1 * 0.145 + KRT14 * 0.098 + SSTR4 * -2.716. According to the median cut-off, all the LUAD patients were divided into different risk groups based on risk score (Figures 4D, I). In TCGA and GSE68465 cohorts, according to KM survival analysis, the LUAD patients with lower risk score had better overall survival (OS) (Figures 4E, J). As illustrated by the time-dependent ROC curve, our model showed a satisfactory prediction performance (1-year accuracy is 0.727, 3-year accuracy is 0.738, 5-year accuracy is 0.726) in the TCGA cohort (Figures 4F-H). For GEO cohort, similar results were observed (1-year accuracy = 0.682, 3-year accuracy = 0.677, 5-year accuracy = 0.739) (Figures 4K-M).

10.3389/fonc.2022.1059253

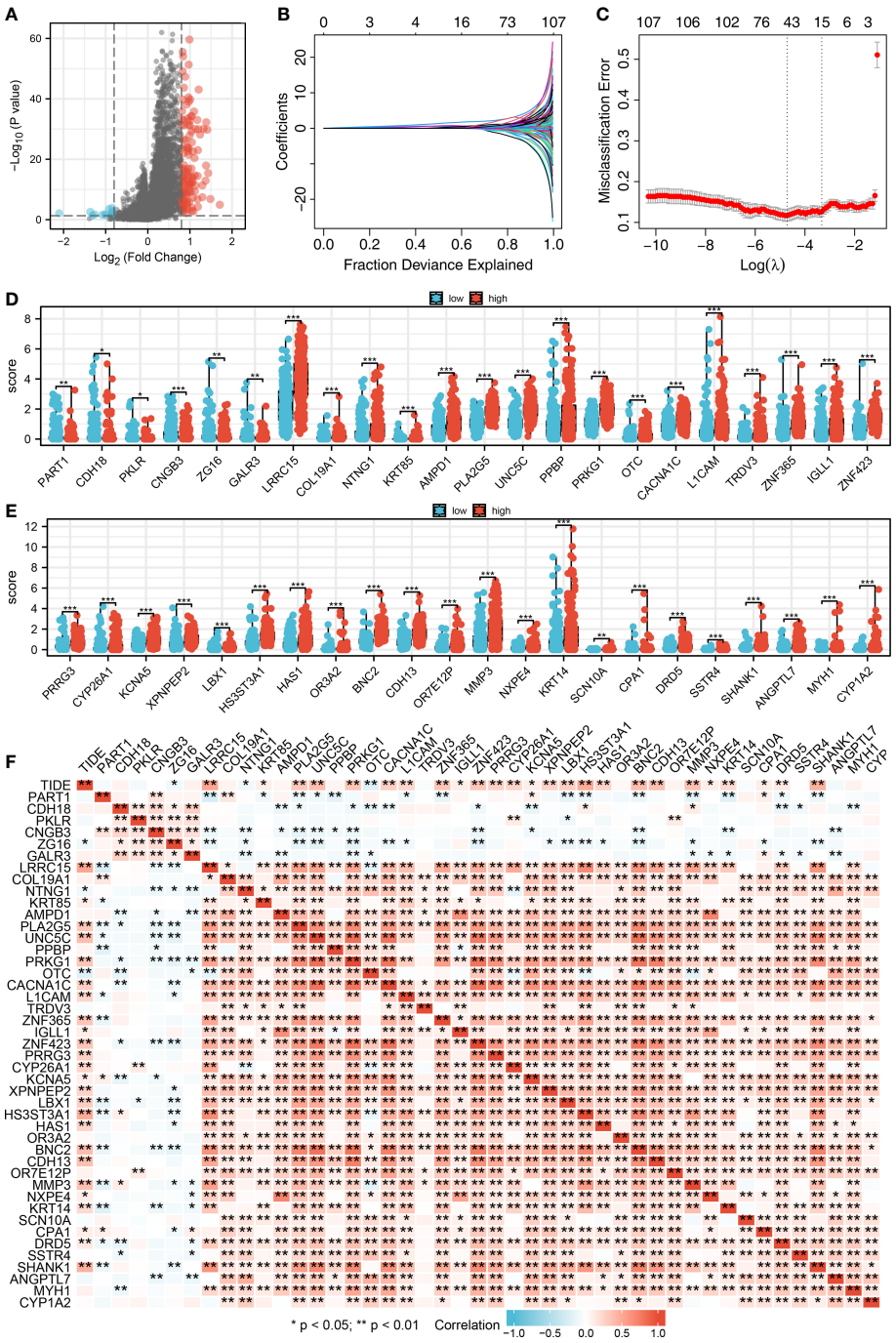


FIGURE 3
Identification of characteristic molecules of lung cancer immunotherapy. **(A)** The volcano map demonstrated the differential expressed genes between high- and low-CAFs groups; **(B-C)** Logistic lasso regression; **(D-E)** The expression level 44 CAF characteristic genes in patients with high- and low-CAFs infiltration, *P < 0.05, **P < 0.01, ***P < 0.001; **(F)** The correlation analysis between 44 CAF-related signatures and TIDE score, *P < 0.05, **P < 0.01.

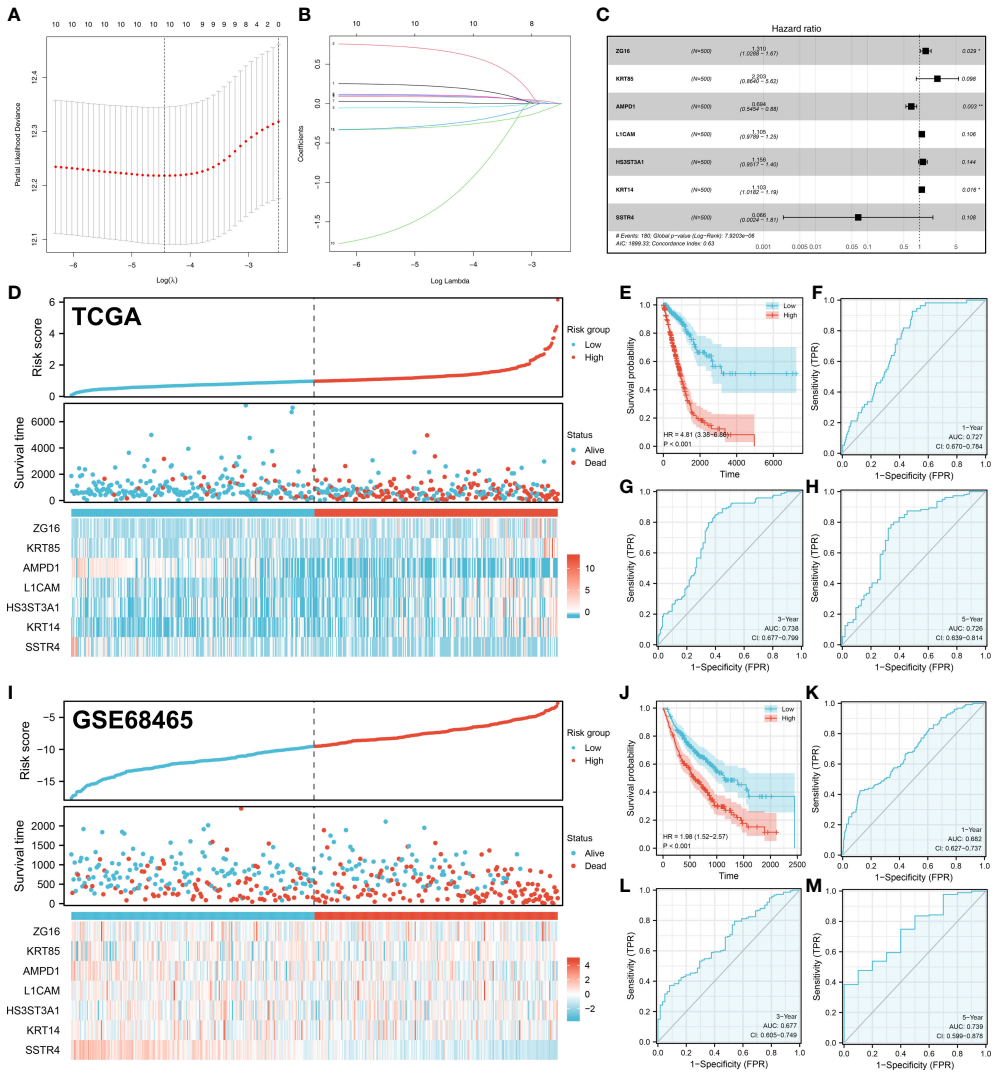


FIGURE 4
Prognosis model construction. **(A-B)** The lasso regression analysis was used for data dimension reduction; **(C)** A prognosis signature consisting of characteristic molecules was constructed by multivariate cox regression analysis; **(D)** LUAD patients involved in TCGA cohort was divided into low-risk and high-risk groups based on the median cut-off; **(E)** The Kaplan-Meier curve was used to demonstrate the OS between low-risk and high-risk groups in TCGA cohort; **(F)** The time-dependent ROC curves reveals the 1-year predictive value of prognostic in TCGA cohort; **(G)** The time-dependent ROC curves reveals the 3-year predictive value of prognostic in TCGA cohort; **(H)** The time-dependent ROC curves reveals the 5-year predictive value of prognostic in TCGA cohort; **(I)** LUAD patients involved in GEO cohort was divided into low-risk and high-risk groups based on the median cut-off; **(J)** The Kaplan-Meier curve was used to demonstrate the OS between low-risk and high-risk groups in GEO cohort; **(K)** The time-dependent ROC curves reveals the 1-year predictive value of prognostic in GEO cohort; **(L)** The time-dependent ROC curves reveals the 3-year predictive value of prognostic in GEO cohort; **(M)** The time-dependent ROC curves reveals the 5-year predictive value of prognostic in GEO cohort;. *P < 0.05, **P < 0.01.

Clinical correlation of the model and seven characteristics molecules

We found that clinical features like the clinical stage, T, N, and M classifications are significantly influenced by the risk score. Additionally, AMPD1 expression levels are closely associated with many clinical characteristics, such as age, stage, T stage, and N stage (Figures 5A-F). Results of the univariate analysis revealed that risk score is an independent prognostic factor in LUAD patients. For multivariate analysis, the same conclusion was get (Figures 5G, H).

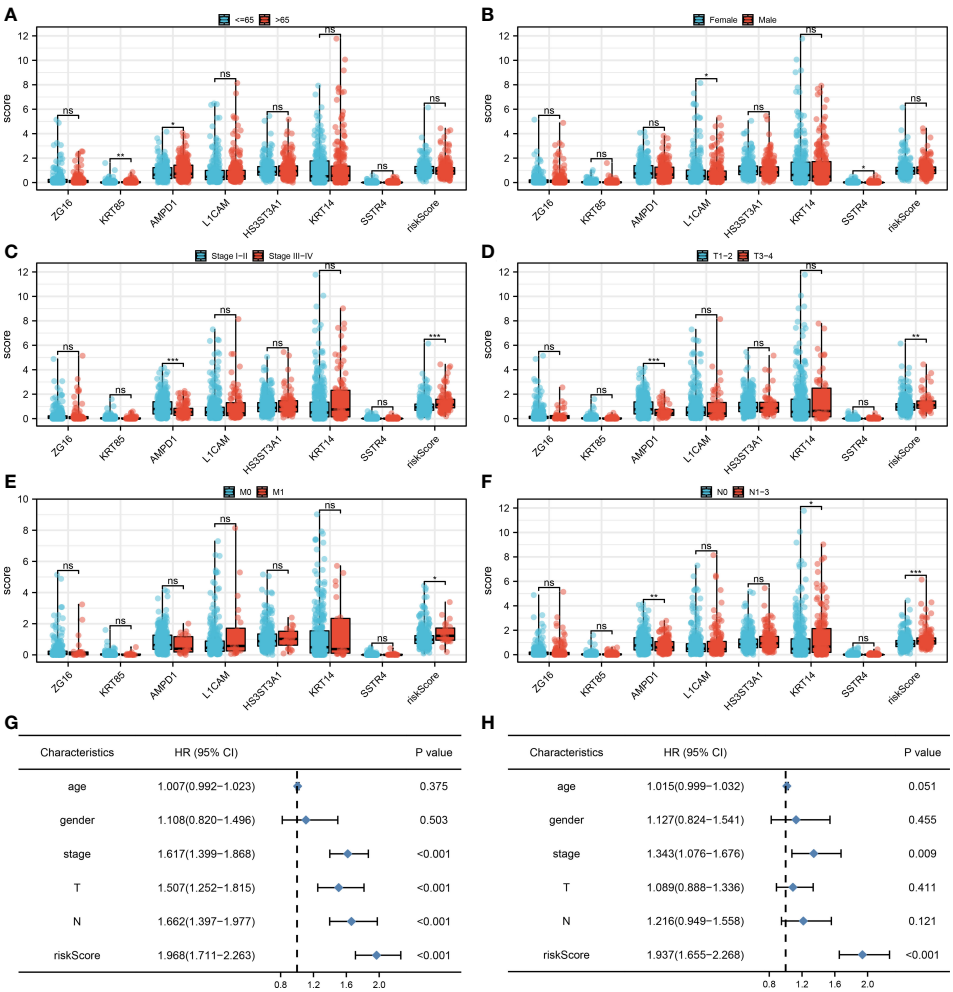


FIGURE 5
Clinical correlation. **(A)** The correlation analysis between age and expression level of seven characteristic molecules, ns P > 0.05, *P < 0.05, **P < 0.01; **(B)** The correlation analysis between gender and expression level of seven characteristic molecules, ns P > 0.05, *P < 0.05; **(C)** The correlation analysis between stage and expression level of seven characteristic molecules, ns P > 0.05, ***P <0.001; **(D)** The correlation analysis between T stage and expression level of seven characteristic molecules, ns P > 0.05, **P < 0.01, ***P <0.001; **(E)** The correlation analysis between M stage and expression level of seven characteristic molecules, ns P > 0.05, *P < 0.05; **(F)** The correlation analysis between N stage and expression level of seven characteristic molecules, ns P > 0.05, *P < 0.05, **P < 0.01, ***P <0.001; **(G)** The univariate independent prognostic analysis revealed that stage, T stage, N stage and risk score are independent prognostic factors in LUAD patients; **(H)** The multivariate independent prognostic analysis reveals that stage and risk score are independent prognostic factors in LUAD patients.

Distribution of immune cell subsets and immunotherapy response

The tumor microenvironment contains many different cells, such as fibroblasts, immune cells, inflammatory mediators, and cancer cells with varying properties and chemical properties. Among them, immune cells are important in diagnosis, survival outcomes, and drug sensitivity in tumors. Figure 6A shows the relationship between risk score and various immune cells. Briefly, the risk score can remarkably affect the LUAD immune microenvironment. Subsequently, we examined the genes encoding human leukocyte antigens (HLA) and immune checkpoint blocking (ICB). Some reduction in immunological HLA gene and ICB-related gene expression was observed in the low-risk group, suggesting that tumor cells involved in the high-risk group evade immune surveillance more (Figure 6B). Also, in TCGA and GSE68465 cohorts, the LUAD patients in the high-risk group were correlated with higher TIDE scores, indicating a worse response to immunotherapy (Figures 6C-F). The correlation heatmap demonstrated significant associations between seven characteristics molecules, risk score and TIDE score (Figure 6G). In addition, the patients with lower risk score

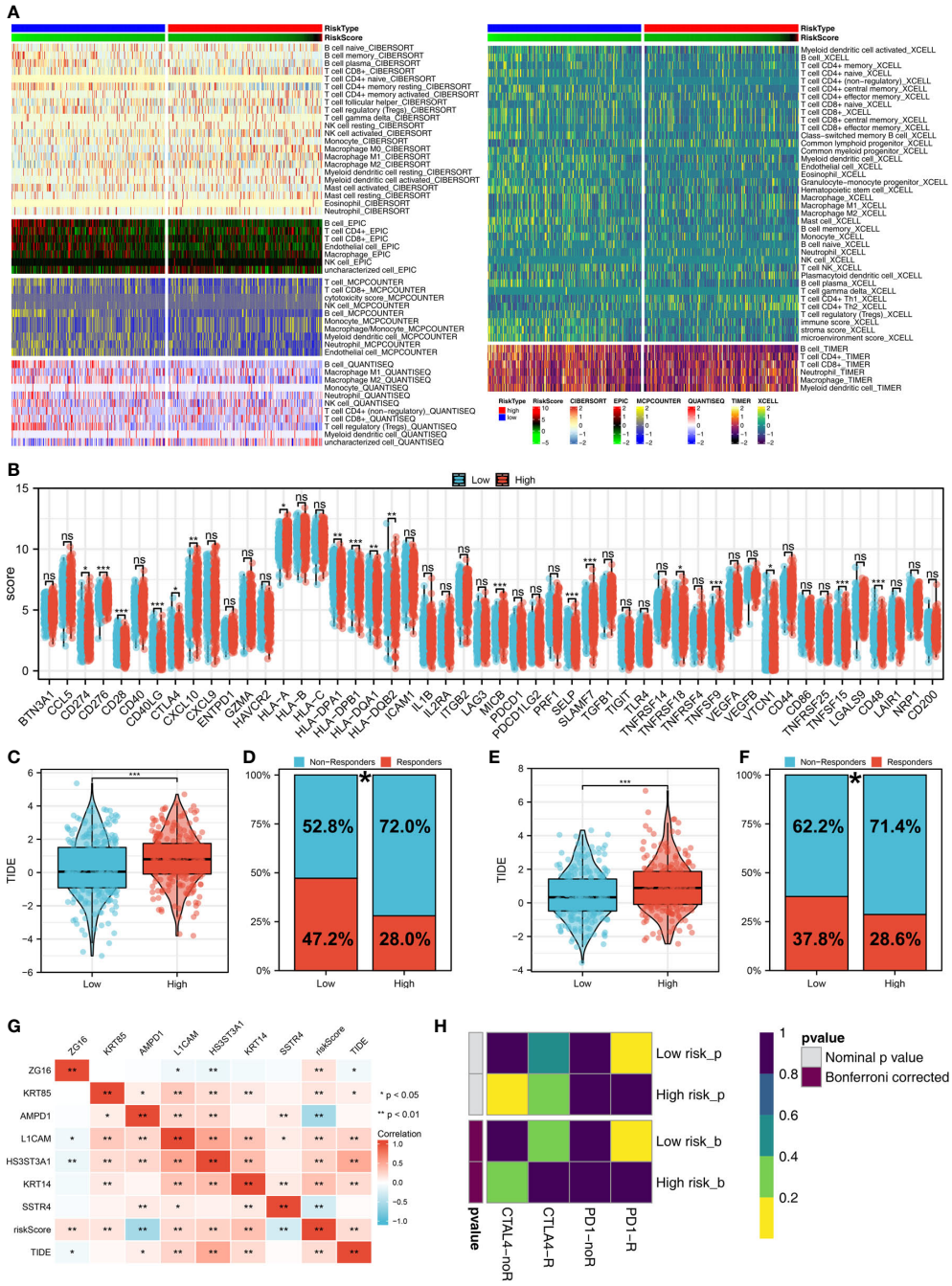


FIGURE 6
Immune-related analysis. **(A)** The correlation between risk score and various immune scores; **(B)** The expression level of immune-related genes in low- and high-risk groups, ns P > 0.05, *P < 0.05, **P < 0.01, ***P < 0.001; **(C)** The level of TIDE score in low- and high-risk groups in the TCGA cohort, ***P < 0.001; **(D)** Percentage of immunotherapy responders and non-responders in patients with high and low risk score (TCGA cohort), *P < 0.05; **(E)** T The level of TIDE score in low- and high-risk groups in the GSE68465 cohort, ***P < 0.001; **(F)** Percentage of immunotherapy responders and non-responders in patients with high and low risk score (GSE68465 cohort), *P < 0.05; **(G)** The correlation analysis between TIDE score, risk score and seven characteristic molecules, *P < 0.05, **P < 0.01; **(H)** The correlation analysis between risk score and the immunotherapy response of PD-1 and CTLA4.

might respond better to the therapy of PD-1 (Figure 6H). Then, we explored the difference in the potential biological function in patients with high- and low-risk groups by GSVA and GO enrichment analysis. The GSVA enrichment analysis revealed that MTORC1 signaling, E2F targets, glycolysis and MYC target are the most enriched terms in the high-risk group, while angiogenesis, hedgehog signaling and inflammatory response are the most enriched terms in low-risk group (Figure 7A). For GO enrichment analysis, the enriched pathways involve biological process (BP), cell component (CC) and molecular function (MF). Keratinocyte differentiation, keratinization, and epidermal cell differentiation are the most enriched GO-BP pathways (Figure 7B). For GO-CC, cornified envelope and desmosome are the most enriched pathways (Figure 7C). The results of GO-MF revealed that structural constituents of skin epidermis, serine-type endopeptidase activity and serine hydrolase activity are the most relative pathways (Figure 7D). Tumor mutational burden (TMB) and microsatellite instability (MSI) are important biomarkers for immunotherapy response. We found that the patients in the high-risk group were associated with higher TMB, but not MSI (Figure 7E, F). The stemness index (SI) is also an indicator that determines whether tumor cells are similar to stem cells and correlates with tumor dedifferentiation. The results revealed that risk score were

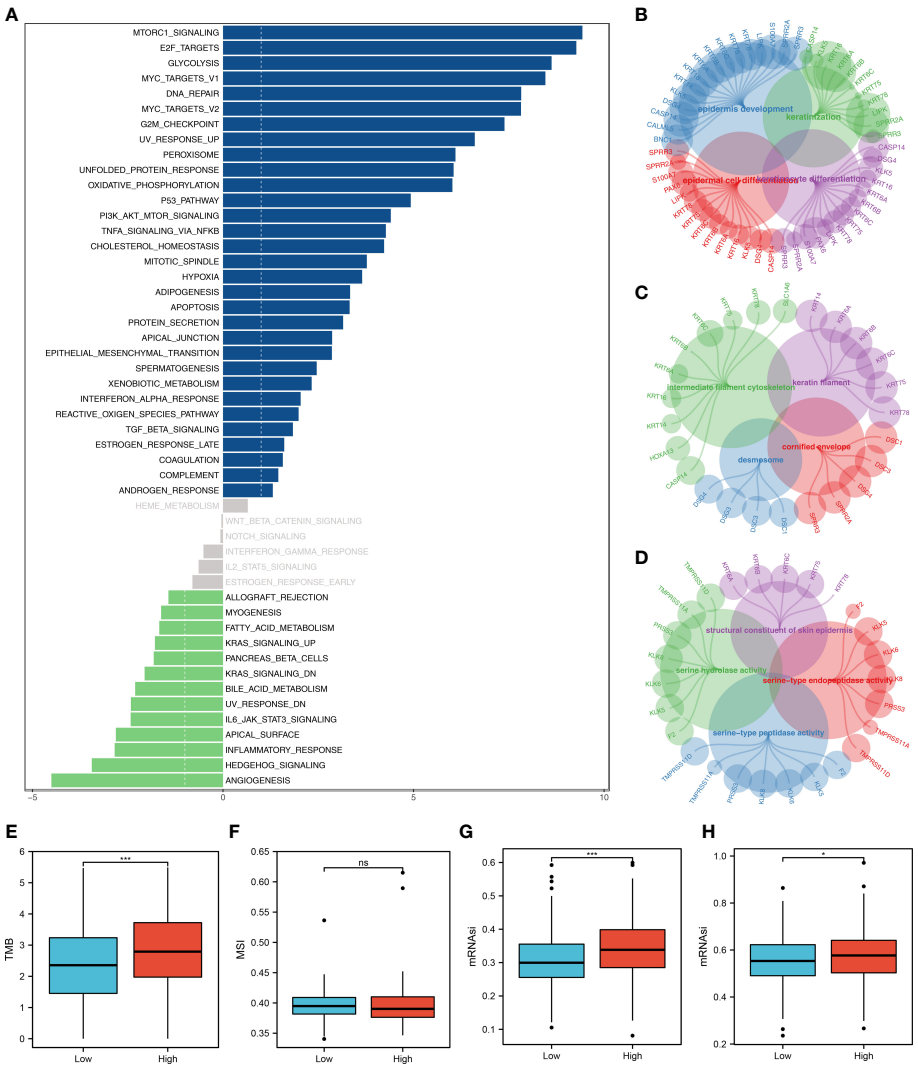


FIGURE 7
Biological enrichment analysis. **(A)** The GSVA analysis between low- and high-risk groups; **(B)** The GO-BP enrichment analysis of our model; **(C)** The GO-CC enrichment analysis of our model; **(D)** The GO-MF enrichment analysis of our model; **(E)** The boxplot demonstrated the TMB score in low- and high-risk groups, ***P < 0.001; **(F)** The boxplot demonstrated the MSI score in low- and high-risk groups, ns P > 0.05; **(G)** The boxplot demonstrated the mRNAsi score in low- and high-risk groups, ***P < 0.001; **(H)** The boxplot demonstrated the EGFR-mRNAsi score in low- and high-risk groups, *P < 0.05.

associated with higher EREG-mRNAsi and mRNAsi (Figures 7G, H).

Role of AMPD1 in LUAD and its correlation with important immune checkpoints

AMPD1 appears to play a crucial role in LUAD, according to the previous analysis. For a deeper understanding the role of AMPD1 in LUAD patients, we then performed differential expressed analysis, survival analysis, immunohistochemistry, PCR, GSEA, immune cell infiltration and TIDE score of AMPD1. No significant difference in AMPD1 mRNA level was found between the paired LUAD and the control tissue (Figure 8A). The survival analysis demonstrated that the patients with high AMPD1 levels might have better OS, disease-specific survival (DSS) and progress-free interval (PFI) (Figures 8B-D). In addition, the clinical analysis demonstrated that AMPD1 is closely associated with some clinical

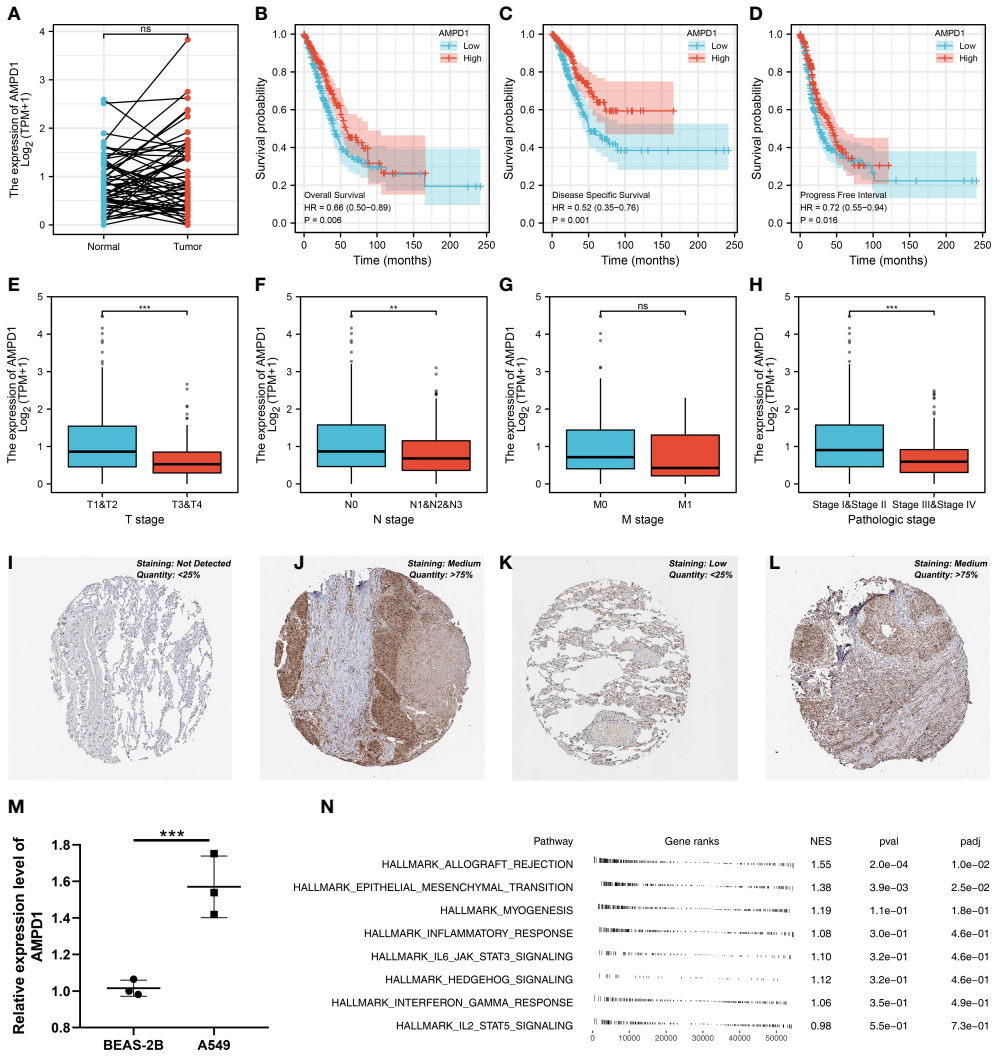


FIGURE 8
Further exploration of AMPD1 in lung cancer. **(A)** The expression level of AMPD1 in paired LUAD and control tissue, ns P > 0.05; **(B)** The OS in patients with low and high AMPD1 expression level; **(C)** The DSS in patients with low and high AMPD1 expression level; **(D)** The PFI in patients with low and high AMPD1 expression level; **(E)** The correlation analysis between expression level of AMPD1 and T stage, ***P <0.001; **(F)** The correlation analysis between expression level of AMPD1 and N stage, **P < 0.01; **(G)** The correlation analysis between expression level of AMPD1 and M stage, ns P > 0.05; **(H)** The correlation analysis between expression level of AMPD1 and stage, ***P < 0.001; **(I)** The AMPD1 protein level in normal lung tissue; **(J)** The AMPD1 protein level in LUAD tissue; **(K)** The AMPD1 protein level in normal lung tissue; **(L)** The AMPD1 protein level in LUAD tissue; **(M)** The qRT-PCR assay demonstrated the expression level of AMPD1 in A549 and BEAS-2B cells, ***P < 0.001; **(N)** The GSEA analysis of AMPD1.

characteristics, such as clinical stage, T stage and N stage (Figures 8E-H). Immunohistochemical showed that the protein level of AMPD1 was higher in the LUAD tissue (Figures 8I-L). Also, the qRT-PCR result indicated that AMPD1 is upregulated in A549 cells compared with BEAS-2B cells at mRNA level (Figure 8M). The GSEA analysis revealed that allograft rejection, EMT, myogenesis, inflammatory response and IL6-JAK-STAT3 signaling may be the key pathways AMPD1 involved in (Figure 8N). The immune cell infiltration also revealed a positive correlation between AMPD1 expression and B cells, TFH, T cells, Th1 cells, cytotoxic cells, and mast cells (Figure 9A). Also, a positive correlation was observed between CAFs and AMPD1 (Figure S2). Further, in GEO and TCGA databases, we found that AMPD1 is associated with higher TIDE scores. LUAD patients with high AMPD1 expression might respond better to immunotherapy (Figures 9B-G). Considering the effect of AMPD1 on immunotherapy response, we next evaluated the correlation between AMPD1 and four key immune checkpoints, CD274, PDCD1, CTLA4 and PDCD1LG2. Results indicated that AMPD1 was positively correlated with these checkpoint molecules, especially CTLA4 (Figures 10A-D). The qRT-PCR result of cell lines showed that

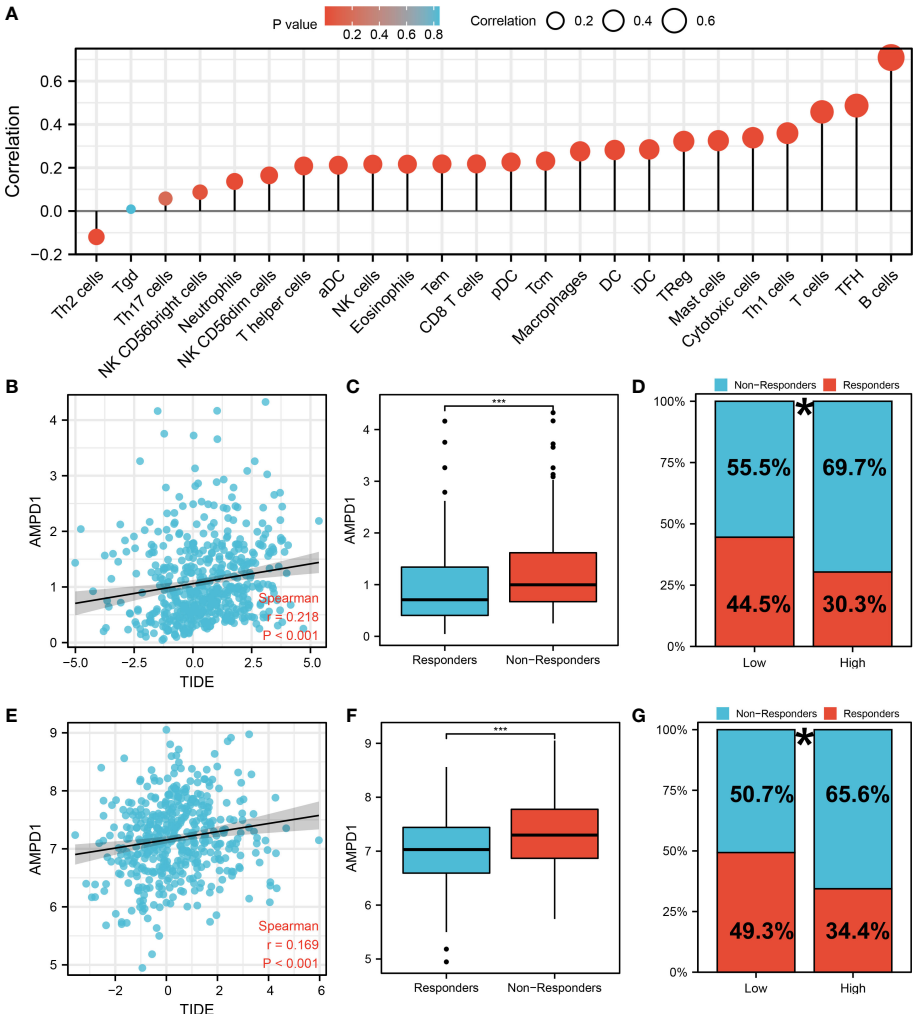


FIGURE 9
Immune-related analysis of AMPD1 in lung cancer. **(A)** The immune cell infiltration reveals that B cells, TFH, T cells, Th1 cells, cytotoxic cells and mast cells are positively correlated with the expression of AMPD1; **(B)** The correlation analysis between TIDE score and expression level of AMPD1 in TCGA cohort; **(C)** The AMPD1 expression in immunotherapy responders and non-responders (TCGA cohort), ***P < 0.001; **(D)** Percentage of immunotherapy responders and non-responders in high- and low-risk patients (TCGA cohort), *P <0.05; **(E)** The correlation analysis between TIDE score and expression level of AMPD1 in GSE68465 cohort; **(F)** The AMPD1 expression in immunotherapy responders and non-responders (GSE68465 cohort), ***P < 0.001; **(G)** Percentage of immunotherapy responders and non-responders in high- and low-risk patients (GSE68465 cohort), *P < 0.05.

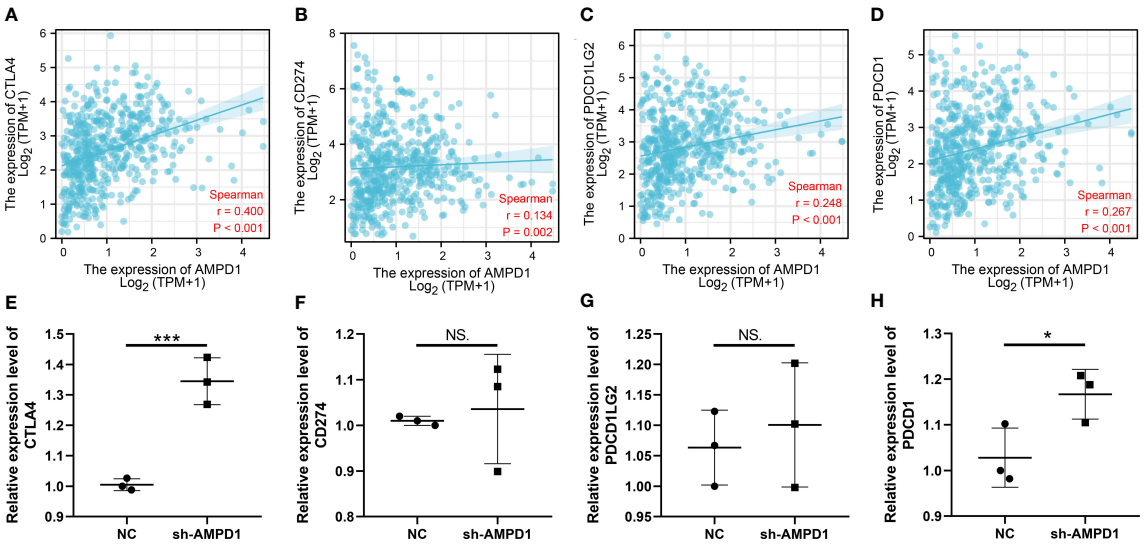


FIGURE 10
Correlation of AMPD1 and key immune checkpoints. **(A-D)** Correlation between AMPD1 and key immune checkpoints; **(E-H)** RNA level of key immune checkpoint in AMPD1 knockdown and control cells, NS P > 0.05, *P < 0.05, ***P < 0.001.

knockdown of AMPD4 can significantly reduce the expression of CTLA4 and PDCD1, but not CD274 and PDCD1LG2 (Figures 10E-H).

Discussion

There are more than one million deaths worldwide each year caused by LUAD. Although new diagnostic and therapeutic technologies have been developed, it remains a serious global public health problem lacking effective advanced diagnostics and treatment options (25). Consequently, it is necessary to develop novel and effective markers to improve the early diagnosis and treatment of LUAD.

Cancer progression is believed to be largely driven by CAFs, which may be valuable therapeutic targets (26). Several studies indicated that CAFs are closely correlated with the development and occurrence of lung cancer. According to previous studies, miR-210 released by CAFs-exosomes has a role in promoting migration and invasion in lung cancer cells (27). In addition, FUT8/CF in CAFs propagated aggressive and malignant tissue microenvironments, which led to the faster proliferation and greater invasiveness of lung cancer in vivo and in vitro (28). In this work, based on the ssGSEA algorithm and single-cell analysis, LUAD patients were divided into different risk groups. Meanwhile, the patients with different CAFs infiltration levels showed different response patterns to immunotherapy. Subsequently, after screening differential expressed genes, we finally constructed a prognostic prediction model based on seven CAFs characteristics molecules.

Researchers have found that CAFs can significantly affect the cancer microenvironment. By inducing immunosuppressive macrophages, CAF may lead to an immunosuppressive environment (29). In a study by Zhao et al., the infiltration of CAFs can induce regulatory T-Cell (Treg) infiltration and indicate poor prognosis of oral squamous cell carcinoma patients (30). Moreover, CAFs promote the accumulation of tumor-promoting macrophages, resulting in immunosuppression (31). In this study, we found that the risk score has closely correlated with many immune-related pathways and patients with lower risk score might respond better to immunotherapy. In addition, the reduction in immunological expression of HLA genes and immune checkpoint genes was higher in the low-risk group. More specifically, the LUAD patients in the low-risk group show better outcomes when receiving immunotherapy of PD-1. Also, the biological enrichment revealed that CAFs was significantly correlated with immune response. In this work, we also discovered that AMPD1 was associated with many immune cells, including B cells and T cells. Moreover, we found that LUAD patients with a high expression level of AMPD1 are associated with higher TIDE scores. Accordingly, lung cancer patients who express high levels of AMPD1 have a worse response to immunotherapy.

In seven characteristics molecules involved in the prognosis prediction model, we found that AMPD1 is highly associated with clinical characteristics, such as age, clinical stage, T stage and N stage. Further, we found that the LUAD patients with higher

AMPD1 expression levels were associated with better OS, DSS and PFI. The qRT-PCR assay demonstrated that the expression level of AMPD1 in A549 cells is higher than in normal lung cells. Immunohistochemistry showed that the protein expression level is inconsistent with the RNA level. According to numerous studies, many tumors are related to AMPD1. In breast cancer, a study showed that AMPD1 expression was closely linked to tumor-infiltrating immune cells and prognosis outcomes in HER2-positive breast cancer and that it may serve as a potential biomarker (32). In thyroid cancer, AMPD1 expression closely correlates with malignant evolution and the clinical prognosis of patients, and it is promising to become an important biomarker for immunotherapy (33).

Considering recent exciting developments in immunotherapy in cancer, it is meaningful to develop new and appealing immunotherapeutic approaches by utilizing innate immune cells. CAFs can influence the microenvironment of tumors. Hence, therapeutic strategies that reactivate the immunosuppressive microenvironment that CAF mediates may enhance conventional treatments as well as immunotherapies. Understanding the interaction between CAFs and tumor progression will allow new therapy clues to be identified. Meanwhile, some limitations should also be noticed. Firstly, the patients included in our study were primarily Western populations and the race bias might reduce the conclusion reliability. Secondly, the detailed information of patients was not clear, for instance, detailed laboratory inspection data, which might bring underlying bias.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://portal.gdc.cancer.gov/, https://www.ncbi.nlm.nih.gov/gds/?term.

Author contributions

WQ, WX, ZX and DK performed the analysis. WQ, WX and ZY performed the experiments and wrote the manuscript. CD and ZL designed this work. All authors participated in the 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/fonc.2022.1059253/full#supplementary-material

SUPPLEMENTARY FIGURE 1
The flow chart of the whole study.

SUPPLEMENTARY FIGURE 2
The correlation analysis between CAFs infiltration and expression level of AMPD1.

References

1. Bodor JN, Boumber Y, Borghaei H. Biomarkers for immune checkpoint inhibition in non-small cell lung cancer (NSCLC). *Cancer* (2020) 126:260–70. doi: 10.1002/cncr.32468

2. Cho H, Seo Y, Loke KM, Kim SW, Oh SM, Kim JH, et al. Cancer-stimulated CAFs enhance monocyte differentiation and protumoral TAM activation *via* IL6 and GM-CSF secretion. *Clin Cancer Res* (2018) 24:5407–21. doi: 10.1158/1078-0432.Ccr-18-0125

3. de Sousa VML, Carvalho L. Heterogeneity in lung cancer. *Pathobiology* (2018) 85:96–107. doi: 10.1159/000487440

4. Ham J, Wang B, Po JW, Singh A, Niles N, Lee CS. Cancer-associated fibroblasts (CAFs) in thyroid papillary carcinoma: molecular networks and interactions. *J Clin Pathol* (2021) 74:759–65. doi: 10.1136/jclinpath-2020-207357

5. Hu JL, Wang W, Lan XL, Zeng ZC, Liang YS, Yan YR, et al. CAFs secreted exosomes promote metastasis and chemotherapy resistance by enhancing cell stemness and epithelial-mesenchymal transition in colorectal cancer. *Mol Cancer* (2019) 18:91. doi: 10.1186/s12943-019-1019-x

6. Hurtado P, Martínez-Pena I, Piñeiro R. Dangerous liaisons: Circulating tumor cells (CTCs) and cancer-associated fibroblasts (CAFs). *Cancers* (2020) 12:2861. doi: 10.3390/cancers12102861

7. Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, et al. Tracking the evolution of non-Small-Cell lung cancer. *New Engl J Med* (2017) 376:2109–21. doi: 10.1056/NEJMoa1616288

8. Ren X, Chen X, Zhang X, Jiang S, Zhang T, Li G, et al. Immune microenvironment and response in prostate cancer using Large population cohorts. *Front Immunol* (2021) 12:686809. doi: 10.3389/fimmu.2021.686809

9. Jang I, Beningo KA. Integrins, CAFs and mechanical forces in the progression of cancer. *Cancers* (2019) 11:721. doi: 10.3390/cancers11050721

10. Kang J, Zhang C, Zhong WZ. Neoadjuvant immunotherapy for non-small cell lung cancer: State of the art. *Cancer Commun* (2021) 41:287–302. doi: 10.1002/cac2.12153

11. Kochetkova M, Samuel MS. Differentiation of the tumor microenvironment: are CAFs the organizer? *Trends Cell Biol* (2022) 32:285–94. doi: 10.1016/j.tcb.2021.11.008

12. Zhang X, Lu Z, Ren X, Chen X, Zhou X, Zhou X, et al. Genetic comprehension of organophosphate flame retardants, an emerging threat to prostate cancer. *Ecotoxicol Environ Saf* (2021) 223:112589. doi: 10.1016/j.ecoenv.2021.112589

13. Shedden K, Taylor JM, Enkemann SA, Tsao MS, Yeatman TJ, Gerald WL, et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med* (2008) 14:822–7. doi: 10.1038/nm.1790

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

15. Hu JY, Wang Y, Tong XM, Yang T. When to consider logistic LASSO regression in multivariate analysis? *Eur J Surg Oncol* (2021) 47:2206. doi: 10.1016/j.ejso.2021.04.011

16. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol Biol* (2018) 1711:243–59. doi: 10.1007/978-1-4939-7493-1_12

17. Racle J, Gfeller D. EPIC: A tool to estimate the proportions of different cell types from bulk gene expression data. *Methods Mol Biol* (2020) 2120:233–48. doi: 10.1007/978-1-0716-0327-7_17

18. Plattner C, Finotello F, Rieder D. Deconvoluting tumor-infiltrating immune cells from RNA-seq data using quanTIseq. *Methods enzymol* (2020) 636:261–85. doi: 10.1016/bs.mie.2019.05.056

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

20. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. TIMER: A web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res* (2017) 77:e108–10. doi: 10.1158/0008-5472.Can-17-0307

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

22. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci United States America* (2005) 102:15545–50. doi: 10.1073/pnas.0506580102

23. Sun D, Wang J, Han Y, Dong X, Ge J, Zheng R, et al. TISCH: a comprehensive web resource enabling interactive single-cell transcriptome visualization of tumor microenvironment. *Nucleic Acids Res* (2021) 49:D1420–d1430. doi: 10.1093/nar/gkaa1020

24. Wei Y, Chen X, Ren X, Wang B, Zhang Q, Bu H, et al. Identification of MX2 as a novel prognostic biomarker for sunitinib resistance in clear cell renal cell carcinoma. *Front Genet* (2021) 12:680369. doi: 10.3389/fgene.2021.680369

25. Collins LG, Haines C, Perkel R, Enck RE. Lung cancer: diagnosis and management. *Am Family phys* (2007) 75:56–63.

26. Biffi G, Tuveson DA. Diversity and biology of cancer-associated fibroblasts. *Physiol Rev* (2021) 101:147–76. doi: 10.1152/physrev.00048.2019

27. Yang F, Yan Y, Yang Y, Hong X, Wang M, Yang Z, et al. MiR-210 in exosomes derived from CAFs promotes non-small cell lung cancer migration and invasion through PTEN/PI3K/AKT pathway. *Cell signal* (2020) 73:109675. doi: 10.1016/j.cellsig.2020.109675

28. Li F, Zhao S, Cui Y, Guo T, Qiang J, Xie Q, et al. α1,6-fucosyltransferase (FUT8) regulates the cancer-promoting capacity of cancer-associated fibroblasts (CAFs) by modifying EGFR core fucosylation (CF) in non-small cell lung cancer (NSCLC). *Am J Cancer Res* (2020) 10:816–37.

29. Saw PE, Chen J, Song E. Targeting CAFs to overcome anticancer therapeutic resistance. *Trends Cancer* (2022) 8:527–55. doi: 10.1016/j.trecan.2022.03.001

30. Zhao X, Ding L, Lu Z, Huang X, Jing Y, Yang Y, et al. Diminished CD68(+) cancer-associated fibroblast subset induces regulatory T-cell (Treg) infiltration and predicts poor prognosis of oral squamous cell carcinoma patients. *Am J Pathol* (2020) 190:886–99. doi: 10.1016/j.ajpath.2019.12.007

31. Gunaydin G. CAFs interacting with TAMs in tumor microenvironment to enhance tumorigenesis and immune evasion. *Front Oncol* (2021) 11:668349. doi: 10.3389/fonc.2021.668349

32. Wang L, Zhang X, Wang M, Li Y, Xu J, Wei J, et al. AMPD1 is associated with the immune response and serves as a prognostic marker in HER2-positive breast cancer. *Front Oncol* (2021) 11:749135. doi: 10.3389/fonc.2021.749135

33. Zha T, Wu H. Expression of serum AMPD1 in thyroid carcinoma and its clinical significance. *Exp Ther Med* (2018) 15:3357–61. doi: 10.3892/etm.2018.5859

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 14 November 2022
DOI 10.3389/fonc.2022.1042766

Check for updates

OPEN ACCESS

EDITED BY
Ping Zhan,
Nanjing University School of Medicine, China

REVIEWED BY
Jing Yang,
Chongqing Technology and Business University, China
Chrysostomi Gialeli,
Lund University, Sweden

*CORRESPONDENCE
Camila Machado Baldavira
camilamachado@usp.br
Vera Luiza Capelozzi
vera.capelozzi@fm.usp.br

SPECIALTY SECTION
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

RECEIVED 13 September 2022
ACCEPTED 27 October 2022
PUBLISHED 14 November 2022

CITATION
Baldavira CM, Prieto TG, Machado-Rugolo J, de Miranda JT, de Oliveira LKR, Velosa APP, Teodoro WR, Ab'Saber A, Takagaki T and Capelozzi VL (2022) Modeling extracellular matrix through histo-molecular gradient in NSCLC for clinical decisions.
Front. Oncol. 12:1042766.
doi: 10.3389/fonc.2022.1042766

COPYRIGHT
© 2022 Baldavira, Prieto, Machado-Rugolo, de Miranda, de Oliveira, Velosa, Teodoro, Ab'Saber, Takagaki and Capelozzi. 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.

Modeling extracellular matrix through histo-molecular gradient in NSCLC for clinical decisions

Camila Machado Baldavira[1]*, Tabatha Gutierrez Prieto[1], Juliana Machado-Rugolo[1,2], Jurandir Tomaz de Miranda[3], Lizandre Keren Ramos de Oliveira[3], Ana Paula Pereira Velosa[3], Walcy Rosolia Teodoro[3], Alexandre Ab'Saber[1], Teresa Takagaki[4] and Vera Luiza Capelozzi[1]*

[1]Department of Pathology, Faculty of Medicine, University of São Paulo, São Paulo, Brazil, [2]Health Technology Assessment Center, Clinical Hospital, Medical School of São Paulo State University, Botucatu, São Paulo, Brazil, [3]Rheumatology Division of the Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, Faculty of Medicine, University of São Paulo, São Paulo, SP, Brazil, [4]Division of Pneumology, Instituto do Coração (Incor), University of São Paulo Medical School (USP), São Paulo, Brazil

Lung cancer still represents a global health problem, being the main type of tumor responsible for cancer deaths. In this context, the tumor microenvironment, and the extracellular matrix (ECM) pose as extremely relevant. Thus, this study aimed to explore the prognostic value of epithelial-to-mesenchymal transition (EMT), Wnt signaling, and ECM proteins expression in patients with non–small-cell lung carcinoma (NSCLC) with clinical stages I-IIIA. For that, we used 120 tissue sections from patients and evaluated the immunohistochemical, immunofluorescence, and transmission electron microscopy (TEM) to each of these markers. We also used *in silico* analysis to validate our data. We found a strong expression of E-cadherin and β-catenin, which reflects the differential ECM invasion process. Therefore, we also noticed a strong expression of chondroitin sulfate (CS) and collagens III and V. This suggests that, after EMT, the basal membrane (BM) enhanced the motility of invasive cells. EMT proteins were directly associated with WNT5A, and collagens III and V, which suggests that the WNT pathway drives them. On the other hand, heparan sulfate (HS) was associated with WNT3A and SPARC, while WNT1 was associated with CS. Interestingly, the association between WNT1 and Col IV suggested negative feedback of WNT1 along the BM. In our cohort, WNT3A, WNT5A, heparan sulfate and SPARC played an important role in the Cox regression model, influencing the overall survival (OS) of patients, be it directly or indirectly, with the SPARC expression stratifying the OS into two

groups: 97 months for high expression; and 65 for low expression. In conclusion, the present study identified a set of proteins that may play a significant role in predicting the prognosis of NSCLC patients with clinical stages I-IIIA.

KEYWORDS

lung cancer, extracellular matrix, epithelial-to-mesenchymal transition, WNT signaling pathway, glycosaminoglycans

1 Introduction

Non-small-cell lung cancer (NSCLC) represents the most frequent malignant epithelial tumor of the lung. It accounts for 85% of the cases and mostly includes three histological subtypes: lung adenocarcinoma (ADC); lung squamous cell carcinoma (SqCC); and large cell carcinoma (LCC) (1). Globally, NSCLC remains the major cause of cancer mortality (2). In Brazil, the five-year overall survival (OS) is estimated at 18% (3). Radically resected NSCLC has a significant risk of progressing to distant metastasis, an outcome seen in 40% of the patients (4). One of the major issues associated with this process of tumor recurrence may be linked to occult metastases, which are still difficult to detect, even with current clinical advances (5).

Metastatic processes require malignant cells to invade the surrounding tissue *via* multiple steps. Two of these steps are of particular significance: the epithelial-mesenchymal transition process (EMT); and cell migration through the tissue to the vessels (6). To successfully complete both steps, tumor cells must express certain features, especially those that impact their ability to adhere to different molecules present in the extracellular matrix (ECM) and to other cell surfaces (7). Furthermore, since the ECM is remodeled, it is also relevant because it enables cell mobility, which results in metastasis (7).

The EMT process is relevant in NSCLC, which usually develops from the epithelial cells lining the bronchiolar and alveolar epithelium (8, 9). This is because it is through the EMT process that the cells change their epithelial phenotype in favor of a mobile mesenchymal phenotype (10). For this, epithelial cells interact with the basement membrane and undergo multiple biochemical changes that allow the alteration of cell phenotype. As a result, tumor cells acquire enhanced migratory capabilities, invasiveness, and significantly increased production of ECM molecules (11). Thus, ECM becomes crucial for cells to progress in this mesenchymal phenotype, thus reverting to their epithelial origins after secondary site invasion (12). Soon, tumor cells become able to re-epithelialize at the metastatic site. This is vital for colonization and development of metastatic extensions (13).

The ECM provides the cells with histoarchitectural support and anchoring. The ECM is composed of a complex network of highly cross-linked components, including fibrous proteins, glycoproteins, proteoglycans, and polysaccharides (14). The biomechanical and biochemical properties of the ECM regulate cell survival, proliferation, differentiation, and motility through the action of proteins such as SPARC, chondroitin sulfate (CS), heparan sulfate (HS), and collagens (15–18). The molecular changes that occur in the ECM have been potentially associated with invasive carcinoma. Furthermore, the modifications undergone by the ECM can modulate important signaling pathways in tissue morphogenesis, such as the Wnt signaling pathway. Abnormal signaling of this pathway is already associated with several types of cancer, where it exerts a tumorigenic effect (19, 20). In addition, this pathway has also been described as influencing the EMT process, which consequently acts on tumor growth and progression (19).

In this regard, several studies have investigated the ECM components individually. However, to the best of our knowledge, there are no studies investigating the structural components of the ECM combined with EMT behavior and the Wnt signaling pathway in NSCLC tissues. Thus, the present study analyzes the ECM patterns in different types of NSCLC and associates them with the expression of EMT markers and WNT proteins, and with the clinicopathologic features and outcome of patients.

2 Methods

2.1 Study cohort

We conducted a retrospective, longitudinal, and unicentric study on a consecutive series of patients with NSCLC who underwent surgery between 2004 and 2012 at the Thoracic Surgery Unit of Hospital das Clínicas, Instituto do Coração (InCor), and Instituto de Câncer de São Paulo (ICESP) linked to the University of São Paulo Medical School. We included chemo-naive patients with a histological diagnosis of NSCLC stage I, II or IIIA, and adequate tissue samples obtained from

thoracic surgery. We excluded from the study patients treated with neoadjuvant chemotherapy and/or radiotherapy, palliative surgical procedure and the specimens inadequately fixed in paraffin.

We collected and managed patient data using REDcap electronic data capture tools at ICESP and included: sex, age, smoking history, histology, and disease stage – according to the 8th edition of the International Union for Cancer Control (UICC) TNM Classification of Malignant Tumors (21) –, as well as subsequent systemic or locoregional treatments, eventual recurrence, and death. All patients were followed up through monthly consultations with the oncologist and submitted to brain, chest, and abdominal CT scans every six months for the first five years, and annually thereafter. Overall survival (OS) served as the primary endpoint and was defined as first contact to death from recurrent lung cancer.

We carried the study out in accordance with the rules of Good Clinical Practice and the principles of the Declaration of Helsinki. The Internal Ethics Committees of all participating institutions approved this study protocol under number 150.443/2019.

2.2 Tissue microarray

The tissue microarray (TMA) slides were constructed with 120 samples of primary tumor tissue collected consecutively using three 1.5 mm tissue cores from the central, intermediate, and peripheral portions of the most representative tumor areas. An experienced pathologist had previously selected these areas and marked on a hematoxylin-eosin-stained sample. We used normal liver and kidney tissues to control and guide the slides and classified the tumors histologically according to the 2015 World Health Organization guidelines on lung cancer classification (22).

2.3 Immunohistochemistry and immunofluorescence assays

To perform the immunohistochemistry (IHC) and immunofluorescence (IF) assays, first we tested the immunostains on both whole tissue and TMA sections to ensure uniformity. We added negative controls to verify that the staining is specific using isotype antibody controls. Then, the TMA sections (N=120) were stained with immunoperoxidase and antibodies against: E-cadherin (1:100; Boster Biological), β-catenin (1:100; Santa Cruz), heparan sulfate (1:500; Santa Cruz), chondroitin sulfate (1:100, Santa Cruz), WNT1 (1:100, Santa Cruz), WNT3A (1:100, Abnova), WNT5A (1:400, Abnova), WNT5B (1:50, Santa Cruz), and SPARC (1:400, BIOSS).

To perform the immunofluorescence assay, we dewaxed the TMA sections (N=120) in xylol, hydrated in graded ethanol, and exposed them to a 0.3% hydrogen peroxide and formic acid solution to inhibit endogenous peroxidase activity. Antigen retrieval was accomplished using a citrate buffer solution at pH 9.0 and heated in a Pascal pressure cooker (125°C for 1 minute). Nonspecific sites were blocked with 5% bovine serum albumin (BSA) in phosphate buffer saline (PBS) for 30 minutes at room temperature. The specimens were incubated overnight at 4°C with antibodies against: E-cadherin (1:100; Boster Biological), β-catenin (1:100; Santa Cruz), heparan sulfate (1:500; Santa Cruz), chondroitin sulfate (1:100, Santa Cruz), anti-human collagen type I (1:700; Rockland Inc.), anti-human collagen type III (1:200; Rockland Inc.), anti-human collagen type IV (1:100; Dako), and anti-human collagen type V (1:1000; Rockland Inc.). These TMA sections were then washed in PBS with Tween 20 at 0.05% and incubated for 60 minutes at room temperature with Alexa 488-conjugated goat anti-mouse IgG (1:200, Invitrogen, Eugene, OR, USA) and Alexa 488-conjugated goat anti-rabbit IgG (1:200, Invitrogen, Eugene, OR, USA). For negative and autofluorescence controls, the sections were incubated with PBS and normal rabbit or mouse serum instead of the specific antibody. The nuclei were counterstained with 0.4 mM/mL 4',6-Diamidino-2-Phenylindole, Dihydrochloride (DAPI; Molecular ProbesTM, Invitrogen, Eugene, OR, USA) for 15 minutes at room temperature. Finally, the specimens were mounted in buffered glycerol and their images were visualized in an immunofluorescence microscope (OLYMPUS BX51), and digitally scanned at ×20 magnification using a Pannoramic 250 whole slide scanner (3DHistech, Budapest, Hungary).

2.4 Quantification by image analysis

To measure the IHC expression of each different marker and quantify protein expression, the TMA slides were digitally scanned at ×40 magnification using a Pannoramic 250 whole slide scanner (3DHistech, Budapest, Hungary). The stained TMA sections were analyzed using QuPath (version 0.2.3; Centre for Cancer Research & Cell Biology, University of Edinburgh, Edinburgh, Scotland), an open-source image analysis software platform (23). During the scoring process, we assessed all cores to manually exclude any invalid samples (less than 10% of tumor per core or artifact).

QuPath allowed us to use a simple, automated, and semi-assisted method to quantify the TMAs. We first submitted each scanned TMA slide to a series of automated evaluations: staining vector analysis; total tissue area detection; tumor separation from non-tumor areas; and cellular detection. We then established the threshold of positivity for each of our markers through trial and error, and sent the cells considered to be positive to validation by an expert pathologist before applying them to the full array. Since GAGs are part of the extracellular matrix, – either on the cell surface or secreted in the form of PGs,

as well as collagen types – the QuPath measurements we adopted were the percentage of positive tissue or expression within the tumor or stroma. Henceforth, we will refer to low expression whenever positive cell density is equal to or below the mean expression in the cohort, and to high expression whenever the positive cell density falls above this mean cut.

2.5 Transmission electron microscopy

Tissues were fixed in 2% glutaraldehyde buffer and post-fixed in 1% OsO4. The samples were then washed overnight in 0.9% saline solution containing uranyl and sucrose and soaked in Epon. Finally, the samples were stained with uranyl acetate and lead citrate and examined with a JEOL JEM-1010 electron microscope.

2.6 Data mining

The UALCAN platform (http://ualcan.path.uab.edu/), a user-friendly web resource, was used to analyze data from The Cancer Genome Atlas (TCGA) (24, 25) to investigate the relative expression of mRNA from our interest genes (E-cadherin, β-catenin, collagens type I, III, IV and V, WNT1, WNT3A, WNT5A, WNT5B, and SPARC) in ADC, SqCC, and normal samples. The mRNA expression level of the analyzed genes was normalized to transcription per million reads, and only a P-value not greater than 0.01, according to Student's T-test, was significant. The UALCAN platform also was consulted to obtain the expression of ADC proteins present in the Clinical Proteomic Tumor Analysis Consortium (CPTAC) dataset (24, 26, 27). Protein expression was normalized according to Z-score. This database did not contain values for WNT1, WNT3A, and WNT5B proteins. Also, the database has not incorporated data for SqCC samples until this moment.

In this study, the clinicopathological characteristics of patients, as well as the expression levels of mRNA for the markers of interest were obtained in cBio Cancer Genomics (cBioPortal) (28, 29). This information was collected from the TCGA (Pan Cancer Atlas) database for both ADC and SqCC. To be consistent with our study model, we selected patients in pathological stages I, II, or IIIA. Data from this database that did not meet this criterion were excluded.

The prognostic significance value of proteins of interest in this study in NSCLC was evaluated using the Kaplan-Meier plotter database (30). Patient samples were divided into two cohorts, according to the median expression of each gene (high vs. low expression). The Kaplan-Meier plotter database calculated the log-rank P value and hazard ratio (HR) with 95% confidence intervals (CI).

The mRNA expression of enzymes involved in GAGs biosynthesis needs to be evaluated otherwise since their building blocks are polysaccharides synthesized in Golgi. Thus, the mRNA analysis for the GAGs on the databases described did not include them.

The String platform was consulted to reveal the functional interactions between the proteins evaluated in this present study and to map their protein-protein interaction (PPI) network (31, 32). We also used Metascape (33) to elucidate the function and biological processes involved in the enrichment of the genes corresponding to the proteins of our interest.

2.7 Data analysis

Since our data presented a distribution close to normal, we used the T-test, ANOVA, and Pearson's chi-square test to associate protein expression, clinicopathologic characteristics, and histotypes. The Cox proportional hazards model was then used to analyze the association between OS rate and other covariances. Any parameters that were thought to be clinically relevant or had a $P \leq 0.02$ in univariate analysis were considered for multivariate analysis. However, the data on the TCGA database presented a non-normal distribution, so we used non-parametric statistical tests instead. Finally, we used the statistical software IBM SPSS (version 22; Armonk, NY, USA) and RStudio to perform the analyses and plot the graphics. A P-value<0.05 was considered significant.

3 Results

3.1 Characterization of the NSCLC study cohort

Table 1 shows the demographic and clinicopathologic characteristics of the 120 patients included in the study. Patients had a mean age of 65 years old (range, 30-80 years old) and were evenly distributed between male (66, 55%) and female (54, 45%). 71 patients (79.8%) had a history of tobacco use. We histologically classified most samples as ADC (73, 60.8%), followed by SqCC (40, 33.3%), and LCC (7, 5.8%). 91 patients (75.8%) were in T1 and T2 stage, with a greater proportion of patients (79, 65.8%) in the N0 lymph node stage. After surgical resection, the mean tumor size was 4.46 cm (range, 1 to 13 cm) and a pathological classification identified 39 patients in stage I (32.5%), 56 in stage II (46.7%), and 25 in stage IIIA (20.8%). 49 patients (42.2%) received adjuvant chemotherapy, and 32 patients (27.6%) received adjuvant radiotherapy. The mean follow-up was 57.3 months (range, 0-181) and, during this period, 31 patients (31.3%) were relapsed, and 65 patients died.

10.3389/fonc.2022.1042766

TABLE 1 Demographic and clinicopathologic characteristics of the patients (N=120).

Characteristics	Number (%) of Patients
Age (years)	
Median (range)	65 (30 – 88)
≤65	64 (53.3%)
>65	56 (46.7%)
Gender	
Male	66 (55.0%)
Female	54 (45.0%)
Smoke Status[a]	
Smoker/Former-Smoker	71 (79.8%)
Non-Smoker	18 (20.2%)
Smoking load[a] (pack/years)	
Mean (range)	53.9 (1.5-150)
Histological subtypes	
Adenocarcinoma	73 (60.8%)
Squamous cell carcinoma	40 (33.3%)
Large cell carcinoma	7 (5.8%)
T stage†	
T1	30 (25.0%)
T2	61 (50.8%)
T3	23 (19.2%)
T4	6 (5.0%)
N stage†	
N0	79 (65.8%)
N1	24 (20.0%)
N2	17 (14.2%)
Tumor size (cm)[a]	
Mean (range)	4.46 (1.0-13.0)
Pathological Stage†	
I	39 (32.5%)
II	56 (46.7%)
IIIA	25 (20.8%)
Adjuvant treatment	
Chemotherapy[a]	
No	67 (57.8%)
Yes	49 (42.2%)
Radiotherapy[a]	
No	84 (72.4%)
Yes	32 (27.6%)
Follow up (months)[a]	
Mean (range)	57.3 (0-181)
Relapse [a]	31 (31.3%)
Status[a]	
Death	65 (65.7%)

[a]Some cases had missing follow-up information: smoke status (31); Smoking load (53); tumor size (2); Relapse (21); Chemotherapy (4); Radiotherapy (4); Status (21).
† According to 8th Edition International Association for the Study of Lung Cancer (21).

3.2 Epithelium-to-mesenchymal morphometric variables

As a first approach, we examined the epithelium-to-mesenchymal transition (EMT) phenotype through E-cadherin and β-catenin protein expression using IHC and IF. In addition, we observed the epithelium junctions and the ultrastructural pattern using TEM.

The mean expression of E-cadherin in tumor cells was 23.99% ± 1.49 positive cells. For β-catenin, the mean expression was 21.24% ± 0.90 positive cells. When we compared different histological subtypes, we observed that LCC samples expressed lower levels of both markers when compared to ADC and SqCC (Figures 1A, B), with a significant difference between LCC and SqCC for E-cadherin expression (P=0.03); whereas ADC and SqCC had similar behaviors of expression for both markers. Supplementary Table 1 shows the distribution of these two markers by histological subtypes.

Morphologically, E-cadherin and β-catenin were expressed at the cell boundary in all three major histotypes (Figures 2A–R; Figures 3A–L - the negative control can be checked in Supplementary Figure 1), reflecting the ultrastructural pattern that is characterized by the presence of functional adherent junctions (Figures 4A1–F1).

In LCC, both EMT markers were arranged just beneath the plasma membrane, forming a thin cortical barrier around each malignant cell (Figures 2C, F, 3A, C), in agreement with the short microvilli and cell interdigitation seen under TEM (Figures 4A1–F1). Finger-like projections interdigitated adjacent cells, whereas desmosomes linked small groups of cells (Figure 4A1). Notably, invasive LCC was characterized by long protrusions of cells dissecting the basement membrane (BM) to invade the surrounding matrix (Figure 4B1). LCC stained for E-cadherin and β-catenin both in the plasma membrane and in the cytoplasm (Figures 2B, E; respectively). This suggests a modest tumor proliferation and isolated invasion of the surrounding matrix.

In ADC, both EMT markers were even more evident in the plasma membrane (Figures 2I, L, 3E, G). However, we also found the EMT markers in the cytoplasm of some cells. This indicates focal invadopodia, which attach to the matrix, as seen under TEM (Figure 4C1). Invasive adenocarcinomatous cells showed strong staining of E-cadherin and β-catenin in the cytoplasm and plasma membrane (Figures 2H, K), which indicates invasion of the surrounding stroma by groups of malignant cells with fusiform shape (Figure 4D1) and high tumor proliferation status.

Lastly, in SqCC, both EMT markers were highly evident and formed dots along the plasma membrane (Figures 2O, R), with

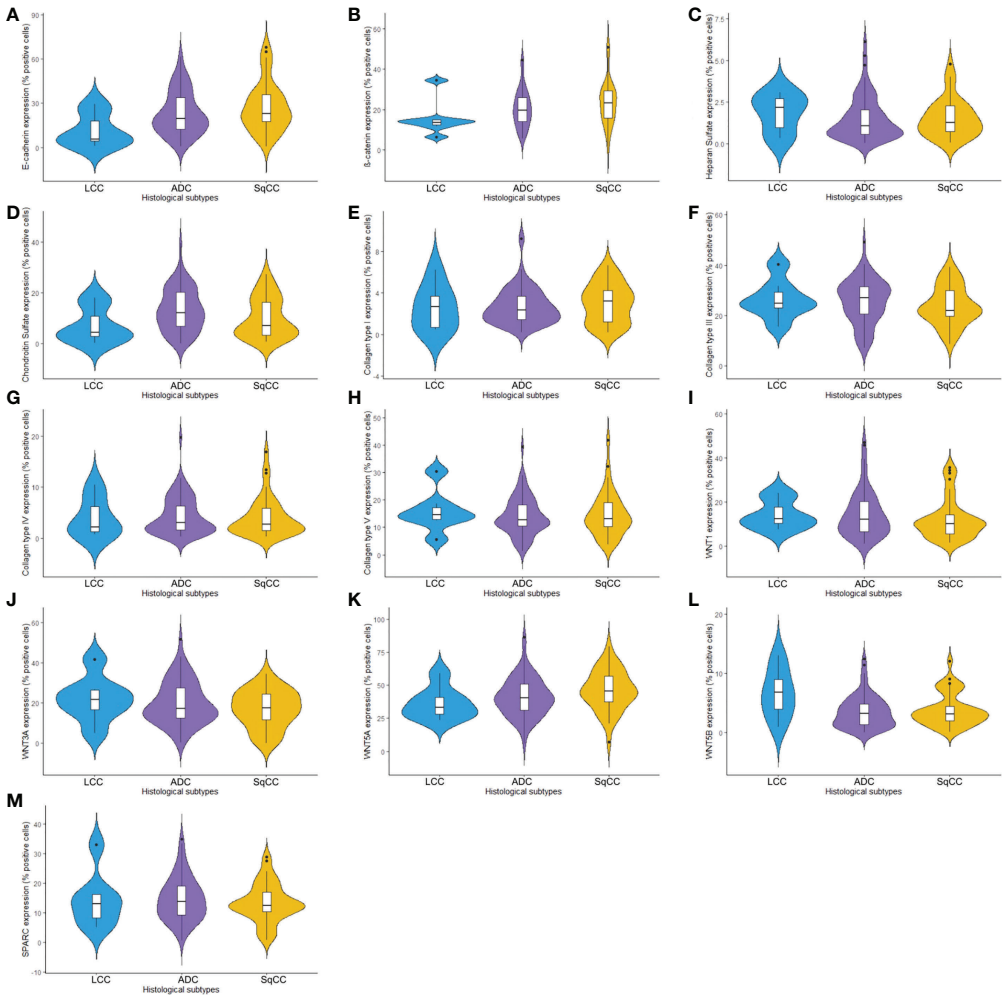


FIGURE 1
Violin plots graphs showing proteins of the epithelial-mesenchymal transition process, matricellular, and the Wnt signaling pathway expression analyzed by QuPath (N=120). The protein expression was demonstrated between three different histologic types in non-small cells lung cancer for **(A)** E-cadherin, **(B)** β-catenin, **(C)** Heparan sulfate, **(D)** Chondroitin sulfate, **(E)** Col I, **(F)** Col III, **(G)** Col IV, **(H)** Col V, **(I)** WNT1, **(J)** WNT3A, **(K)** WNT5A, **(L)** WNT5B, and **(M)** SPARC. LCC, large cell carcinoma; ADC, lung adenocarcinoma; SqCC, lung squamous cell carcinoma; Col, collagen type.

invasive squamous cells also showing expression of E-cadherin and β-catenin in the cytoplasm (Figures 2N, Q). The TEM of SqCC samples showed adjacent cells connected by junctional complexes consisting of desmosomes (Figure 4F1). During the invasion of the surrounding matrix, we observed an evident spindle cell transformation with spindle-like projections extending into the BM, and the detachment of squamous cells.

The above results then suggest that, during EMT, the strong expression of E-cadherin and β-catenin seen at light microscopy reflects the ultrastructure of fragmentation and loss of continuity of adherent epithelial junctions. This enables ECM invasion by individual malignant cells, in the case of LCC, and cells groups, in the case of ADC and SqCC.

3.3 Extracellular matrix morphometric variables

Next, we examined morphometric variables linked to the ECM, including the glycosaminoglycans (GAGs) heparan sulphate (HS) and chondroitin sulphate (CS), and collagen type I (Col I), type III (Col III), type IV (Col IV), and type V (Col V).

Starting our analysis with the GAGs, we observed that the mean expression of HS was 1.55% ± 0.11 positive cells. This behavior coincided with weak immunostaining on all three major histotypes (Figure 5). Most ADC and SqCC samples showed lower expression of HS when compared with LCC, but with no statistical significance (Figure 1C; Supplementary Table 1).

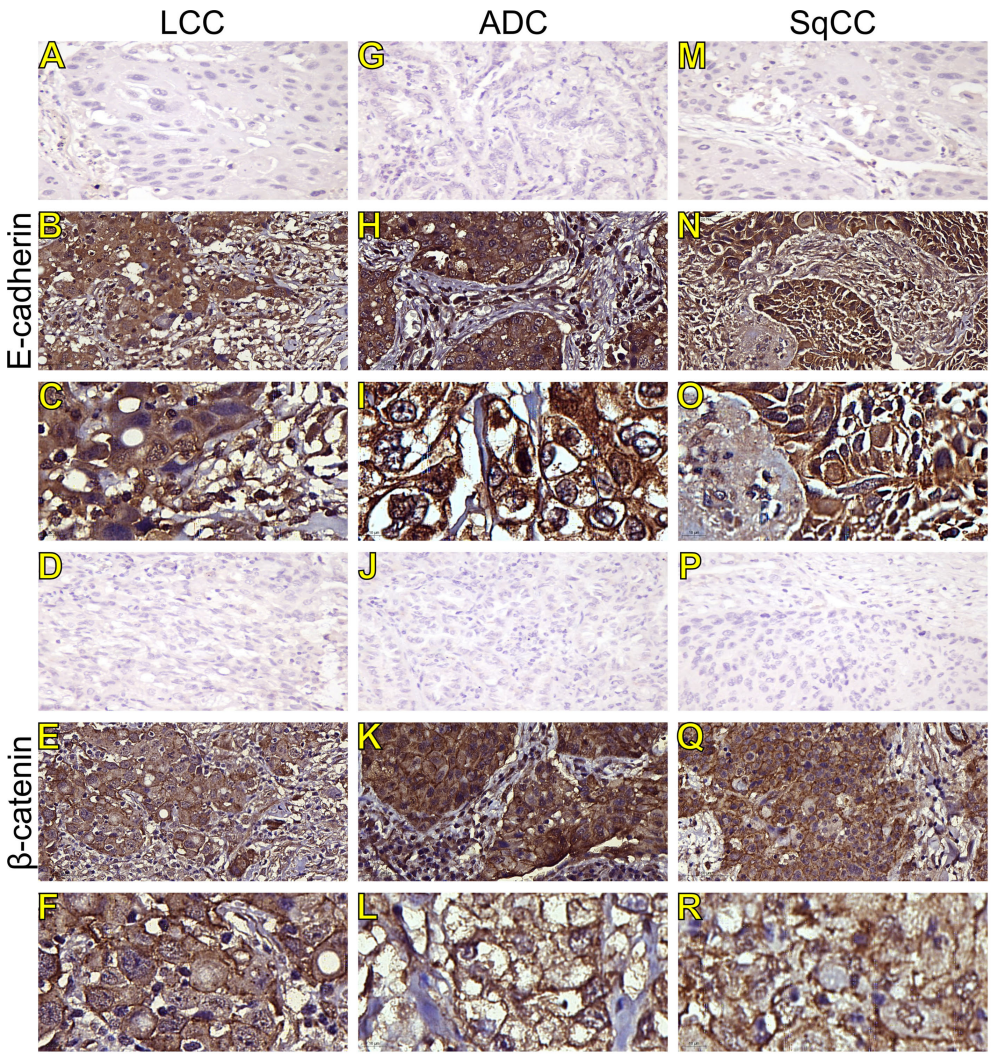


FIGURE 2
Immunohistochemistry expression of epithelial-to-mesenchymal transition-related markers in tumors and invasive groups of large cell carcinoma (**B, C**, E-cadherin; **E, F**, β-catenin; respectively), adenocarcinoma (**H, I**, E-cadherin; **K, L**, β-catenin; respectively) and squamous cell carcinoma (**N, O**, E-cadherin; **Q, R**, β-catenin; respectively) (N=120). For both markers, the first line illustrates the negative control **(A, D, G, J, M, P)**. Original magnification: 40X and 100X. LCC, large cell carcinoma; ADC, lung adenocarcinoma; SqCC, lung squamous cell carcinoma.

In contrast, the mean expression of CS was 11.91% ± 0.75 positive cells, which agrees with the strong immunostaining similarly observed in all three histological subtypes (Figure 1D; Supplementary Table 1).

CS arrangement in LCC took place just beneath the plasma membrane. This CS arrangement formed a thick cortical barrier around it (Figures 5E, F), in agreement with the thick BM seen under TEM (Figure 4A2). Furthermore, there was strong CS staining in both the plasma membrane and in the cytoplasm (Figures 5E, F; respectively), a sign of rapid proliferation of the malignant cells invading the interstitial ECM.

In ADC, the presence of CS was even more evident in the BM (Figures 5K, L), although also being found in the cytoplasm of some cells. This indicates focal invadopodia, which attach to the BM, as described under TEM (Figure 4C2). Invasive adenocarcinomatous cells presented intense CS staining in the cytoplasm and plasma membrane (Figures 5K, L) The intense CS staining shows a rapid invasion of the surrounding ECM by cell groups with high tumor proliferation status. In contrast, CS in invasive SqCC was considerably more evident than in others histotypes, forming a thick barrier along the BM with strong CS expression in the cytoplasm (Figures 5Q, R).

When we quantify the different collagen types in this context, the mean collagen expression was 2.80% ± 0.16 positive fibers for Col I, 25.04% ± 0.76 positive fibers for Col III, 4.29% ± 0.33 positive fibers for Col IV, and 14.41% ± 0.66

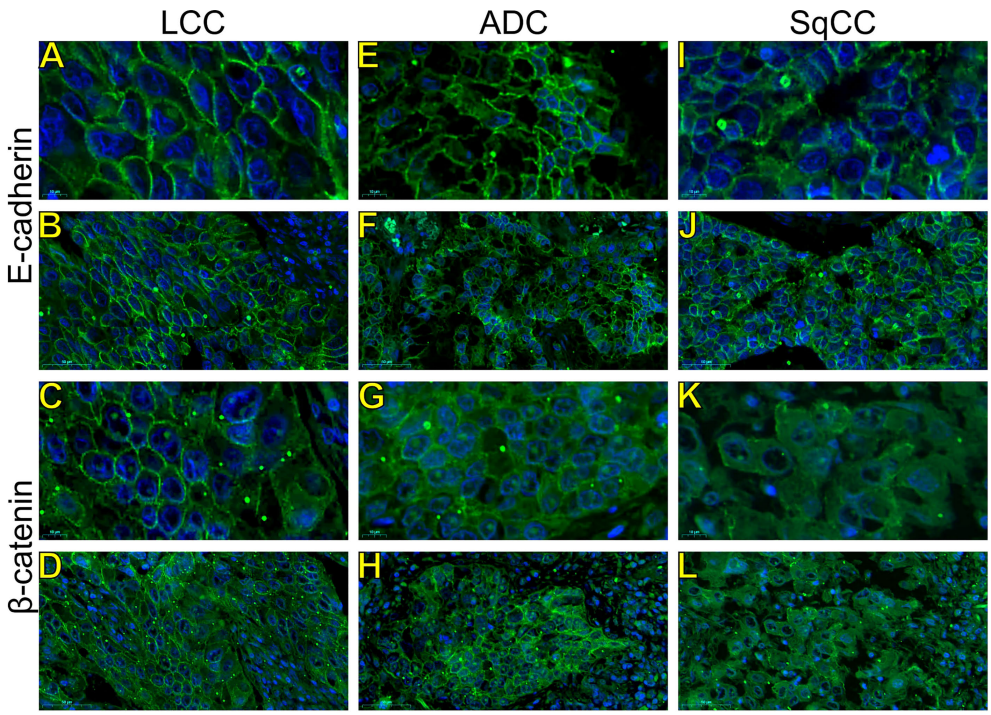


FIGURE 3
Expression of epithelial-to-mesenchymal transition-related markers. Images visualized under immunofluorescence microscope showing tumors and invasive groups of large cell carcinoma (**A, B**, E-cadherin; **C, D**, β-catenin; respectively), adenocarcinoma (**E, F**, E-cadherin; **G, H**, β-catenin; respectively) and squamous cell carcinoma (**I, J**, E-cadherin; **K, L**, β-catenin; respectively) (N=120). Original magnification: 100X and 40X. LCC, large cell carcinoma; ADC, lung adenocarcinoma; SqCC: lung squamous cell carcinoma.

positive fibers for Col V. The expression of all four makers was similar across the three histological subtypes (Figures 1E–H, Supplementary Table 1).

Figures 6–9 show the IF co-analyses of CS/HS and collagen types (the respectively negative control can be found in Supplementary Figure 2). The interstitial matrix of ADC and SqCC showed a strong reddish fluorescence of Col I fibers; conversely, CS is represented by a strong greenish fluorescence along the BM in ADC and LCC (Figure 6). While Col III fibers are seen as in a strongly reddish fluorescent color in the interstitial matrix of LCC, ADC, and SqCC (Figures 7B, F, J), CS showed a weak greenish fluorescence in the BM of the three histotypes (Figures 7C, G, K). The intensity of refringence of Col IV (red) and HS (green) in the BM was strong in ADC and weak in SqCC and LCC (Figure 8). Notably, Col V (red) and CS (green) were more fluorescent in LCC compared to ADC and SqCC (Figure 9).

The above results suggest that, after EMT, the strong expression of CS, with the strong refringence of Col III and Col V, may enhance the motility of invasive cells from the three histotypes through the BM into the interstitial ECM.

3.4 Wnt signaling pathway and SPARC morphometric expression

We then evaluated Wnt signaling through WNT1, WNT3A, WNT5A, and WNT5B expression in NSCLC histotypes. The Wnt protein family triggers a relevant cascade which regulates development and is associated with cancer. We also examined the morphometric variables of SPARC, a multifunctional glycoprotein involved with the EMT, ECM remodeling, and Wnt proteins.

The mean expression of WNT1, WNT3A, WNT5A, and WNT5B by tumor cells was respectively 13.55% ± 0.89, 19.16% ± 0.94, 42.47% ± 1.35, and 3.74% ± 0.26 positive cells, which coincides with a moderate expression of WNT1 and WNT3A, a weak expression of WNT5B, but an intense cytoplasmic expression of WNT5A in the three histologic subtypes (Figure 10). However, only the expression of WNT5A showed a significant statistical difference between histological subtypes (SqCC versus non-squamous tumors, Figure 1K; Supplementary Table 1).

For SPARC, the mean expression in tumor stroma was 14.24% ± 0.65 positive cells, with similar proportions among histotypes. Although cytoplasmic staining was predominantly

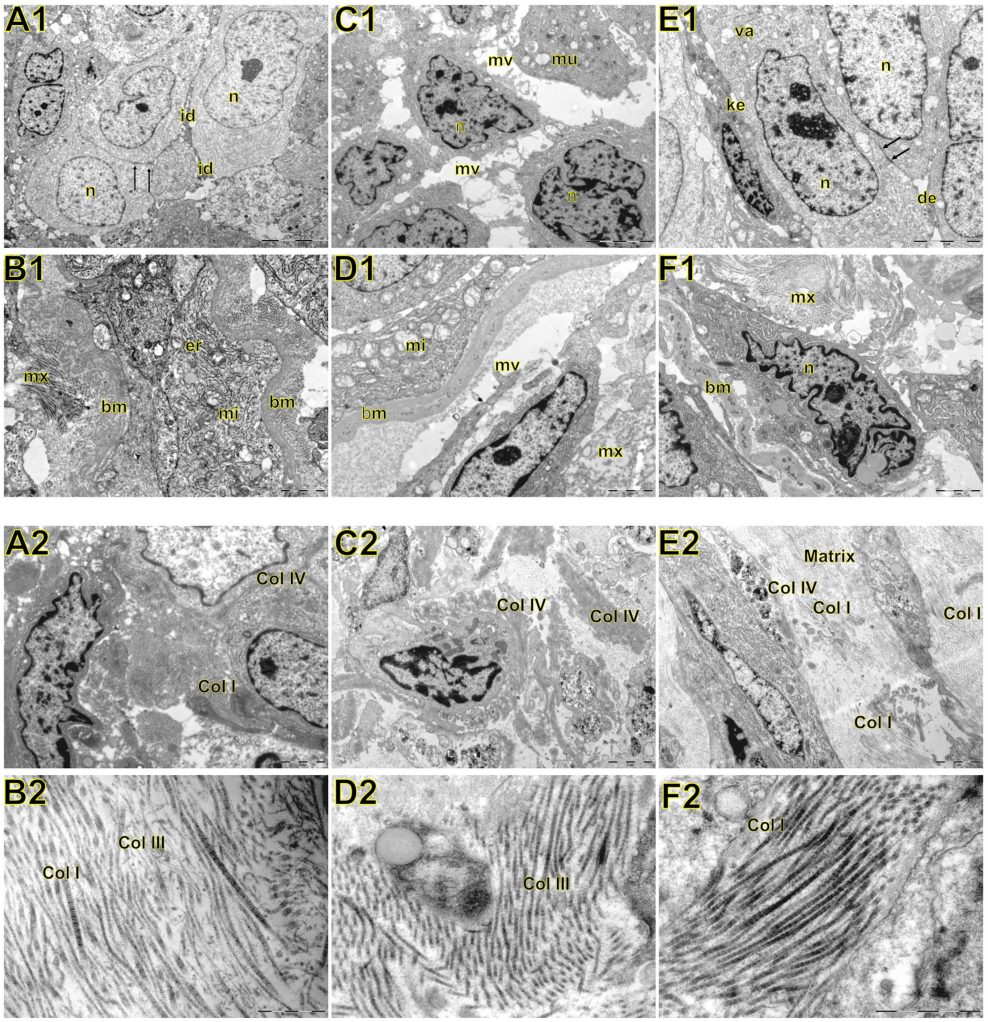


FIGURE 4
Transmission electron microscopy of large cell carcinoma **(A, B)**, adenocarcinoma **(C, D)**, and squamous cell carcinoma **(E, F)** visualized inside the tumor, basement membrane, and interstitial matrix. Above **A1:** large cell carcinoma showing large tumour cells with abundant light or little dark cytoplasm. Nuclei (n) are euchromatic and frequently display more or less deep invaginations with one to two prominent nucleoli. The cell membrane presented very short microvilli (mi) densely packed. Adjacent cells were interdigitated (id) with neighboring cells. In **B1**, note invasive large cell carcinoma with prominent organelles including mitochondria (mi) and endoplasmic reticulum (er) protruding and dissecting the basement membrane (bm) to invade the surrounding matrix (mx). **C1:** adenocarcinoma showing round cells with marked irregularity of nucleus, light cytoplasm, and interdigitation by sparse and short microvilli (mv). Mucin vacuoles (mu) can be visualized disperses in the cytoplasm. In **D1**, invasive cells protrude the basement membrane (bm) showing numerous organelles such as mitochondria (mi), and short microvilli projected toward the matrix (mx) assuming the fusiform phenotype. **E1**: squamous cell showing cells partially cohesive by junctional complexes (desmosomes, des). Some cells exhibit a pale and regular nucleus (n) with evident nucleolus. Adjoined cells are connected by numerous adherens junctions (arrows) in their lateral domains showing cellular polarity. Some autophagosomes (va) can be observed in the cytoplasm. The dark cells exhibit keratin filaments in the cytoplasm (ke). In **F1**, invasive squamous cells assume elongated form, marked irregularity of nucleus dissecting basement membrane (bm) and collagen fibers of the matrix (mx). Below **A2:** at low magnification large cell carcinoma showing abnormal large tumour cells with abundant light or little dark cytoplasm, densely cohesive. In **B2**, at high magnification note the interstitial matrix composed buy fibrillar collagens (fc) intermixed in a scant amorphous matrix. **C2**: a low magnification of adenocarcinoma showing aberrant round cells with marked irregularity of nucleus, thick basement membrane matrix in the junction of tumor cells with the interstitial matrix suggesting invadopodium. In **D2**, a high magnification shows fibrillar collagen (fc). **E2**: a low magnification showing elongated invasive squamous cells with the amorphous collagen (ac) along the basement membrane matrix. The interstitial matrix is abundant and composed of fibrillar collagen (fc) immersed in an amorphous matrix composed of heparan sulfate and chondroitin sulfate. In **F2**, a high magnification exhibited mainly thick and distorted fibers and microfibrils of fibrillar collagen (fc) in the interstitial matrix.

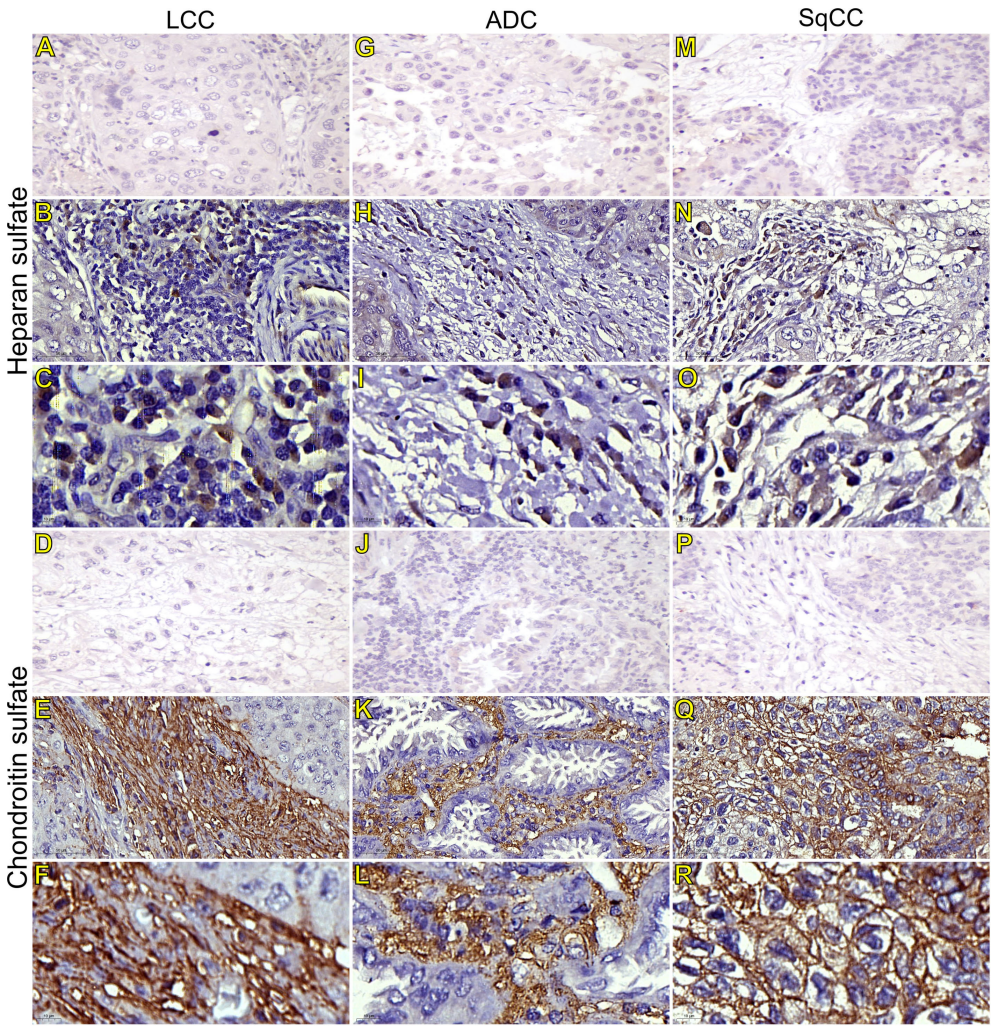


FIGURE 5
Immunohistochemistry expression of glycosaminoglycans markers in tumors and invasive groups of large cell carcinoma (**B**, **C**, heparan sulfate; **E**, **F**, chondroitin sulfate; respectively), in adenocarcinoma (**H**, **I**, heparan sulfate; **K**, **L**, chondroitin sulfate; respectively), and in squamous cell carcinoma (**N**, **O**, heparan sulfate; **Q**, **R**, chondroitin sulfate; respectively) (N=120). For both markers, the first line illustrates the negative control **(A, D, G, J, M, P)**. Original magnification: 40X and 100X. LCC, large cell carcinoma; ADC, lung adenocarcinoma; SqCC: lung squamous cell carcinoma.

positive in the stroma cells in all NSCLC cases, we also observed weak SPARC staining in tumor cells across all three histological subtypes (Figure 10). There was no statistical difference in SPARC levels between the three histotypes (Figure 1M, Supplementary Table 1).

These findings show the importance of WNT5A expression across histotypes, and its possible association with tumor progression.

3.5 Correlation between EMT, WNTs, SPARC, and ECM morphometric variables

The next step was to explore whether the WNT proteins or SPARC mechanistically orchestrated the phenotypic and molecular changes in NSCLC. This would suggest EMT and ECM as intrinsic links between development and cancer progression.

Figure 11 shows the correlation between the tumor and stroma morphometric variables. E-cadherin correlated strongly with β-catenin (ρ=0.617, P<0.001) and WNT5A (ρ=0.672, P<0.001), and moderately with Col III (ρ=0.336, P<0.001). β-catenin also correlated strongly with WNT5A (ρ=0.693, P<0.001).

HS moderately correlated with WNT3A (ρ=0.401, P<0.001) and SPARC (ρ=0.376, P<0.001), while CS moderately correlated with WNT1 (ρ=0.406, P<0.001). Col IV was moderately inversely associated with WNT1 (ρ =-0.346, p<0.001), whereas Col V moderately correlated with Col III (ρ=0.329, P<0.001).

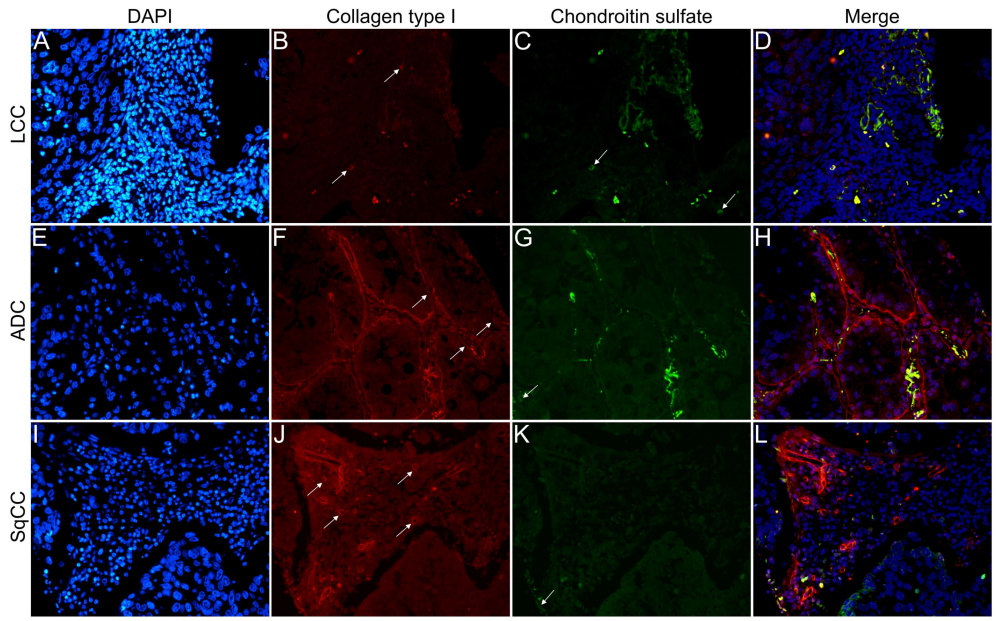


FIGURE 6
Co-analysis of immunofluorescence of chondroitin sulfate (green; **C**, **G**, **K**) and collagen type I (red; **B**, **F**, **J**) in three different histological subtypes of non-small cell lung carcinoma (N=120). The stained nuclei are represented in blue (DAPI; **A**, **E**, **I**). Images **D**, **H**, **L** represent the merge of the same field of these three stains. White arrows indicate positive expression of the markers. Original magnification: 40x. LCC, large cell carcinoma; ADC, lung adenocarcinoma; SqCC: lung squamous cell carcinoma.

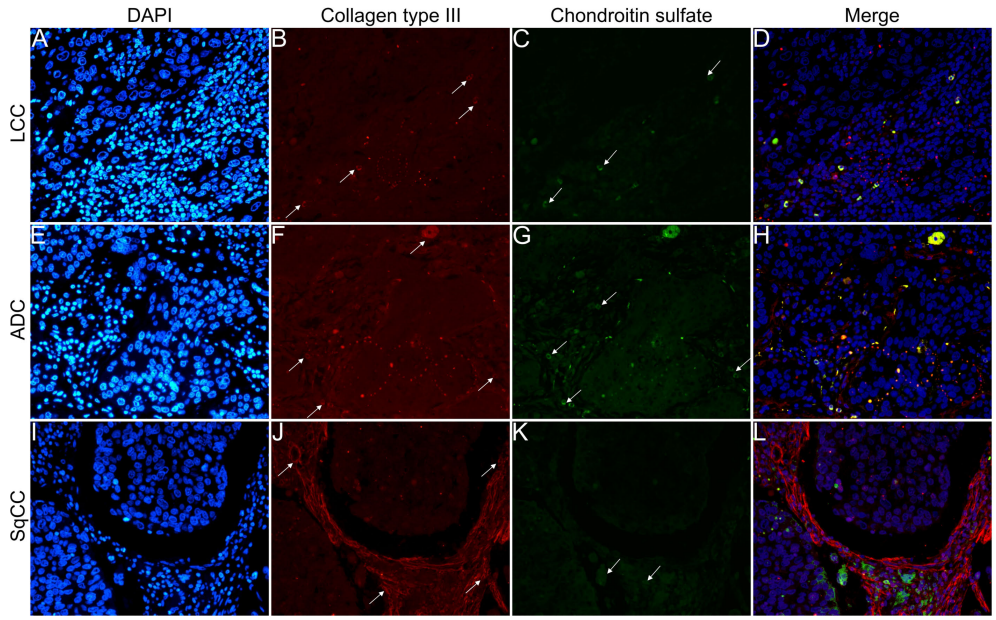


FIGURE 7
Co-analysis of immunofluorescence of chondroitin sulfate (green; **C**, **G**, **K**) and collagen type III (red; **B**, **F**, **J**) in three different histological subtypes of non-small cell lung carcinoma (N=120). The stained nuclei are represented in blue (DAPI; **A**, **E**, **I**). Images **D**, **H**, **L** represent the merge of the same field of these three stains. White arrows indicate positive expression of the markers. Original magnification: 40x. LCC, large cell carcinoma; ADC, lung adenocarcinoma; SqCC: lung squamous cell carcinoma.

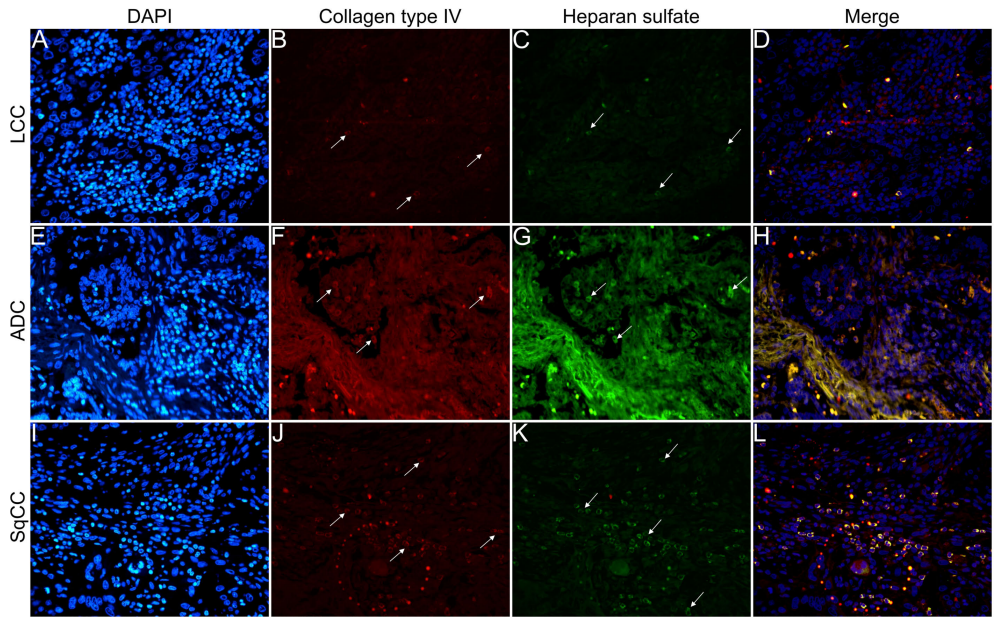


FIGURE 8
Co-analysis of immunofluorescence of heparan sulfate (green; **C**, **G**, **K**) and collagen type IV (red; **B**, **F**, **J**) in three different histological subtypes of non-small cell lung carcinoma (N=120). The stained nuclei are represented in blue (DAPI; **A**, **E**, **I**). Images **D**, **H**, **L** represent the merge of the same field of these three stains. White arrows indicate positive expression of the markers. Original magnification: 40x. LCC, large cell carcinoma; ADC, lung adenocarcinoma; SqCC: lung squamous cell carcinoma.

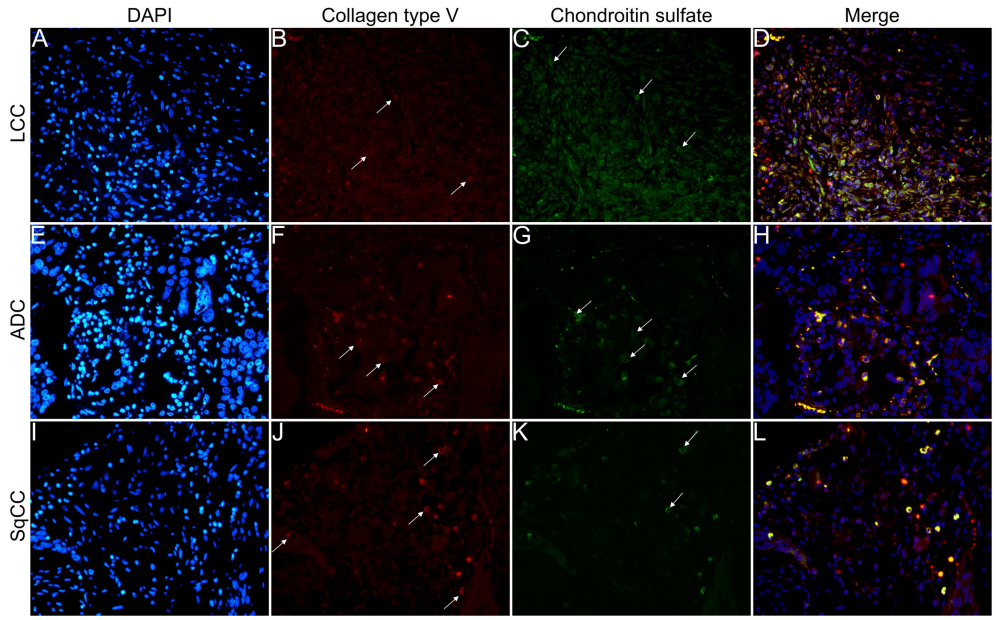


FIGURE 9
Co-analysis of immunofluorescence of chondroitin sulfate (green; **C**, **G**, **K**) and collagen type V (red; **B**, **F**, **J**) in three different histological subtypes of non-small cell lung carcinoma (N=120). The stained nuclei are represented in blue (DAPI; **A**, **E**, **I**). Images **D**, **H**, **L** represent the merge of the same field of these three stains. White arrows indicate positive expression of the markers. Original magnification: 40x. LCC, large cell carcinoma; ADC, lung adenocarcinoma; SqCC: lung squamous cell carcinoma.

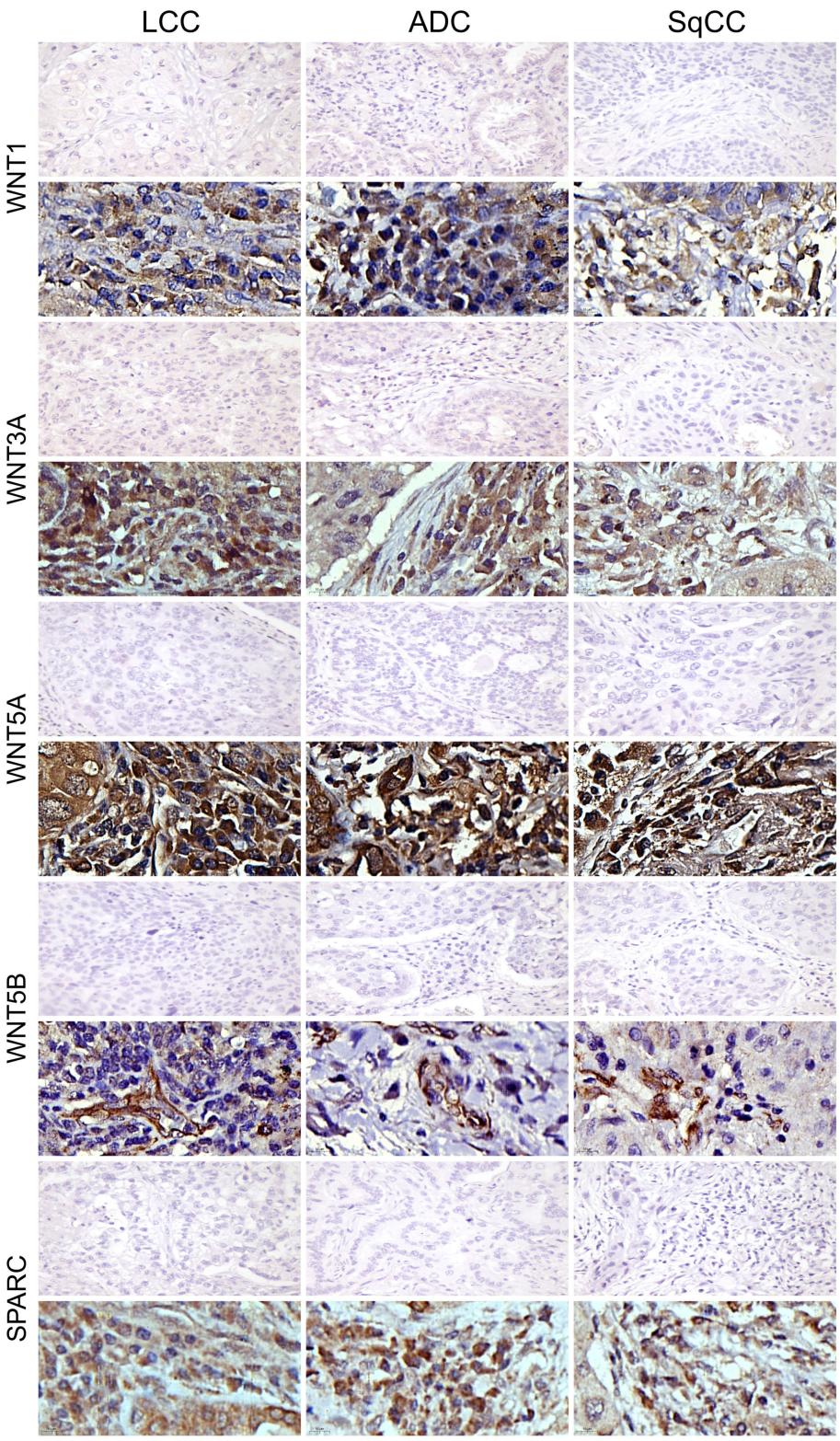


FIGURE 10
Immunohistochemistry expression of WNTs and SPARC in large cell carcinoma, adenocarcinoma, and squamous cell carcinoma (N=120). For both markers, the first line illustrates the negative control. Original magnification: 40X and 100X. LCC, large cell carcinoma; ADC, lung adenocarcinoma; SqCC: lung squamous cell carcinoma.

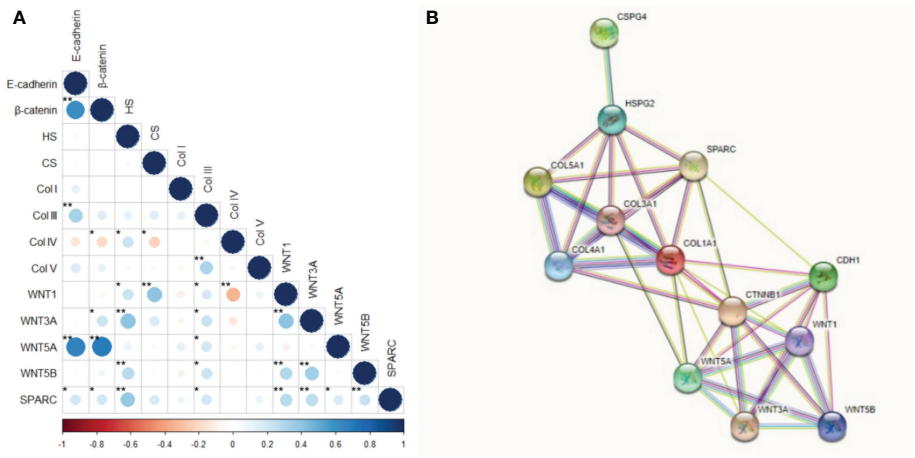


FIGURE 11
Correlation and protein-protein interaction network. **(A)** Correlation between the interest markers. The graduation of colors represents the positive or negative correlation. The size of the dot represents the Spearman's rho, larger dots have values closer to |1|, therefore, more strong correlation. *P<0.05; **P<0.01. **(B)** Protein-protein interaction network obtained into the STRING tool for our interest proteins.

WNT1 also had a moderate correlation with WNT3A (ρ=0.403, P<0.001), and WNT3A moderately correlated with WNT5B (ρ=0.340, P<0.001).

These results suggest the WNT5A pathway drove the EMT and the synthesis of both Col III signaling and Col V. On the other hand, WNT3A and SPARC regulated HS, while WNT1 directly regulated CS. A WNT1 negative feedback loop controlled the synthesis of Col IV along the BM.

3.6 Associations between clinicopathological features and morphometric variables

Table 2 shows the clinicopathological characteristics stratified by E-cadherin, β-catenin, WNTs signaling proteins, and SPARC. Statistical significance was found between lower E-cadherin expression and younger patients (P=0.04), patients with LCC (P=0.03), and tumors smaller than 3 cm (P=0.02). We also found statistical significance between tumors smaller than 3 cm and a higher WNT1 expression (P=0.05) and lower WNT5A expression (P=0.04). Patients with LCC showed higher WNT5B expression (P=0.01).

We found borderline significance between lower WNT1 expression in male patients (P=0.07) and stage II tumor versus stage I (P=0.06). This borderline significance suggests that increased WNT1 expression occurs at earlier stages of carcinogenesis, which then decreases as the tumor grows. There was also borderline significance between lower expression of WNT3A and younger patients (P=0.07), and between tumors larger than 3 cm (P=0.07). Finally, lower SPARC expression had a borderline significance in patients who had received adjuvant chemotherapy (P=0.07).

We did not find a significant association between tobacco history, T stage, N stage, relapse, and the makers under investigation. β-catenin expression failed to show any significant association with clinicopathologic characteristics.

Table 3 shows the association between clinicopathological features, GAGs, and collagen types. There was a significant association between lower CS expression and tumors larger than 3 cm (P=0.05). In turn, lower Col IV expression was associated with younger (P=0.05) and female (P=0.004) patients. Tumors in the N2 stage tended to express lower Col V in the stroma (p=0.05) when compared to N0, as did tumors from patients who received adjuvant chemotherapy (P=0.03).

When compared to stage I (P=0.06), we found a borderline significance between lower CS expression and tumor stage II, and between lower Col IV expression in patients who received chemotherapy (P=0.06). We did not find any statistical differences between clinicopathologic characteristics and HS, Col I, and Col III.

3.7 Prognostic value of morphometric variables

Out of 120 patients in our cohort, 65 progressed to death. In a univariable analysis (Table 4), OS for the entire cohort was significantly influenced by: gender (HR 0.39 for male versus female, CI 0.20-0.75, P=0.004), T stage (HR 2.69 for T3-T4 versus T1-T2, CI 1.46 – 4.94, P=0.001), tumor size (HR 1.95 for > 3cm versus ≤ 3cm, CI 1.08-3.53, P=0.026), metastases (HR 2.98 for present versus absent, CI 1.60-5.53, P=0.001), and radiotherapy

TABLE 2 Association between clinicopathologic characteristics and mean expression (% positive expression) of E-cadherin, β-catenin, WNTs signaling proteins, and SPARC (*t* test and ANOVA, P<0.05).

Characteristics	E-cadherin	β-catenin	WNT1	WNT3A	WNT5A	WNT5B	SPARC
Age (years)							
≤65	**21.35 ± 1.66***	21.01 ± 1.26	13.89 ± 1.29	17.54 ± 1.27&	41.56 ± 1.82	3.86 ± 0.39	14.05 ± 0.84
> 65	**27.00 ± 2.30**	21.51 ± 1.32	13.19 ± 1.26	20.96 ± 1.39	43.52 ± 2.03	3.61 ± 0.36	14.47 ± 1.03
Gender							
Male	25.24 ± 1.91	21.10 ± 1.14	12.11 ± 1.04&	18.57 ± 1.25	43.21 ± 1.84	3.79 ± 0.32	13.83 ± 0.86
Female	22.46 ± 2.08	21.43 ± 1.48	15.39 ± 1.53	19.92 ± 1.46	41.58 ± 2.01	3.69 ± 0.43	14.77 ± 1.01
Tobacco History							
Smoker/Former-smoker	28.15 ± 1.79	23.06 ± 1.23	13.28 ± 1.28	18.45 ± 1.18	44.61 ± 1.69	3.74 ± 0.33	14.62 ± 0.86
Non-smoker	27.85 ± 4.10	22.00 ± 2.22	13.80 ± 2.52	21.27 ± 2.64	43.16 ± 4.05	3.16 ± 0.63	13.44 ± 1.53
Histotypes							
ADC	**23.23 ± 1.70**	20.59 ± 1.08	14.24 ± 1.22	20.00 ± 1.26	41.04 ± 1.71	**3.45 ± 0.32**	14.69 ± 0.86
SqCC	**27.57 ± 2.63**	23.48 ± 1.73	12.19 ± 1.48	17.11 ± 1.49	46.18 ± 2.39	**3.79 ± 0.44**	13.42 ± 1.03
LCC	**11.52 ± 4.17*#**	15.54 ± 3.35	14.17 ± 2.45	22.13 ± 4.30	36.28 ± 4.58	**6.61 ± 1.57*##**	14.35 ± 3.53
T stage							
T1	21.55 ± 2.55	24.53 ± 1.94	15.76 ± 2.32	20.12 ± 2.05	41.37 ± 2.67	3.93 ± 0.60	13.61 ± 1.18
T2	24.78 ± 2.06	19.78 ± 1.09	13.50 ± 1.12	20.24 ± 1.24	42.64 ± 1.87	3.83 ± 0.32	15.26 ± 1.01
T3	25.20 ± 3.58	20.86 ± 2.24	11.63 ± 1.89	16.51 ± 2.30	42.27 ± 3.45	3.32 ± 0.65	13.24 ± 1.16
T4	23.57 ± 3.92	21.55 ± 6.23	10.81 ± 2.92	14.08 ± 3.86	47.14 ± 5.70	3.63 ± 1.94	11.08 ± 3.14
N stage							
N0	25.22 ± 1.79	22.11 ± 1.13	13.55 ± 1.21	19.55 ± 1.13	42.37 ± 1.63	3.69 ± 0.31	14.81 ± 0.77
N1	20.11 ± 2.85	19.66 ± 2.29	13.63 ± 1.59	20.31 ± 2.49	41.77 ± 3.74	4.44 ± 0.73	12.59 ± 1.63
N2	23.73 ± 3.65	19.42 ± 0.91	13.48 ± 1.98	15.90 ± 2.24	43.96 ± 2.69	3.03 ± 0.59	13.59 ± 1.81
Pathologic stage							
I	26.39 ± 2.66	24.03 ± 1.65	16.48 ± 2.00	20.71 ± 1.64	43.21 ± 2.46	3.85 ± 0.47	15.85 ± 1.20
II	23.02 ± 2.08	19.99 ± 1.30	11.83 ± 2.00&&	19.59 ± 1.34	41.78 ± 2.05	3.91 ± 0.33	13.95 ± 0.90
IIIA	22.05 ± 2.75	19.38 ± 1.89	12.19 ± 1.42	15.60 ± 2.29	42.97 ± 2.77	3.22 ± 0.74	12.36 ± 1.50
Tumor size							
≤3 cm	**20.60 ± 1.86***	22.18 ± 1.41	**15.80 ± 1.66***	21.29 ± 1.54	**39.42 ± 2.08***	4.07 ± 0.42	15.15 ± 1.05
>3 cm	**27.02 ± 1.96**	20.68 ± 1.20	**12.02 ± 0.96**	17.84 ± 1.17&	**45.05 ± 1.76**	3.60 ± 0.34	13.75 ± 0.85
Relapse							
No	23.96 ± 1.90	21.70 ± 1.25	14.02 ± 1.28	19.19 ± 1.20	42.60 ± 1.90	3.90 ± 0.37	15.09 ± 0.88
Yes	24.38 ± 2.71	19.98 ± 1.94	12.79 ± 1.37	17.05 ± 1.87	42.85 ± 2.66	3.47 ± 0.44	12.98 ± 1.29
Adjuvant treatment							
Chemotherapy							
No	22.39 ± 1.84	21.40 ± 1.23	14.27 ± 1.33	19.75 ± 1.24	41.45 ± 1.87	3.89 ± 0.36	15.48 ± 0.92
Yes	26.60 ± 2.33	21.24 ± 1.46	12.85 ± 1.26	18.97 ± 1.56	43.61 ± 2.07	3.72 ± 0.41	12.97 ± 0.91$
Radiotherapy							
No	24.48 ± 1.94	22.06 ± 1.02	13.93 ± 1.16	19.78 ± 1.17	42.57 ± 1.59	3.86 ± 0.32	14.55 ± 0.76
Yes	23.33 ± 3.09	19.45 ± 2.06	12.95 ± 1.45	18.48 ± 1.75	41.80 ± 2.82	3.71 ± 0.48	14.08 ± 1.39

*P>0.05; #SqCC vs. LCC - P=0.03; ##ADC vs. LCC - P=0.01, and SqCC vs. LCC - P=0.04; &P=0.07; &&Pathological Stage I vs. II – P=0.06; $P=0.06
ADC, lung adenocarcinoma; SqCC: lung squamous cell carcinoma; LCC, large cell carcinoma.
Bolded values refer to a P-value with statistical significance (P<0.05).

(HR 0.40 for No versus Yes, CI 0.21-0.76, P=0.005). We also observed that high SPARC and WNT3A expressions in the tumor stroma had a significant influence on OS (HR 0.55 for higher versus low expression, CI 0.30-1.00, P=0.050; HR 0.54 for higher versus lower expression, CI 0.30–0.99, P=0.046; respectively).

Conversely, in a multivariable analysis, gender, T stage, tumor size, metastases, adjuvant therapy, and SPARC were significantly associated with OS, whereas high HS expression had only a borderline association with OS (P=0.055). WNT3A and WNT5A were co-variables in this mathematical model (Chi-square 33.223; P<0.001). Mean OS was 97 months for patients with SPARC expression >14.25% compared to 65 months for patients with expression ≤14.25% (Supplementary Figure 3).

TABLE 3 Association between clinicopathologic characteristics and mean expression (% positive expression) of GAGs and collagen types (*t* test and ANOVA, P<0.05).

Characteristics	HS	CS	Col I	Col III	Col IV	Col V
Age (years)						
≤65	1.52 ± 0.16	12.21 ± 1.06	2.85 ± 0.21	23.67 ± 1.00	**3.64 ± 0.35***	13.63 ± 0.87
> 65	1.58 ± 0.15	11.59 ± 1.09	2.75 ± 0.28	26.60 ± 1.13	**5.03 ± 0.58**	15.32 ± 1.02
Gender						
Male	1.59 ± 0.16	10.95 ± 0.97	3.09 ± 0.23	25.52 ± 1.01	**5.12 ± 0.53**	13.94 ± 0.93
Female	1.50 ± 0.16	13.10 ± 1.18	2.47 ± 0.23	24.45 ± 1.16	**3.29 ± 0.34****	15.01 ± 0.96
Tobacco History						
Smoker/Former-smoker	1.63 ± 0.16	10.71 ± 0.93	2.91 ± 0.21	26.69 ± 0.99	3.82 ± 0.40	15.15 ± 0.93
Non-smoker	1.24 ± 0.22	14.52 ± 2.10	2.94 ± 0.46	25.78 ± 2.15	4.20 ± 0.66	15.79 ± 1.87
Histotypes						
ADC	1.52 ± 0.15	13.31 ± 0.98	2.72 ± 0.20	25.58 ± 1.03	4.41 ± 0.43	14.17 ± 0.82
SqCC	1.55 ± 0.18	10.20 ± 1.23	3.01 ± 0.30	23.80 ± 1.20	4.14 ± 0.60	14.63 ± 1.26
LCC	1.86 ± 0.43	7.18 ± 2.62	2.57 ± 0.82	26.50 ± 2.97	4.07 ± 1.43	15.85 ± 2.86
T stage						
T1	1.19 ± 0.16	13.98 ± 1.67	2.46 ± 0.34	24.47 ± 1.56	3.72 ± 0.52	15.50 ± 1.41
T2	1.76 ± 0.17	11.74 ± 1.07	2.99 ± 0.25	26.15 ± 1.13	4.92 ± 0.54	13.00 ± 0.82
T3	1.23 ± 0.20	9.67 ± 1.46	2.50 ± 0.31	23.44 ± 1.46	3.63 ± 0.68	16.61 ± 1.91
T4	2.43 ± 0.69	12.05 ± 2.55	3.91 ± 0.28	22.83 ± 2.55	3.44 ± 0.74	14.98 ± 1.54
N stage						
N0	1.49 ± 0.13	12.31 ± 0.97	2.78 ± 0.20	25.79 ± 0.93	4.41 ± 0.43	**15.50 ± 0.88**
N1	1.50 ± 0.28	10.69 ± 1.51	2.80 ± 0.37	23.10 ± 1.52	4.41 ± 0.75	**13.40 ± 1.13**
N2	1.91 ± 0.35	11.84 ± 1.97	2.95 ± 0.46	24.30 ± 2.33	3.59 ± 0.75	**10.85 ± 1.34*#**
Pathologic stage						
I	1.35 ± 1.16	14.30 ± 1.38	2.83 ± 0.34	26.90 ± 1.45	4.31 ± 0.67	14.84 ± 1.27
II	1.53 ± 0.16	10.34 ± 1.08&##	2.60 ± 0.22	24.68 ± 1.12	4.63 ± 0.48	14.93 ± 0.98
IIIA	1.95 ± 032	11.47 ± 1.57	3.38 ± 0.32	23.00 ± 1.38	3.66 ± 0.59	12.65 ± 1.30
Tumor size						
≤3 cm	1.56 ± 0.19	**13.72 ± 1.27**	2.67 ± 0.25	25.43 ± 1.17	3.84 ± 0.42	14.51 ± 0.97
>3 cm	1.57 ± 0.14	**10.65 ± 0.92***	2.90 ± 0.23	24.90 ± 1.03	4.64 ± 0.50	14.48 ± 0.93
Relapse						
No	1.60 ± 0.16	11.71 ± 0.95	3.02 ± 0.23	25.57 ± 1.05	4.21 ± 0.47	15.10 ± 0.98
Yes	1.55 ± 0.20	13.19 ± 1.75	2.50 ± 0.29	23.38 ± 1.47	4.40 ± 0.59	14.24 ± 1.19
Adjuvant treatment						
Chemotherapy						
No	1.56 ± 0.15	12.56 ± 1.10	2.69 ± 0.24	25.58 ± 1.06	4.75 ± 0.50	**15.68 ± 0.89**
Yes	1.51 ± 0.17	11.03 ± 1.07	2.94 ± 0.24	24.48 ± 1.16	3.57 ± 0.41&	**12.71 ± 1.04***
Radiotherapy						
No	1.52 ± 0.14	12.00 ± 0.93	2.95 ± 0.20	24.89 ± 0.93	4.40 ± 0.41	14.78 ± 0.88
Yes	1.58 ± 0.19	11.70 ± 1.43	2.38 ± 0.28	25.72 ± 1.47	3.86 ± 0.57	13.50 ± 0.93

*P>0.05; **P>0.01; #N0 vs. N2, P=0.05; &P=0.07; ##Stage I vs. Stage II P=0.06
HS, heparan sulfate; CS, chondroitin sulfate; Col I, collagen type I; Col III, collagen type III, Col IV, collagen type IV; Col V, collagen type V; ADC, lung adenocarcinoma; SqCC: lung squamous cell carcinoma; LCC, large cell carcinoma.
Bolded values refer to a P-value with statistical significance (P<0.05).

3.8 Validation of study cohort by *in silico* data mining

3.8.1 mRNA and protein expression

To create a possibility of the comparison between our data, normal samples, and other results, we used the UALCAN to analyze the TCGA database and to obtain levels of mRNA expression of our markers of interest, except the GAGs, in ADC and SqCC. The database did not include data for LCC.

Compared to normal tissues, in ADC (Supplementary Figure 4) the mRNA expression level of E-cadherin, β-catenin, Col I, Col III, Col V, WNT1, and WNT5B showed significant upregulation

10.3389/fonc.2022.1042766

TABLE 4 Variables associated with overall survival in 120 non-small cell lung cancer patients.

Variables	Mean OS (months)	Univariate Analysis[b]			Multivariate Analysis[c]	
		HR (95% CI)	HR	P value	HR (95% CI)	P value
Age (years, median)						
≤ 65 (reference)	74					
> 65	82	0.96 (0.54-1.71)	-0.039	0.895		
Gender						
Male (reference)	59					
Female	107	0.39 (0.20-0.75)	-0.942	**0.004**	0.39(0.17-0.90)	**0.028**
Smoker Status[a]						
Non-Smoker (reference)	95					
Smoker/Former-smoker	84	1.38(0.67-2.89)	0.327	0.382		
Histological type						
Adenocarcinoma (reference)	83			0.245		
Squamous cell carcinoma	74	1.65(0.89-3.09)	0.498	0.120		
Large cell carcinoma	60	1.88(0.45-7.94)	0.633	0.389		
T stage†						
T1+T2 (reference)	97					
T3+T4	45	2.69(1.46-4.94)	0.989	**0.001**	2.14(1.03-4.74)	**0.041**
N stage†						
N0 (reference)	85			0.497		
N1	55	1.63(0.71-3.70)	0.486	0.247		
N2	75	0.992(0.45-2.16)	-0.008	0.983		
Pathological Stage†						
I+II (reference)	87					
III	66	1.35(0.71-2.58)	0.301	0.362		
Tumor Size						
≤ 3 cm (reference)	100					
> 3 cm	61	1.95(1.08-3.53)	0.669	**0.026**	3.65(1.58-8.40)	**0.002**
Relapse[a]						
Absent (reference)	99					
Present	48	2.98(1.60-5.53)	1.091	**0.001**	2.65(1.27-5.51)	**0.009**
Adjuvant therapy						
Chemotherapy [a]						
Yes (reference)	64					
No	91	0.62(0.35-1.11)	-0.469	0.110	4.61(1.74-12.20)	**0.002**
Radiotherapy [a]						
Yes (reference)	45					
No	91	0.40(0.21-0.76)	-0.906	**0.005**	0.22(0.08-0.59)	**0.003**
Protein expression (≤ mean vs > mean)						
Heparan Sulfate [a]						
≤ 1.55% (reference)	72					
> 1.55%	91	0.77(0.42-1.41)	-0.258	0.402	0.44(0.19-1.02)	**0.055**
Chondroitin Sulfate						
≤ 11.92% (reference)	85					
> 11.92%	74	1.02(0.57-1.82)	0.022	0.941		
Col I						
≤ 2.81% (reference)	78					
> 2.81%	78	1.12(0.62-2.01)	0.114	0.703		

(Continued)

TABLE 4 Continued

Variables	Mean OS (months)	Univariate Analysis[b]			Multivariate Analysis[c]	
		HR (95% CI)	HR	P value	HR (95% CI)	P value
Col III						
≤ 25.04% (reference)	81					
> 25.04%	77	0.92(0.52-1.64)	-0.080	0.786		
Col IV						
≤ 4.30% (reference)	86					
> 4.30%	67	1.30(0.70-2.42)	0.264	0.405		
Col V						
≤ 14.42% (reference)	72					
> 14.42%	87	0.83(0.46-1.49)	-0.184	0.535		
E-cadherin						
≤ 23.99% (reference)	91					
> 23.99%	71	1.35(0.75-2.44)	0.300	0.320		
β-catenin [a]						
≤ 21.25% (reference)	87					
> 21.25%	77	1.08(0.60-1.93)	0.076	0.799		
WNT1 [a]						
≤ 13.55% (reference)	85					
> 13.55%	73	1.11(0.62-2.00)	0.108	0.715		
WNT3A [a]						
≤ 19.17% (reference)	61					
> 19.17%	97	0.54(0.30-0.99)	-0.612	**0.046**	1.44(0.69-3.01)	0.334
WNT5A						
≤ 42.48% (reference)	88					
> 42.48%	68	1.36(0.73-2.51)	0.306	0.329	0.768(0.35-1.66)	0.501
WNT5B						
≤ 3.75% (reference)	80					
> 3.75%	77	0.97(0.54-1.73)	-0.034	0.908		
SPARC						
≤ 14.25% (reference)	65					
> 14.25%	98	0.55(0.30-1.00)	-0.601	**0.050**	0.31(0.13-0.75)	**0.009**

[a]Some cases had missing follow-up information: Smoke status (31); Tumor size (2); Relapse (21); Chemotherapy (4); Radiotherapy (4); Status (21); Heparan Sulfate (1); β-catenin (2); WNT1 (2); WNT3A (2).
[b]Univariate analysis was carried out without any adjustment in order to generate hazard ratios with confidence intervals for individual risk for each of the parameters on survival.
[c]Multivariate analysis was carried out to analyze the effects of several risk parameters on survival.
† According to the 8th Edition International Association for the Study of Lung Cancer (21).
OS, Overall Survival; HR, Hazard Ratio (β coefficient); CI, Confidence Interval; Col I, collagen type I; Col III, collagen type III, Col IV, collagen type IV; Col V, collagen type V.
Bolded values refer to a P-value with statistical significance (P<0.05).
The univariate and multivariate analysis employed a Cox proportional hazards model (Chi-square 33.223; P<0.001).

upregulated (P<0.01, for all), whereas WNT3A mRNA expression levels showed significant downregulation (P<0.01).

For SqCC (Supplementary Figure 5), E-cadherin, Col I, Col III, Col V, WNT5A, and WNT5B mRNA expression levels showed significant upregulation (P<0.01, for all), and only WNT3A mRNA expression levels showed significant downregulation (P<0.01) when compared to the normal tissue.

We used the same platform to analyze protein expression data from the Clinical Proteomic Tumor Analysis Consortium (CPTAC). Until the time of this study, the consortium database only covered ADC samples but included data on most of the molecules explored in our study, except for WNT1, WNT3A, and WNT5B. Compared to normal tissues, E-cadherin, Col I, Col III, Col V, WNT5A, and SPARC showed significant overexpression (P<0.01, except SPARC, P=0.02; Supplementary Figure 6), whereas β-catenin, HS, CS, and Col IV showed significant under-expression (P<0.01; Supplementary Figure 6).

Consistent with our data, the results of E-cadherin, Col I, Col III, Col V, WNT5A, and SPARC proteins were overexpressed in ADC, whereas β-catenin, HS, CS, and Col IV were under-expressed.

3.8.2 Association between expression and clinicopathological parameters

Using the TCGA (Pan Cancer Atlas) data, we collected data from the cBio Cancer Genomics Portal which included clinical data of 939 patients with ADC and SqCC in pathological stages I to IIIA, like our cohort. We also collected the mRNA expression of all our markers of interest for this same group of individuals.

When correlating this data, we noted the association between gender and β-catenin, WNT1, WNT3A, WNT5A, and WNT5B expression; and a significant difference between histotypes and β-catenin, WNT1, WNT3A, WNT5A, and WNT5B expression. We also observed a significant difference between the expression of WNT1 and T2 stage and N1 stage. In patients who developed metastasis during follow-up, we observed a significant difference between the expression of β-catenin, WNT3A, and WNT5A. We did not find a statistical difference between other clinicopathologic characteristics and the expression of these markers (Supplementary Table 2).

Using the same analysis for the stoma markers, we observed a significant difference between the histotypes and the expression of Col III. We did not find significant differences between any other clinicopathologic characteristic and the expression of these markers (Supplementary Table 3).

3.8.3 Prognostic value GAGs, SPARC, EMT and collagen types

We used a Kaplan-Meier plotter analysis to find the correlation between expression levels and OS in NSCLC patients for all makers analyzed in our study (Supplementary Figure 7).

We set the cutoff for high or low expression using the group median expression. As shown in Supplementary Figure 7, patients with high Col IV, WNT3A, WNT5A, WNT5B, and SPARC expression had longer OS ($P<0.05$). On the other hand, low Col I and WNT1 expression correlated with longer OS ($P \leq 0.05$). We found no statistical significance effect of E-cadherin, β-catenin, HS, CS, Col III, and Col V on OS.

3.8.4 Biologic interaction among GAGs, SPARC, EMT and collagen types

Given the above, our next step was to conduct a functional enrichment analysis using the STRING database, a search tool for protein interaction, to find a significant protein-protein interaction (PPI) network. Figure 11 shows the molecular organization of this network. The network is made of differentially connected nodes, each node represents a protein, and the edges represent their dynamic interactions. The PPI enrichment P-value was <1.0e-16. This shows that these proteins are at least partially biologically connected as a group.

We also performed an analysis in Metascape to assess the function and the biological process of the genes corresponding to the proteins of our interest. Supplementary Figure 8A shows the heatmap of enriched terms across input gene lists. The main terms we observed were: "Epithelial to mesenchymal transition in colorectal cancer", "ECM proteoglycans", and "proteoglycans in cancer". Supplementary Figure 8B shows the network formed by these enriched terms. Supplementary Figure 8C shows the top-level Gene Ontology biological processes. The most statistically significant terms within this group were "development process", "signaling", "response to stimulus", and "cellular process".

Taken together, these *in silico* results suggest a strong integration between our proteins of interest and fundamental cellular processes in carcinogenesis, which confirms our experimental results.

4 Discussion

Under the scenario of our study cohort, the locoregional and distant metastases not previewed by TNM stage and histological classification are the possible reasons for surgical resection failures at curing some early-stage NSCLC patients. The query of interest is whether ancillary information gathered from either the tumor cells or its tumor stroma can help us to improve risk stratification and patient selection for adjuvant systemic treatment. The development of cancer cell invasion and metastases certainly encompasses a series of complex and sequential stages. Among them are the EMT, loss of basement membranes, and remodeling of the interstitial extracellular matrix barriers by tumor cells. These processes are considered important because the tumor-reprogrammed lung microenvironment promotes both primary lung tumors and metastasis by contributing mainly to mechanical and functional barriers (4). The loss of these barriers facilitates the migration of tumor cells and penetration of tumor by blood vessels (34–37). Other important glycoproteins present in lung cancer are the SPARC and WNTs. These glycoproteins act on the remodeling of the extracellular matrix and the EMT and provide tumor growth and metastasis (20, 38–40).

Therefore, to understand the relationship between EMT, matricellular barriers, and the metastatic process, we used a step-stage design. We first used IHC, TEM, and IF to characterize EMT proteins, glycosaminoglycans, collagen types, SPARC, and WNT proteins in ADC, SqCC, and LCC histological subtypes. We then examined the clinical association between these markers and the data of 120 patients with surgically excised NSCLC. Afterwards, we analyzed the impact of these markers on

patients' survival. Lastly, we validated the study cohort using *in silico* data mining. While one of the major limitations of our study is the small number of NSCLC cases used, the data obtained using IHC, IF and TEM and the image analysis applied minimized this limitation. Thus, we provide new evidence that NSCLC cells can express EMT and matricellular proteins with known mechanical barrier function. Expression of those proteins is associated both with the Wnt pathway and with significantly longer overall patient survival. We also found that there is strong integration between our proteins of interest, their expression/behavior is like what we obtained in this work, and they act on fundamental cellular processes in carcinogenesis.

While most of the studies evaluated NSCLC progression by either loss of tumor suppressor genes and/or activation of oncogenes (41), we described NSCLC progression during the EMT phenotypic changes in rendering tumor cells invasive and able to metastasize distant organs. Another main innovative feature of our study was the description of the relationship between EMT-ECM components-Wnt signal pathway with histological subtypes, TNM stage, and survival. Throughout our study, we showed that, during the EMT process, the low expression of E-cadherin and B-catenin created a poor tumor portion barrier against tumor invasion. Wnt signaling, mainly by WNT5A, and SPARC enhanced this barrier and facilitated tumor progression.

We also found that tumor-associated GAGs and collagen mechanical barriers reinforced the functional barriers between EMT, WNT proteins and malignant cells. The collagen mechanical barriers correspond to different levels of HS, CS, and collagen fibers, which are reorganized to locate and characterize malignant cells. In this scenario, we inferred that CS, Col III, and Col V also have a high chance to create a mechanical barrier against malignant cells and prevent the invasion of the interstitial ECM. Importantly, the high expression of WNT1, CS, and Col V was associated with tumors in stages I, and N0-N1. These findings suggest that increased expression of these markers occurs at early stages of carcinogenesis, which decreases with tumor growth. Therefore, these markers emerge as promising for therapeutic decisions before surgery. In addition, we also observed the influence, direct and indirect, of WNT3A, WNT5A, HS, and SPARC on the overall survival of patients with early-stage NSCLC. This influence shows that there is a relevant regulation between these components in tumor progression. However, there are some major points which need to be addressed, as discussed below.

Firstly, the issue to be addressed is the significance of EMT-like phenotypic changes for the interaction between collagen matrix and malignant cells. We observed that the scattering intensity was higher in EMT-positive malignant cells from SqCC and ADC compared to LCC, to promote ECM invasion by individual malignant cells, in case of LCC, and by cells groups in case of ADC and SqCC. Thus, these findings describe a partial EMT – e.g., a hybrid EMT (42, 43) – as the predominant, hierarchical immune phenotype in SqCC and ADC of the lung compared to LCC. This suggests the emerging notion that a partial EMT, but not necessarily a complete EMT, is associated with aggressive tumor progression (44). As recently reported, there is no mesenchymal transition in a hybrid EMT, as the tumor cells retain E-cadherin (45). This could explain the high expression of E-cadherin found in our work, mostly in SqCC and ADC. We also observed a strong correlation between TEM proteins and WNT5A. WNT5A can interact with the tumor, functioning as both a suppressor and a promoter (46, 47). When acting on cell adhesion, motility, and cell polarity, WNT5A interacts with intracellular effectors through the Wnt signaling pathway and with ECM structures. Thus, it acts in different ways on the process of EMT and on the B-catenin/E-cadherin complex (48, 49).

Moreover, these EMT modifications resulted from the dissemination of cancer and invasion of ECM, leading to changes in GAGs and fibril structural organization of collagens visualized at IF and TEM These changes then caused the endogenous GAGs/collagen to degenerate and the emergence of new GAGs/collagen of a diminished structural organization, as previously described in breast cancer (50). In addition, at TEM, we demonstrated appreciably larger cell-cell boundaries, suggesting abnormal adherent junctions because of cancer propagation into the interstitial ECM. As mentioned, the tissue loses cell-cell adhesion, causing diminished, long-range intermolecular bonding rigidity to Col I, Col III and Col V, with the disruption of collagen fibrils structure as well as HS/CS. Previous studies described that some cancer cells can produce collagen types I, III, IV, V and VI (51–53). The alteration in the deposition of these fibrillar collagens alters the rigidity of the tumor matrix and may contribute to the metastatic behavior of tumor cells, mediated, for example, by cancer-associated fibroblasts (54). Besides, tissue remodeling is a crucial step during carcinogenesis, a transformation of epithelial cells being associated with metalloproteinases in collagen degradation by synthesis of fibrillary and non-fibrillary matrix proteins (55).

In our cohort, we noted a high expression of Col III and V, and a low expression of Col I and IV. We observed the same increased expression of Col III in pancreatic cancer tissue (56). Concerning the Col I expression, it has been proposed that the interaction of epithelial cells with Col I contributes to increased cell motility accompanying EMT, critical in disease progression (57). In colorectal cancer, Col I was described to downregulate the E-cadherin and B-catenin expression (58). However, we did not observe a correlation between Col I and ETM proteins. The expression of these markers in our study was inverse, suggesting a possible response to changes in ECM. In addition, the reduction of the structural assembly, and consequent decrease in expression of HS, and Col I and IV within ECM, can be associated with the activity of collagenases, a group of collagen-degrading enzymes (59). Furthermore, not only there was a

change in the concentration of fibrillar collagens, but also in the levels of immunoreactive GAGs, such as HS and CS, and their proteoglycans. This change can alter the stiffness of the ECM and participate in the EMT process (60, 61). Although these compounds do not alter the biomechanical characteristics of ECM, they can play an important role in tumor ECM remodeling (60).

In the current study, it also became evident that, at the supramolecular level changes in collagen in cancer-invaded ECM, there was an association, direct or indirect, with WNT1, WNT3A, and WNT5A signaling. This paved the way for NSCLC classification and improved understanding of mechanisms of cancer growth. These findings can be confirmed by recent studies that have shown the Wnt/β-catenin signaling pathway integrates signals from other proteins and signaling pathways, such as the possibility of the pathway being modulated by integrins (62). These studies suggest that this modulation process can be done through the expression of WNT ligands, receptors, and inhibitors, or through the modulation of β-catenin concentration in different cell types (63).

We also highlight that we found an important association between HS with SPARC and WNT3A. In tumor stroma, the composition of the ECM and the population of cells present there are quite different when compared to normal tissue. As the tumor develops, a series of processes also occur that remodel the stromal tissue to regulate tumor progression. In general, several glycoproteins, collagens, GAGs, proteoglycans, and many other proteins that promote cell proliferation and motility drive this modification process in the ECM (64–66). Among the GAGs, HS, when cleaved by heparanase, alters its structure and function, and contributes to tumor-mediated remodeling of both cell surfaces and the ECM (67–69). It is also known that HS proteoglycans extracellularly regulate WNT signaling, including WNT3A (70). Thus, these activities increase the bioavailability of HS-linked growth factors (71) that recruit metastatic malignant cells, and support their survival and growth, thus driving the metastatic process. Previous studies report that HS plays a crucial role in cell proliferation and metastasis in breast cancer (72), rhabdomyosarcoma (73), and NSCLC (74). Furthermore, changes in the microenvironment also affect the expression and function of other molecules, such as SPARC. This is a matricellular glycoprotein that directly participates in the ECM remodeling process, regulating processes, such as metalloproteinase secretion and cell-matrix interactions (75). Recent studies showed that SPARC favored the migration and invasion of endometrial carcinoma cells *in vitro* and *in vivo* (76).

We also found that CS was correlated with WNT1. CS is a transmembrane glycoprotein with a large extracellular domain and a short intracellular domain (77). Its extracellular domain includes subdomains that can interact with various components of the ECM, such as Col V and Col VI. Thus, they promote the activation of oncogenic pathways, growth factors, and EMT, increasing the migration of malignant cells (78–80). As it is widely accepted, GAGs can shape morphogenesis gradients and modulate morphogenesis signaling through their binding affinities with a variety of signaling molecules due to their various structures (81). Furthermore, as we could observe in this study, an increase in CS expression in tissues has been described in several tumor types (82–85). Thus, it can be suggested that this increase could influence ECM remodeling (as previously mentioned for HS), altering the tumor microenvironment and modifying cell-signaling processes. Even though we failed to locate any other studies that demonstrate this direct affinity between CS and WNT1, our finding may be supported by the study that showed that CS-E (a CS with an increased level of 4,6-O-disulfated disaccharides) inhibited the Wnt signaling pathway *in vitro* assays using breast cancer cells (86).

Another important issue to address is the impact of the morphometric variables on clinicopathologic features. Histologically, we observed that NSCLC encompassed a bimodal spectrum of malignancies. On one side, there are ADC and SqCC subtypes, both characterized by EMT process, associated with considerable desmoplasia due to accumulation of ECM components, which were closely associated with basement membrane invasion by groups of cells at TEM. On the opposite side, there is LCC, a NSCLC subtype composed by malignant cells immersed in a poor desmoplastic stroma with low EMT process, and low levels of basement membrane and interstitial ECM components, associated with invasion by isolated cells at TEM. Clinically, we found that younger patients presented more small-sized aggressive tumors, LCC-histotype with low expression of E-cadherin and Col IV, and high expression of WNT1 and WNT5B. Moreover, large tumors (bigger than 3.0 cm) showed low expression of CS and higher expression of WNT5A. Notably, the high expression of WNT1, CS and Col V was associated with tumors in stage I and II, and N0-N1, suggesting that increased expression of these markers occurs at very earlier stages of carcinogenesis. Considering the findings discussed so far, it is possible to see that, sometimes, the molecules studied in the microenvironment here presented may not act directly on the clinical characteristics, affecting the aggressiveness of the tumor. However, the observed behavior is very consistent with the remodeling of this environment for progression and metastasis. This makes us speculate that the study of these compounds is important for the analysis of early-stage patients regarding progression and the intention to prevent operated patients from having relapses due to possible occult metastasis.

We observed a consistent result in our Cox regression analysis. We performed this analysis to examine the impact of morphometric variables on survival. We observed that OS, for the entire cohort, was significantly influenced by gender, T stage, tumor size, metastases, and radiotherapy. We also observed a significant influence for better OS by higher SPARC and WNT3A

expression in tumor stroma. In multivariable analysis, there was a significant association between gender, T stage, tumor size, metastasis, adjuvant therapy, and SPARC and OS. High expression of HS presented borderline influence for OS. WNT3A and WNT5A were co-variables in this mathematical model. Mean OS was 97 months for patients with SPARC expression >14.25% compared to 65 months for patients with expression ≤14.25%. These findings are consistent with the literature, in which there is an association between low SPARC expression and a worse prognosis in endometrial carcinoma (76), colorectal cancer (87, 88), and NSCLC (89). Furthermore, it is described that, depending on the tumor microenvironment, SPARC can act both as a tumor suppressor and as an oncogene (76). Thus, we speculate that the HS and WNT proteins, which influenced our regression model, may affect these questions about the paradoxical effect of SPARC. Since HS proteoglycans modulate WNT signaling, HS modifications influence disease progression (90–92). In addition, we have the entire ECM modification process discussed previously.

To complete these results of the roles of E-cadherin, β-catenin, Col I, Col III, Col V, WNT1, and WNT5B in NSCLC, we explored, *in silico*, whether the mRNA level of these proteins related to clinical outcome of the patients. We confirmed that in ADC and SqCC there was a significant upregulation of the mRNA expression levels of E-cadherin, β-catenin, Col I, Col III, Col V, WNT1, and WNT5B, while there was a significant downregulation of Wnt3A mRNA expression. By Clinical Proteomic Tumor Analysis Consortium (CPTAC) data for protein expression only available for ADC, we confirmed that there was a significant overexpression of E-cadherin, Col I, Col III, Col V, WNT5A, and SPARC protein, while there was a significant under-expression of β-catenin, HS, CS, and Col IV. The behavior of this protein expression was similar in our study, proving the consistency of our results. As for clinicopathological characteristics, we observed some correlations between β-catenin, WNT proteins, and histological subtypes (ADC and SqCC), which we observed only with β-catenin and WNT5A in our study. However, except for WNT1 expression and T and N stages, the data also showed no significant correlations with pathological stage or TNM as in our study. Kaplan-Meier curves showed that patients with higher Col IV, WNT3A, WNT5A, WNT5B, and SPARC expression had a longer OS, whereas patients with higher Col I and WNT1 expression had a shorter OS. The function and the biological process of the genes corresponding to the proteins of our interest showed that the top-level Gene Ontology biological processes involved were “development process”, “signaling”, “response to stimulus”, and “cellular process”.

In summary, the data presented provide important hierarchical evidence that genes and proteins associated with EMT, WNT signal pathway, and ECM are involved in the proliferative signal of cancer cells, spaced desmosomes, and facilitating cell motility. This evidence suggests sequential steps for primary tumor invasion and metastasis in patients that were in early-stages and who underwent surgical resection. Importantly, this study indicates that NSCLC with increased expression of mechanical barrier proteins and low expression of the functional proliferative barrier presents a low risk of patient mortality due to metastasis and promising new therapeutic targets. In addition, mechanistic insight into the major findings needs to be complemented with *in vitro* data. Therefore, in this emerging scenario of personalized treatments, future studies are necessary to include such observations in the clinic, as the basis of a biomarker measured in circulation and/or urine for selecting patients who may benefit from these.

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: EMPIAR, accession number EMPIAR-11226.

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee of the Faculty of Medicine, University of São Paulo (protocol number 150.443/2019). Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

Conception and design: VC and CB. Writing, review, and editing: VC, CB, and TP. Data analysis and interpretation: VC, CB, TP. Investigation: CB, TP, JM-R, JM, LO, AV, and WT. Provision of study materials or patients: AA and TT. Supervision: VC. Administrative support: VC. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by Sao Paulo Research Foundation (FAPESP; 2018/20403-6), the National Council for Scientific and Technological Development (CNPq 303735/2021-0), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES; Finance Code 001).

Acknowledgments

We are grateful to Ms. Esmeralda Miristeni Eher and Ms. Sandra de Morais Fernezlian for their expertise in immunohistochemical protocols, and to Aline Assato for her expertise to construct of the TMAs slides.

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.2022.1042766/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Negative controls of immunofluorescence for E-cadherin and β-catenin in the different histological subtypes of NSCLC. The stained nuclei are represented in blue (DAPI). Original magnification: 40X. LCC, large cell carcinoma; ADC, lung adenocarcinoma; SqCC: lung squamous cell carcinoma.

SUPPLEMENTARY FIGURE 2
Negative controls of immunofluorescence for chondroitin sulfate and heparan sulfate (green), and collagen types I, III, IV, and V (red). The stained nuclei are represented in blue (DAPI). Original magnification: 40X.

SUPPLEMENTARY FIGURE 3
Survival curve for SPARC expression. The blue line represents the high expression of SPARC, and the yellow line the low expression of SPARC.

SUPPLEMENTARY FIGURE 4
mRNA expression of the interest markers in lung adenocarcinoma from data obtained from "The Cancer Genome Atlas" (TCGA). **P<0.01. ADC, lung adenocarcinoma; Col, collagen type.

SUPPLEMENTARY FIGURE 5
mRNA expression of the interest markers in lung squamous cell carcinoma from data obtained from "The Cancer Genome Atlas" (TCGA). **P<0.01. SqCC: lung squamous cell carcinoma; Col, collagen type.

SUPPLEMENTARY FIGURE 6
Expression of the interest proteins in adenocarcinoma from data obtained from the Clinical Proteomic Tumor Analysis Consortium (CPTAC). *P<0.05, **P<0.01. ADC, lung adenocarcinoma; HS, heparan sulfate; CS, chondroitin sulfate; Col, collagen type.

SUPPLEMENTARY FIGURE 7
Kaplan Meier curves for our interest markers obtained in KM-Plotter data base. The black line represents a low expression of the marker and the red line, the high expression.

SUPPLEMENTARY FIGURE 8
Metascape enrichment analysis. **(A)** Heatmap of enriched terms across input gene lists. **(B)** Network formed by these enriched terms. **(C)** Biological processes were involved the gene list colored according to the P-value.

References

1. Dela Cruz CS, Tanoue LT, Matthay RA. Lung cancer: epidemiology, etiology, and prevention. *Clinics chest Med* (2011) 32(4):605–44. doi: 10.1016/j.ccm.2011.09.001

2. Duma N, Santana-Davila R, Molina JR. Non-small cell lung cancer: Epidemiology, screening, diagnosis, and treatment. *Mayo Clinic Proc* (2019) 94 (8):1623–40. doi: 10.1016/j.mayocp.2019.01.013

3. Araujo LH, Baldotto C, Castro G. C.OMMAJ.R.X.X.X, Katz A, Ferreira CG, Mathias C, et al. Lung cancer in Brazil. *Jornal brasileiro pneumologia: publicacao oficial da Sociedade Bras Pneumologia e Tisilogia* (2018) 44(1):55–64. doi: 10.1590/S1806-37562017000000135

4. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. *Nat Med* (2013) 19(11):1423–37. doi: 10.1038/nm.3394

5. Coello MC, Luketich JD, Litle VR, Godfrey TE. Prognostic significance of micrometastasis in non-small-cell lung cancer. *Clin Lung Cancer* (2004) 5(4):214–25. doi: 10.3816/CLC.2004.n.002

6. Grivennikov SI, Greten FR, Karin M. Immunity, inflammation, and cancer. *Cell* (2010) 140(6):883–99. doi: 10.1016/j.cell.2010.01.025

7. Pudełko A, Wisowski G, Olczyk K, Koźma EM. The dual role of the glycosaminoglycan chondroitin-6-sulfate in the development, progression and metastasis of cancer. *FEBS J* (2019) 286(10):1815–37. doi: 10.1111/febs.14748

8. Dunnill MS. *Pulmonary pathology*. New York: Churchill Livingstone (1982) p. 293–334.

9. Askin FB, Kaufman DG. Histomorphology of human lung cancer. in carcinogenesis, a comprehensive survey. In: Mass MJ, Kaufman DG, Siegfried JM, Steele VE, Nesnow S, editors. *Cancer of the respiratory tract: Predisposing factors*, vol. 8 . New York: Raven Press (1985). p. 17–21.

10. Wang Y, Zhou BP. Epithelial-mesenchymal transition–a hallmark of breast cancer metastasis. *Cancer hallmarks* (2013) 1(1):38–49. doi: 10.1166/ch.2013.1004

11. Kuo DS, Labelle-Dumais C, Gould DB. COL4A1 and COL4A2 mutations and disease: insights into pathogenic mechanisms and potential therapeutic targets. *Hum Mol Genet* (2012) 21(R1):R97–R110. doi: 10.1093/hmg/dds346

12. Yang J, Antin P, Berx G, Blanpain C, Brabletz T, Bronner M, et al. Guidelines and definitions for research on epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol* (2020) 21(6):341–52. doi: 10.1038/s41580-020-0237-9

13. Dongre A, Weinberg RA. New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. *Nat Rev Mol Cell Biol* (2019) 20(2):69–84. doi: 10.1038/s41580-018-0080-4

14. Hynes RO. The extracellular matrix: not just pretty fibrils. *Sci (New York N.Y.)* (2009) 326(5957):1216–9. doi: 10.1126/science.1176009

15. Fabrizio FP, Sparaneo A, Fontana A, Mazza T, Graziano P, Pantalone A, et al. Potential prognostic role of SPARC methylation in non-Small-Cell lung cancer. *Cells* (2020) 9(6):1523. doi: 10.3390/cells9061523

16. Mizumoto S, Watanabe M, Yamada S, Sugahara K. Expression of n-acetylgalactosamine 4-sulfate 6-o-sulfotransferase involved in chondroitin sulfate synthesis is responsible for pulmonary metastasis. *BioMed Res Int* (2013) 2013:656319. doi: 10.1155/2013/656319

17. Bourgot I, Primac I, Louis T, Noël A, Maquoi E. Reciprocal interplay between fibrillar collagens and collagen-binding integrins: Implications in cancer progression and metastasis. *Front Oncol* (2020) 10:1488. doi: 10.3389/fonc.2020.01488

18. Miao L, Zhu S, Wang Y, Li Y, Ding J, Dai J, et al. Discoidin domain receptor 1 is associated with poor prognosis of non-small cell lung cancer and promotes cell invasion *via* epithelial-to-mesenchymal transition. *Med Oncol (Northwood London England)* (2013) 30(3):626. doi: 10.1007/s12032-013-0626-4

19. Patel S, Alam A, Pant R, Chattopadhyay S. Wnt signaling and its significance within the tumor microenvironment: Novel therapeutic insights. *Front Immunol* (2019) 10:2872. doi: 10.3389/fimmu.2019.02872

20. Chen Y, Chen Z, Tang Y, Xiao Q. The involvement of noncanonical wnt signaling in cancers. *Biomed pharmacother = Biomed pharmacotherapie* (2021) 133:110946. doi: 10.1016/j.biopha.2020.110946

21. Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WE, et al. The IASLC lung cancer staging project: Proposals for revision of the TNM stage groupings in the forthcoming (Eighth) edition of the TNM classification for lung cancer. *J Thorac oncology: Off Publ Int Assoc Study Lung Cancer* (2016) 11(1):39–51. doi: 10.1016/j.jtho.2015.09.009

22. Travis WD, Brambilla E, Burke AP, Marx A, Nicholson AG. Introduction to the 2015 world health organization classification of tumors of the lung, pleura, thymus, and heart. *J Thorac oncology: Off Publ Int Assoc Study Lung Cancer* (2015) 10(9):1240–2. doi: 10.1097/JTO.0000000000000663

23. Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, et al. QuPath: Open source software for digital pathology image analysis. *Sci Rep* (2017) 7(1):16878. doi: 10.1038/s41598-017-17204-5

24. Chandrashekar DS, Karthikeyan SK, Korla PK, Patel H, Shovon AR, Athar M, et al. UALCAN: An update to the integrated cancer data analysis platform. *Neoplasia (New York N.Y.)* (2022) 25:18–27. doi: 10.1016/j.neo.2022.01.001

25. Chandrashekar DS, Bashel B, Balasubramanya S, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, et al. UALCAN: A portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia (New York N.Y.)* (2017) 19(8):649–58. doi: 10.1016/j.neo.2017.05.002

26. Zhang Y, Chen F, Chandrashekar DS, Varambally S, Creighton CJ. Proteogenomic characterization of 2002 human cancers reveals pan-cancer molecular subtypes and associated pathways. *Nat Commun* (2022) 13(1):2669. doi: 10.1038/s41467-022-30342-3

27. Chen F, Chandrashekar DS, Varambally S, Creighton CJ. Pan-cancer molecular subtypes revealed by mass-spectrometry-based proteomic characterization of more than 500 human cancers. *Nat Commun* (2019) 10(1):5679. doi: 10.1038/s41467-019-13528-0

28. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discovery* (2012) 2(5):401–4. doi: 10.1158/2159-8290.CD-12-0095

29. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signaling* (2013) 6(269):l1. doi: 10.1126/scisignal.2004088

30. Lánczky A, Győrffy B. Web-based survival analysis tool tailored for medical research (KMplot): Development and implementation. *J Med Internet Res* (2021) 23(7):e27633. doi: 10.2196/27633

31. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* (2021) 49(D1):D605–12. doi: 10.1093/nar/gkaa1074

32. Snel B, Lehmann G, Bork P, Huynen MA. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res* (2000) 28(18):3442–4. doi: 10.1093/nar/28.18.3442

33. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* (2019) 10(1):1523. doi: 10.1038/s41467-019-09234-6

34. Huang Z, Zhang Z, Zhou C, Liu L, Huang C. Epithelial-mesenchymal transition: The history, regulatory mechanism, and cancer therapeutic opportunities. *MedComm* (2022) 3(2):e144. doi: 10.1002/mco2.144

35. Miller RT. Mechanical properties of basement membrane in health and disease. *Matrix biology: J Int Soc Matrix Biol* (2017) 57-58:366–73. doi: 10.1016/j.matbio.2016.07.001

36. Mouw JK, Ou G, Weaver VM. Extracellular matrix assembly: a multiscale deconstruction. *Nat Rev Mol Cell Biol* (2014) 15(12):771–85. doi: 10.1038/nrm3902

37. Cox TR. The matrix in cancer. *Nat Rev Cancer* (2021) 21(4):217–38. doi: 10.1038/s41568-020-00329-7

38. Koukourakis MI, Giatromanolaki A, Brekken RA, Sivridis E, Gatter KC, Harris AL, et al. Enhanced expression of SPARC/osteonectin in the tumor-associated stroma of non-small cell lung cancer is correlated with markers of hypoxia/acidity and with poor prognosis of patients. *Cancer Res* (2003) 63(17):5376–80.

39. Suzuki M, Hao C, Takahashi T, Shigematsu H, Shivapurkar N, Sathyanarayana UG, et al. Aberrant methylation of SPARC in human lung cancers. *Br J Cancer* (2005) 92(5):942–8. doi: 10.1038/sj.bjc.6602376

40. Andriani F, Landoni E, Mensah M, Facchinetti F, Miceli R, Tagliabue E, et al. Diagnostic role of circulating extracellular matrix-related proteins in non-small cell lung cancer. *BMC Cancer* (2018) 18(1):899. doi: 10.1186/s12885-018-4772-0

41. Cooper WA, Lam DC, O'Toole SA, Minna JD. Molecular biology of lung cancer. *J Thorac Dis* (2013) 5 Suppl 5(Suppl 5):S479–90. doi: 10.3978/j.issn.2072-1439.2013.08.03

42. Jolly MK, Boareto M, Huang B, Jia D, Lu M, Ben-Jacob E, et al. Implications of the hybrid Epithelial/Mesenchymal phenotype in metastasis. *Front Oncol* (2015) 5:155. doi: 10.3389/fonc.2015.00155

43. Huang B, Jolly MK, Lu M, Tsarfaty I, Ben-Jacob E, Onuchic JN. Modeling the transitions between collective and solitary migration phenotypes in cancer metastasis. *Sci Rep* (2015) 5:17379. doi: 10.1038/srep17379

44. Jolly MK, Tripathi SC, Jia D, Mooney SM, Celiktas M, Hanash SM, et al. Stability of the hybrid epithelial/mesenchymal phenotype. *Oncotarget* (2016) 7(19):27067–84. doi: 10.18632/oncotarget.8166

45. Zacharias M, Brcic L, Eidenhammer S, Popper H. Bulk tumour cell migration in lung carcinomas might be more common than epithelial-mesenchymal transition and be differently regulated. *BMC Cancer* (2018) 18(1):717. doi: 10.1186/s12885-018-4640-y

46. Kikuchi A, Yamamoto H, Sato A, Matsumoto S. Wnt5a: its signalling, functions and implication in diseases. *Acta physiologica (Oxford England)* (2012) 204(1):17–33. doi: 10.1111/j.1748-1716.2011.02294.x

47. Endo M, Nishita M, Fujii M, Minami Y. Insight into the role of Wnt5a-induced signaling in normal and cancer cells. *Int Rev Cell Mol Biol* (2015) 314:117–48. doi: 10.1016/bs.ircmb.2014.10.003

48. Matsumoto S, Fumoto K, Okamoto T, Kaibuchi K, Kikuchi A. Binding of APC and dishevelled mediates Wnt5a-regulated focal adhesion dynamics in migrating cells. *EMBO J* (2010) 29(7):1192–204. doi: 10.1038/emboj.2010.26

49. Medrek C, Landberg G, Andersson T, Leandersson K. Wnt-5a-CKI{alpha} signaling promotes {beta}-catenin/E-cadherin complex formation and intercellular adhesion in human breast epithelial cells. *J Biol Chem* (2009) 284(16):10968–79. doi: 10.1074/jbc.M804923200

50. Pucci Minafra I, Minafra S, Tomasino RM, Sciarrino S, Tinervia R. Collagen changes in the ductal infiltrating (scirrhous) carcinoma of the human breast. a possible role played by type I trimer collagen on the invasive growth. *J submicroscopic cytology* (1986) 18(4):795–805.

51. Naba A, Clauser KR, Lamar JM, Carr SA, Hynes RO. Extracellular matrix signatures of human mammary carcinoma identify novel metastasis promoters. *eLife* (2014) 3:e01308. doi: 10.7554/eLife.01308

52. Ohlund D, Lundin C, Ardnor B, Oman M, Naredi P, Sund M. Type IV collagen is a tumour stroma-derived biomarker for pancreas cancer. *Br J Cancer* (2009) 101(1):91–7. doi: 10.1038/sj.bjc.6605107

53. Qiu S, Deng L, Liao X, Nie L, Qi F, Jin K, et al. Tumor-associated macrophages promote bladder tumor growth through PI3K/AKT signal induced by collagen. *Cancer Sci* (2019) 110(7):2110–8. doi: 10.1111/cas.14078

54. Kwa MQ, Herum KM, Brakebusch C. Cancer-associated fibroblasts: how do they contribute to metastasis? *Clin Exp metastasis* (2019) 36(2):71–86. doi: 10.1007/s10585-019-09959-0

55. Coussens LM, Werb Z. Matrix metal loproteinases and the development of cancer. *Chem Biol* (1996) 3(11):895–904. doi: 10.1016/S1074-5521(96)90178-7

56. Bachem MG, Schneider E, Gross H, Weidenbach H, Schmid RM, Menke A, et al. Identification, culture, and characterization of pancreatic stellate cells in rats and humans. *Gastroenterology* (1998) 115(2):421–32. doi: 10.1016/s0016-5085(98)70209-4

57. Liotta LA, Kohn EC. The microenvironment of the tumour-host interface. *Nature* (2001) 411(6835):375–9. doi: 10.1038/35077241

58. Kirkland SC. Type I collagen inhibits differentiation and promotes a stem cell-like phenotype in human colorectal carcinoma cells. *Br J Cancer* (2009) 101(2):320–6. doi: 10.1038/sj.bjc.6605143

59. Round AR, Wilkinson SJ, Hall CJ, Rogers KD, Glatter O, Wess T, et al. A preliminary study of breast cancer diagnosis using laboratory based small angle x-

ray scattering. *Phys Med Biol* (2005) 50(17):4159–68. doi: 10.1088/0031-9155/50/17/017

60. Brauchle E, Kasper J, Daum R, Schierbaum N, Falch C, Kirschniak A, et al. Biomechanical and biomolecular characterization of extracellular matrix structures in human colon carcinomas. *Matrix biology: J Int Soc Matrix Biol* (2018) 68-69:180–93. doi: 10.1016/j.matbio.2018.03.016

61. Yang J, Liao Q, Price M, Moriarity B, Wolf N, Felices M, et al. Chondroitin sulfate proteoglycan 4, a targetable oncoantigen that promotes ovarian cancer growth, invasion, cisplatin resistance and spheroid formation. *Trans Oncol* (2022) 16:101318. doi: 10.1016/j.tranon.2021.101318

62. Astudillo P, Larraín J. Wnt signaling and cell-matrix adhesion. *Curr Mol Med* (2014) 14(2):209–20. doi: 10.2174/1566524014666140128105352

63. Astudillo P. Extracellular matrix stiffness and wnt/β-catenin signaling in physiology and disease. *Biochem Soc Trans* (2020) 48(3):1187–98. doi: 10.1042/BST20200026

64. Werb Z, Lu P. The role of stroma in tumor development. *Cancer J (Sudbury Mass.)* (2015) 21(4):250–3. doi: 10.1097/PPO.0000000000000127

65. Theocharis AD, Skandalis SS, Gialeli C, Karamanos NK. Extracellular matrix structure. *Advanced Drug delivery Rev* (2016) 97:4–27. doi: 10.1016/j.addr.2015.11.001

66. Gialeli C, Theocharis AD, Karamanos NK. Roles of matrix metalloproteinases in cancer progression and their pharmacological targeting. *FEBS J* (2011) 278(1):16–27. doi: 10.1111/j.1742-4658.2010.07919.x

67. Karamanos NK, Theocharis AD, Neill T, Iozzo RV. Matrix modeling and remodeling: A biological interplay regulating tissue homeostasis and diseases. *Matrix biology: J Int Soc Matrix Biol* (2019) 75-76:1–11. doi: 10.1016/j.matbio.2018.08.007

68. Sanderson RD, Bandari SK, Vlodavsky I. Proteases and glycosidases on the surface of exosomes: Newly discovered mechanisms for extracellular remodeling. *Matrix biology: J Int Soc Matrix Biol* (2019) 75-76:160–9. doi: 10.1016/j.matbio.2017.10.007

69. Theocharis AD, Karamanos NK. Proteoglycans remodeling in cancer: Underlying molecular mechanisms. *Matrix biology: J Int Soc Matrix Biol* (2019) 75-76:220–59. doi: 10.1016/j.matbio.2017.10.008

70. Mii Y, Takada S. Heparan sulfate proteoglycan clustering in wnt signaling and dispersal. *Front Cell Dev Biol* (2020) 8:631. doi: 10.3389/fcell.2020.00631

71. Hammond E, Khurana A, Shridhar V, Dredge K. The role of heparanase and sulfatases in the modification of heparan sulfate proteoglycans within the tumor microenvironment and opportunities for novel cancer therapeutics. *Front Oncol* (2014) 4:195. doi: 10.3389/fonc.2014.00195

72. Zhang C, Zhang S, Zhang D, Zhang Z, Xu Y, Liu S. A lung cancer gene GPC5 could also be crucial in breast cancer. *Mol Genet Metab* (2011) 103(1):104–5. doi: 10.1016/j.ymgme.2011.02.005

73. Li F, Shi W, Capurro M, Filmus J. Glypican-5 stimulates rhabdomyosarcoma cell proliferation by activating hedgehog signaling. *J Cell Biol* (2011) 192(4):691–704. doi: 10.1083/jcb.201008087

74. Yang X, Zhang Z, Qiu M, Hu J, Fan X, Wang J, et al. Glypican-5 is a novel metastasis suppressor gene in non-small cell lung cancer. *Cancer Lett* (2013) 341(2):265–73. doi: 10.1016/j.canlet.2013.08.020

75. Arnold SA, Brekken RA. SPARC: a matricellular regulator of tumorigenesis. *J Cell communication Signaling* (2009) 3(3-4):255–73. doi: 10.1007/s12079-009-0072-4

76. Wang L, Wang W, Xu Y, Wang Q. Low levels of SPARC are associated with tumor progression and poor prognosis in human endometrial carcinoma. *OncoTargets Ther* (2020) 13:11549–69. doi: 10.2147/OTT.S277795

77. Nishiyama A, Dahlin KJ, Prince JT, Johnstone SR, Stallcup WB. The primary structure of NG2, a novel membrane-spanning proteoglycan. *J Cell Biol* (1991) 114(2):359–71. doi: 10.1083/jcb.114.2.359

78. Burg MA, Nishiyama A, Stallcup WB. A central segment of the NG2 proteoglycan is critical for the ability of glioma cells to bind and migrate toward type VI collagen. *Exp Cell Res* (1997) 235(1):254–64. doi: 10.1006/excr.1997.3674

79. Wang X, Wang Y, Yu L, Sakakura K, Visus C, Schwab JH, et al. CSPG4 in cancer: multiple roles. *Curr Mol Med* (2010) 10(4):419–29. doi: 10.2174/156652410791316977

80. Yang J, Price MA, Wanshura L, He J, Yi M, Welch DR, et al. Chondroitin sulfate proteoglycan 4 enhanced melanoma motility and growth requires a cysteine in the core protein transmembrane domain. *Melanoma Res* (2019) 29(4):365–75. doi: 10.1097/CMR.0000000000000574

81. Yan D, Lin X. Shaping morphogen gradients by proteoglycans. *Cold Spring Harbor Perspect Biol* (2009) 1(3):a002493. doi: 10.1101/cshperspect.a002493

82. Brauchle E, Kasper J, Daum R, Schierbaum N, Falch C, Kirschniak A, et al. Biomechanical and biomolecular characterization of extracellular matrix structures in human colon carcinomas. *Matrix biology: J Int Soc Matrix Biol* (2018) 68-69:180–93. doi: 10.1016/j.matbio.2018.03.016

83. Rangel MP, de Sá VK, Prieto T, Martins J, Olivieri ER, Carraro D, et al. Biomolecular analysis of matrix proteoglycans as biomarkers in non small cell lung cancer. *Glycoconjugate J* (2018) 35(2):233–42. doi: 10.1007/s10719-018-9815-x

84. Marolla AP, Waisberg J, Saba GT, Waisberg DR, Margeotto FB, Pinhal MA. Glycomics expression analysis of sulfated glycosaminoglycans of human colorectal cancer tissues and non-neoplastic mucosa by electrospray ionization mass spectrometry. *Einstein (Sao Paulo Brazil)* (2015) 13(4):510–7. doi: 10.1590/S1679-45082015AO3477

85. Svensson KJ, Christianson HC, Kucharzewska P, Fagerström V, Lundstedt L, Borgquist S, et al. Chondroitin sulfate expression predicts poor outcome in breast cancer. *Int J Oncol* (2011) 39(6):1421–8. doi: 10.3892/ijo.2011.1164

86. Willis CM, Klüppel M. Chondroitin sulfate-e is a negative regulator of a pro-tumorigenic wnt/beta-catenin-Collagen 1 axis in breast cancer cells. *PLos One* (2014) 9(8):e103966. doi: 10.1371/journal.pone.0103966

87. Liu QZ, Gao XH, Chang WJ, Wang HT, Wang H, Cao GW, et al. Secreted protein acidic and rich in cysteine expression in human colorectal cancer predicts postoperative prognosis. *Eur Rev Med Pharmacol Sci* (2015) 19(10):1803–11.

88. Liang JF, Wang HK, Xiao H, Li N, Cheng CX, Zhao YZ, et al. Relationship and prognostic significance of SPARC and VEGF protein expression in colon cancer. *J Exp Clin Cancer research: CR* (2010) 29(1):71. doi: 10.1186/1756-9966-29-71

89. Huang Y, Zhang J, Zhao YY, Jiang W, Xue C, Xu F, et al. SPARC Expression and prognostic value in non-small cell lung cancer. *Chin J Cancer* (2012) 31(11):541–8. doi: 10.5732/cjc.012.10212

90. Capurro MI, Xiang YY, Lobe C, Filmus J. Glypican-3 promotes the growth of hepatocellular carcinoma by stimulating canonical wnt signaling. *Cancer Res* (2005) 65(14):6245–54. doi: 10.1158/0008-5472.CAN-04-4244

91. Lund ME, Campbell DH, Walsh BJ. The role of glypican-1 in the tumour microenvironment. *Adv Exp Med Biol* (2020) 1245:163–76. doi: 10.1007/978-3-030-40146-7_8

92. Zittermann SI, Capurro MI, Shi W, Filmus J. Soluble glypican 3 inhibits the growth of hepatocellular carcinoma *in vitro* and *in vivo*. *Int J Cancer* (2010) 126(6):1291–301. doi: 10.1002/ijc.24941

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 15 November 2022
DOI 10.3389/fonc.2022.1058436

Check for updates

OPEN ACCESS

EDITED BY
Ping Zhan,
Nanjing University School of Medicine, China

REVIEWED BY
Vera Luiza Capelozzi,
University of São Paulo, Brazil
Xiaocheng Charlie Dong,
Indiana University, United States
Fushun Zhang,
The University of Texas Health Science Center at San Antonio, United States

*CORRESPONDENCE
Mei Shi
clubstone2022@163.com

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

SPECIALTY SECTION
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

RECEIVED 30 September 2022
ACCEPTED 01 November 2022
PUBLISHED 15 November 2022

CITATION
Liu B, Li Y, Suo L, Zhang W, Cao H, Wang R, Luan J, Yu X, Dong L, Wang W, Xu S, Lu S and Shi M (2022) Characterizing microbiota and metabolomics analysis to identify candidate biomarkers in lung cancer. *Front. Oncol.* 12:1058436. doi: 10.3389/fonc.2022.1058436

COPYRIGHT
© 2022 Liu, Li, Suo, Zhang, Cao, Wang, Luan, Yu, Dong, Wang, Xu, Lu and Shi. 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.

Characterizing microbiota and metabolomics analysis to identify candidate biomarkers in lung cancer

Bo Liu[1,2,3,4†], Yige Li[1†], Lijun Suo[2,3†], Wei Zhang[5], Hongyun Cao[2,3], Ruicai Wang[6], Jiahui Luan[2,3], Xiaofeng Yu[2,3], Liang Dong[4], Wenjing Wang[1], Shiyang Xu[1,3], Shiyong Lu[3] and Mei Shi[1,3*]

[1]State Key Laboratory of Microbial Technology, Shandong University, Qingdao, China, [2]Department of Pulmonary and Critical Care Medicine, Department of Clinical Microbiology, Zibo City Key Laboratory of Respiratory Infection and Clinical Microbiology, Zibo City Engineering Technology Research Center of Etiology Molecular Diagnosis, Zibo Municipal Hospital, Zibo, China, [3]Shandong University-Zibo Municipal Hospital Research Center of Human Microbiome and Health, Zibo, China, [4]Department of Pulmonary and Critical Care Medicine, Shandong Institute of Respiratory Diseases, The First Affiliated Hospital of Shandong First Medical University, Shandong Provincial Qianfoshan Hospital, Shandong University, Jinan, China, [5]Department of Thoracic Surgery, Zibo Municipal Hospital, Zibo, China, [6]Department of Pathology, Zibo Municipal Hospital, Zibo, China

Background: Lung cancer is the leading malignant disease and cause of cancer-related death worldwide. Most patients with lung cancer had insignificant early symptoms so that most of them were diagnosed at an advanced stage. In addition to factors such as smoking, pollution, lung microbiome and its metabolites play vital roles in the development of lung cancer. However, the interaction between lung microbiota and carcinogenesis is lack of systematically characterized and controversial. Therefore, the purpose of this study was to excavate the features of the lung microbiota and metabolites in patients and verify potential biomarkers for lung cancer diagnosis.

Methods: Lung tissue flushing solutions and bronchoalveolar lavage fluid samples came from patients with lung cancer and non-lung cancer. The composition and variations of the microbiota and metabolites in samples were explored using muti-omics technologies including 16S rRNA amplicon sequencing, metagenomics and metabolomics.

Results: The metabolomics analysis indicated that 40 different metabolites, such as 9,10-DHOME, sphingosine, and cysteinyl-valine, were statistically significant between two groups (VIP > 1 and $P < 0.05$). These metabolites were significantly enriched into 11 signal pathways including sphingolipid, autophagy and apoptosis signaling pathway ($P < 0.05$). The analysis of lung microbiota showed that significant changes reflected the decrease of microbial diversity, changes of distribution of microbial taxa, and variability of the correlation networks of lung microbiota in lung cancer patients. In particular,

we found that oral commensal microbiota and multiple probiotics might be connected with the occurrence and progression of lung cancer. Moreover, our study found 3 metabolites and 9 species with significantly differences, which might be regarded as the potential clinical diagnostic markers associated with lung cancer.

Conclusions: Lung microbiota and metabolites might play important roles in the pathogenesis of lung cancer, and the altered metabolites and microbiota might have the potential to be clinical diagnostic markers and therapeutic targets associated with lung cancer.

KEYWORDS

lung cancer, lung microbiota, 16S rRNA amplicon sequencing, metabolomics, biomarker

Introduction

Lung cancer (LC) is the main cause of death from cancer worldwide, and its incidence has continued to rise in recent years. Each year, more than 2.2 million people are diagnosed with the disease and 1.79 million die from it (1). The incidence is still rising in China, compared with declines in some western countries, which constitutes a major public health problem and causes a huge social burden (2). In contrast to small cell lung cancer (SCLC), which has been declining in many countries, non-small cell lung cancer (NSCLC) has accounted for the largest proportion of LC (80%-90%) currently (3). For now, common diagnostic methods contain x-ray film, positron emission computed tomography (PET), computed tomography (CT), and computer aided detection and diagnosis (CAD) (4). However, these common techniques, which are conveniently used by medical staff, lack specificity and accuracy. Biomarkers were found to have the potential to assist in the early diagnosis of LC. The most widely used and reliable biomarkers are protein biomarkers found in blood and bronchoalveolar lavage fluid (5). Combining biomarkers, imaging omics and artificial intelligence to constitute an integrated model for LC screening and diagnosis might be the progression orientation for ameliorating LC prediction in the future.

The crucial risk factors of LC contain tobacco smoking, environmental and occupational pollution exposure, chronic lung disease, and lifestyle factors (6). Emerging studies have indicated the lung microbiota and metabolites could affect pulmonary health and diseases of the lungs. Lungs were considered as a sterile environment for a long time due to the limit of the culture-based techniques. However, the use of 16S ribosomal RNA (rRNA) amplicon sequencing has led to the increase of the interest in the lung microbiota (7). Numerous studies have shown that the lung microbiota might play a crucial part in the pathogenesis of pulmonary diseases. Liu et al. showed that the lung microbial composition and community structures of smokers with LC were distinct from that of emphysema-only patients: the abundance of Proteobacteria in the lungs of patients with LC was significantly lower and the abundances of *Streptococcus* and *Prevotella* were higher compared to patients with emphysema only (8). Tsay et al. found that *Streptococcus* and *Veillonella* were up-regulated in the lower airways of LC patients, which was related to the promotion of ERK and PI3K signal pathways (9). Moreover, studies had demonstrated a lower alpha diversity of lung microbiota in LC patients compared with that in patients with non-lung cancer (10). In general, compared with the studies on the gut, there were fewer studies on the correlation between lung microbiota and pulmonary homeostasis and diseases.

Recently, the relationships between the progression of chronic inflammatory diseases and the variations of microbiota have been gradually discovered, and the lung diseases involved include cystic fibrosis, asthma and chronic obstructive pulmonary disease (COPD) (11). The relationship between smoking, airflow obstruction, and LC was well recognized. Previously study showed that COPD was an important element in LC risk in smokers and smokers with COPD had a 3- to 10-fold increased risk of developing LC compared to smokers without emphysema or significant airway obstruction (12). Prolonged exposure to environmental pollutants could stimulate inflammatory factors, promote the formation of the environment suitable for the survival of pathogenic bacteria, and lead to the dysbiosis of lung microbiota; the dysbiosis could further induce inflammation and tissue damage, ultimately leading to accelerated decline in lung function (13). However, current researches had focused on the association of the lung microbiota with chronic inflammatory disease of the lungs, there have been fewer studies on LC.

In recent years, several studies have generated interest in the relationships between metabolites of the lung microbiota and lung health. Microbial components can contribute to the progression of the pulmonary diseases by producing metabolites with oncogenic potential. Gao et al. showed that the metabolites produced by *Pseudomonas aeruginosa* might be related to the pathogenesis of cystic fibrosis (14). In addition, the lung microbiota and metabolites contribute to the maintenance of the balance of the host lung immune system, which is an important contributor to defend against infection. Steed et al. found that desaminotyrosine (DAT), which was a metabolite associated with microbiota, helped the host defend against influenza by positively stimulating type I IFN (15). However, the current LC related metabolomics studies mostly targeted metabolites such as plasma proteins, which might not characterize the metabolism of lung microenvironment clearly.

Despite recent emerging studies on the correlation between lung microbiota and metabolites associated with LC, the mechanisms still need to be further clarified. In addition, few studies have considered both lung microbiota and metabolites to explore their possible associations and their roles in the pathogenesis of LC. Therefore, in our study, the differences in lung microbiota and metabolites between LC patients and patients with non-lung cancer were explored by 16S rRNA amplicon sequencing and metagenomics. Moreover, we used the samples of lung tissue flushing solutions, which could characterize the metabolic changes of lung microenvironment more clearly, for the analysis of metabolomics to explore their effects on the development of LC. The results indicated that lung microbiota and metabolites might play key roles in the development of LC, and the altered metabolites and microbiota might have the potential to be clinical diagnostic markers and therapeutic targets associated with LC.

Materials and methods

Participants

From 2020 to 2021, patients with LC were recruited in the Zibo Municipal Hospital. The exclusion criteria included the uses of antibiotics, corticoids, probiotics, prebiotics or immunosuppressive drugs in the past 3 months; hypertension; diabetes; previous airway surgery; preoperative radiotherapy and chemotherapy; and atomization treatment. Non-lung cancer patients were set as the control group of the metabolomics, and the exclusion criteria were the same as those in the LC group. This study was approved by Ethics Committee of Zibo Municipal Hospital (No. 20201102), and each subject signed a voluntary informed consent before the study. The clinical information was summarized in Supplementary Tables S1–S3.

Sample collection

Nine LC patients with unilateral tumors were selected from patients examined by bronchoscopy for the tests of 16S rRNA amplicon sequencing and metagenomics. All patients underwent routine examinations before operation, including electrocardiogram, pulmonary function, blood routine. Sterile saline samples of bilateral lungs were obtained by bronchoscopy in patients with LC. Paired samples of bronchoalveolar lavage fluid (BALF) included the one from the cancerous lobe and the other from the contralateral noncancerous lobe. Thirty LC patients with lung tumors and thirteen patients with non-lung cancer who underwent lobectomy were selected for the test of metabolomics. A whole tumor of 1 cm^3 and healthy tissue located 5 cm from tumor in the same pulmonary region were extracted for each patient. The removed tumors and tissues were immediately flushed with sterile normal saline and collected in sampling tubes.

All samples were immediately stored at -80 °C until DNA extraction was performed.

Non-targeted metabolomics profiling

Metabolites were extracted from the lung tissue flushing solutions and tested with a liquid chromatography-tandem mass spectrometry (LC-MS/MS). The metabolomics analysis was performed by UHPLC -Q Exactive HF-X system with a ACQUITY UPLC HSS T3 column (Waters, Milford, USA). The temperature of the column was set to 40°C and the injection volume was 2L. The flow rate of helium carrier gas was 0.4 mL/min, and the MS scanning range was m/z 70 - 1050. Progenesis QI (Waters Corporation, Milford, USA) was used to preprocess the MS raw data, and the obtained data matrix included retention time (RT), mass/charge ratio (M/Z) and peak intensity.

Principal component analysis (PCA) and orthogonal partial least squares-discriminant analysis (OPLS-DA) were used to explore whether all samples can be significantly clustered in different groups. The variable importance in projection (VIP) values of OPLS-DA and the P-values (Wilcoxon rank-sum test) were calculated to check the metabolites with statistically significant differences between two groups (16). Metabolites with P-values below 0.05 and VIP values above 1.00 were identified as differentially expressed metabolites. Metabolic pathway analysis was carried out to recognize the enriched pathways based on the altered metabolites. Altered metabolites were annotated through the KEGG database (https://www.kegg.jp/kegg/pathway.html) and the Python package SciPy was used for the pathway enrichment analysis. P-value were corrected by false discovery rate (FDR) with FDR ≤ 0.01 as the threshold.

16S rRNA amplicon sequencing

Microbial DNA was obtained from BALF samples using the FastDNA Spin Kit (MP Biomedicals, Shanghai, China) and tested for DNA purity using Nanodrop microspectrophotometer (Nanodrop 2000, Thermo Fisher Scientific, America). Finally, DNA integrity was determined using agarose gel electrophoresis. PCR was performed to amplification of the V3-V4 hypervariable regions of the bacterial 16S rDNA gene according to universal primers (338F: 5'- ACTCCTACGGGAGGCAGCAG- 3', 806R: 5'- GGACTACHVGGGTWTCTAAT- 3') that contained barcode (data available at Sequence Read Archive: PRJNA858534). PCR products were purified using the AxyPrep DNA Glue recovery kit, and the quantification and qualification of PCR products were detected on 2% agarose gels. Miseq libraries were constructed using NEXTFLEX DNA rapid Sequencing Kit and sequenced using Illumina's Miseq PE300 high-throughput sequencing platform. Raw data were demultiplexed, quality-filtered by fastp (https://github.com/OpenGene/fastp), and merged using FLASH (http://www.cbcb.umd.edu/software/flash). UPARSE (version 7.1 http://drive5.com/uparse/) software was used to perform OTU clustering on all sequences with 97% similarity as standard.

The three diversity indices (Shannon, Chao, ace) of the samples were calculated and averaged to assess the level of alpha diversity in different groups which were obtained by Mothur and visualized by R. The β-diversity was analyzed by weighted UniFrac phylogenetic distance matrices, visualized in non-metric multidimensional scaling analysis (NMDS) plots and determined by Partial Least Squares Discriminant Analysis (PLS-DA) for statistical significance. The effect of the abundance of the species on the discrepancy between groups was estimated using linear discriminant analysis (LDA) and formed a table (LDA 2.0, $P < 0.05$). Wilcoxon rank-sum test was carried out to compare species differences between groups ($P < 0.05$).

Correlation networks were used to show changes in interactions between microbial communities. Degree (DC), closeness (CC), and betweenness centrality (BC) were used to describe the characteristics of multiple networks.

Metagenomics analysis

Microbial DNA was obtained from BALF samples using the FastDNA Spin Kit (MP Biomedicals, Shanghai, China) and tested for DNA purity using Nanodrop microspectrophotometer (Nanodrop 2000, Thermo Fisher Scientific, America). Finally, DNA integrity was determined using agarose gel electrophoresis. DNA was fragmented to an average size of approximately 400 bp using Covaris M220 (Gene Company Limited, China) for paired-end library construction. DNA libraries were subsequently constructed and assessed using the NEXTFLEX Rapid DNA-Seq kit (Bioo Scientific, USA). The metagenomics sequencing was carried out on Illumina NovaSeq/Hiseq Xten (Illumina, USA, data available at Sequence Read Archive: PRJNA858501). The raw sequence reads were trimmed, and the clean reads were assembled *via* MEGAHIT. Gene prediction was performed using MetaGene (http://metagene.cb.k.u-tokyo.ac.jp/), and CD-HIT software (version 4.6.1 http://www.bioinformatics.org/cd-hit/) was used for predicting gene sequence clustering. Redundant gene sets were constructed using the longest sequence of each group clustered in DNA as representative. DIAMOND (https://github.com/bbuchfink/diamond) was employed to compare the sequences of non-redundant gene sets with Eggnog database (http://eggnog.embl.de/) to obtain the Clusters of Orthologous Groups (COG) functions corresponding to genes, and the relative abundance of the COG was calculated using the sum of gene abundances corresponding to COG.

Linear regression analysis was carried out to estimate the consistency between species and function. Significantly differences of COG categories between groups were detected by Wilcoxon rank-sum test ($P < 0.01$).

Biomarker identification

Biomarker identification was performed by MetaboAnalyst (https://www.metaboanalyst.ca/) (17). Based on the differential metabolites and microbiota obtained by the above analysis, the Receiver Operating Characteristic curve (ROC) analysis was used to obtain curve and calculate the area under the curve (AUC). In addition, we combined the obtained biomarkers to further explore the predictive ability of the model.

Results

Metabolomics profiles change in LC patients

The samples of lung tissue flushing solutions were used for the analysis of metabolomics. Based on the processing of the raw data, the area under the curve was used to quantify peaks. In positive (ESI+) modes, 8,650 positive peaks were detected, 428 metabolites were identified, and 125 metabolites were annotated compared with KEGG database. In negative (ESI-) modes, 5,580 negative peaks were detected, 178 metabolites were identified, and 55 metabolites were annotated compared with KEGG database. The data was normalized and relative standard deviation (RSD) was used to evaluate the exclusion of data with poor stability during the experiment. Results indicated that favorable stability was tested from samples in the positive and negative modes (Supplementary Figures S1A, B). PCA analysis revealed that QC samples were clearly differentiated, indicating the metabolomics datasets had satisfying stability and

repeatability (Supplementary Figures S1C, D). OPLS-DA analysis showed that the separation of metabolites in the samples of the two groups was obvious (Figure 1A). Metabolite features that distinguished LC patients from controls were selected based on a log2 fold change cutoff at 1, and VIP scores determined by OPLS-DA (VIP > 1, $P < 0.05$, Supplementary Table S4). We obtained 40 metabolites with significant differences in relative abundance between LC patients and controls (Figures 1B, C), which included 4 organic oxygen compounds, 4 fatty acyls, 3 organoheterocyclic compounds, 3 prenol lipids, 10 glycerophospholipids, 4 benzene and substituted derivatives, 2 carboxylic acids and derivatives, 1 benzenoid, 2 lipids and lipid-like molecules, 2 organic acids and derivatives, 1 purine nucleoside, and 4 other compounds (Supplementary Figure S1E). Overall, 14 and 26 metabolites were significantly up-regulated and down-regulated in LC patients, respectively. Several fatty acyls such as 9,10-DHOME, Erucic acid and N-Isobutyl-2,4,8,10,12-tetradecapentaenamide presented at higher levels in LC patients, and some glycerophospholipids such as PC (14:0/16:0) and PE (14:1(9Z)/14:1(9Z)) were down-regulated in LC patients.

KEGG pathway enrichment analysis was performed to explore the metabolic pathways associated with differential metabolites in LC patients and controls. 130 metabolic pathways were identified, among which 24 metabolic pathways had significant differences, including ABC transporters, protein digestion and absorption, central carbon metabolism in cancer ($P < 0.01$, Supplementary Figure S1F). The significantly different metabolites were enriched into a total of 15 signal pathways, of which 11 signaling pathways were observably changed in LC patients, comprising autophagy, apoptosis, necroptosis and sphingolipid signaling pathway ($P < 0.05$, Figure 1D).

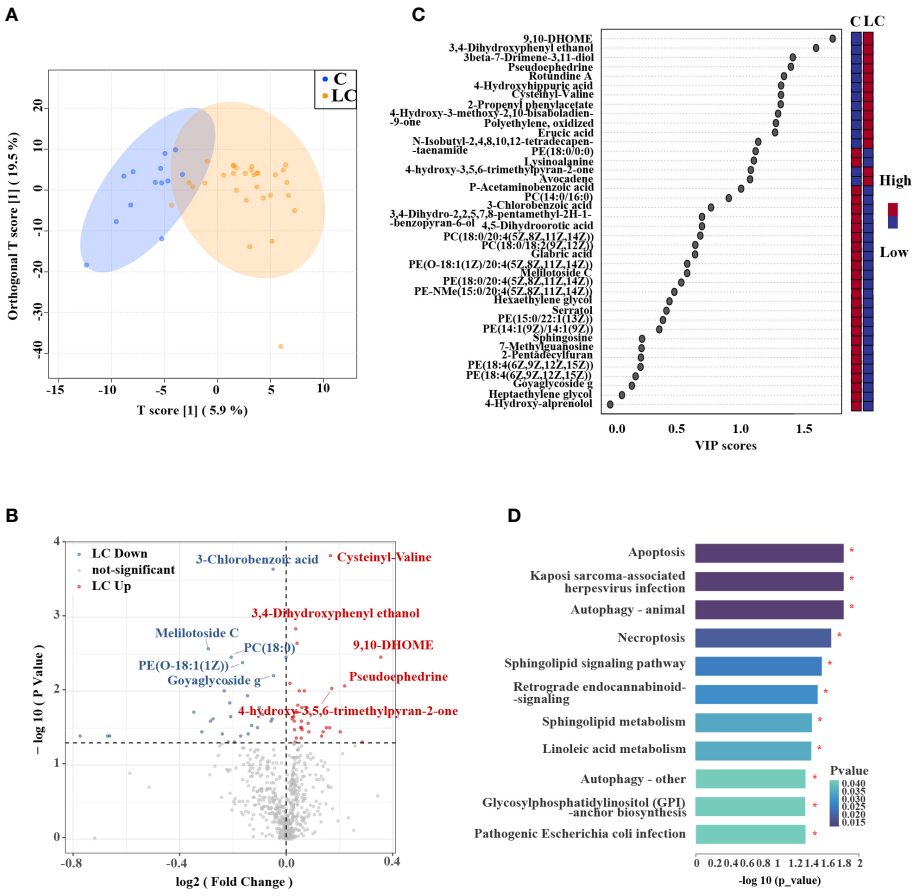


FIGURE 1
Metabolic profiles in LC patients and controls. **(A)** OPLS-DA showed that LC patients and controls were separated into two distinct clusters. **(B)** Volcano plot of metabolites of LC patients compared to controls. The y-axis represented p-value converted to negative log 10 scale and the x-axis represented log2 fold change. **(C)** Variable Importance in Projection (VIP) plot generated from the OPLS-DA analysis showed the most discriminative metabolites in descending order of importance. **(D)** KEGG pathway enrichment analysis of significantly different metabolites showed that there were 11 pathways had significant changes.

Altered composition of the lung microbiota in LC patients

BALF samples were used for 16S rRNA amplicon sequencing to explore the changes of lung microbiota in LC patients and a total of 16 samples passed quality control and were included in the study. According to Usearch statistics, in the raw data of 16S rDNA sequencing using primers 338F and 806R, the total reading of each sample was 888,409 pairs. The original data were filtered by QIIME software and then spliced by FLASH software to generate tags sequence. A total of 16 qualified samples were obtained by BALF sample sequencing, with an average length of 425 bases. Finally, Uparse software was used to cluster the spliced sequences into OTUs according to 97% similarity, and the total number of OTUs obtained was 1,711.

Species cumulative curve and rarefaction curve at the OTU level indicated that the vast majority of microbial diversity was obtained in all samples (Supplementary Figures S2A, B). Venn diagram was used to show the variation in OTUs between the two groups (Supplementary Figure S2C). Overall, 453 OTUs were shared between groups and there were more unique OTUs in controls (973) than in the LC patients (285). The results of PLS-DA model analysis reflecting the clustering of the two groups showed that the separation between LC patients and controls was obvious (Figure 2A). The alpha diversities of two groups did not show a significant difference (Supplementary

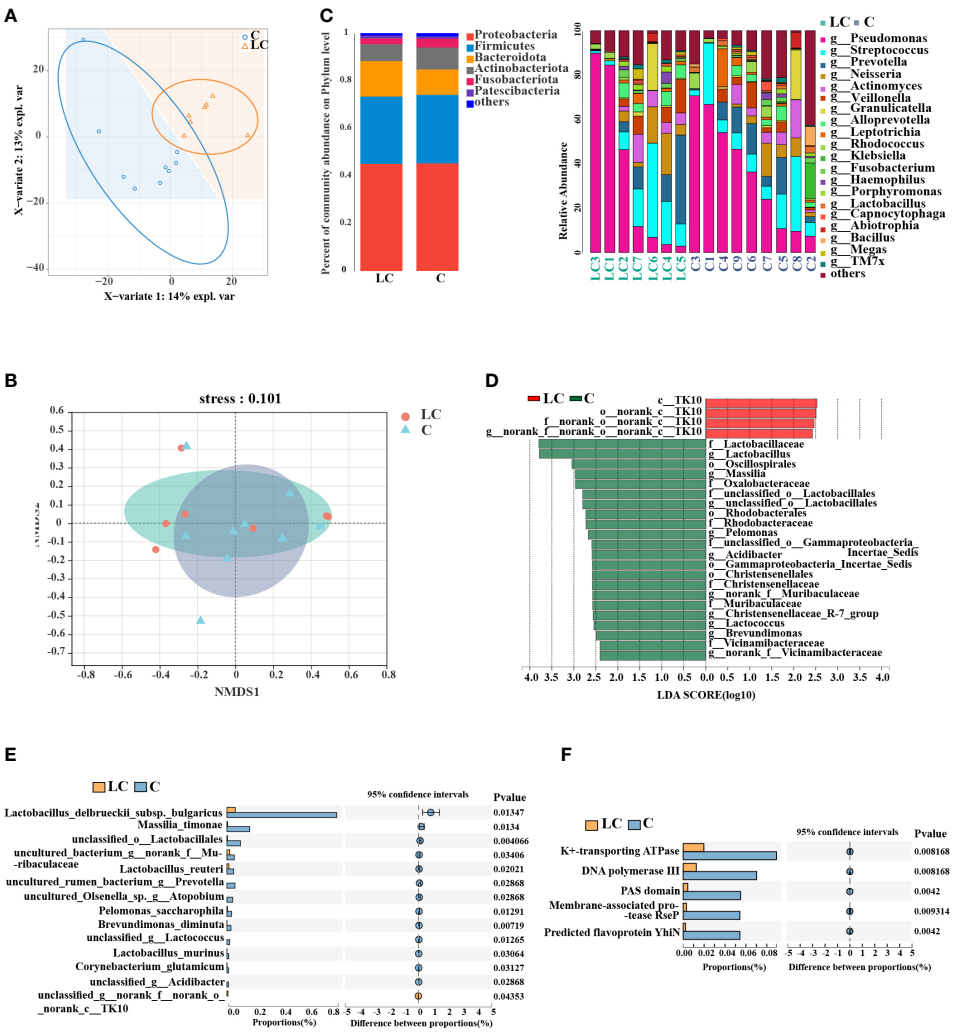


FIGURE 2
Lung microbiota composition profiles in LC patients and controls. **(A)** PLS-DA score plot of LC patients and controls showed clear distinction. **(B)** Non-metric multidimensional scaling analysis (NMDS) of the weighted UniFrac phylogenetic distance matrices demonstrated that the LC patients and controls showed two distinct clusters. **(C)** Relative abundance of major phyla and genera across BALF samples. **(D)** Differential taxa at the genus level identified by linear discriminant analysis (LDA) effect size (LEfSe) analysis (LDA > 2.0, $P < 0.05$). **(E)** Differential taxa at the species level identified by the Wilcoxon rank test ($P < 0.05$). **(F)** COG functional analysis of the microbiota between LC patients and controls.

Figure S2D). NMDS analysis on the basis of Bray-Curtis similarity distance indicated that the two groups were apart from each other on the ordination (stress<0.2, Figure 2B). A taxonomic analysis of sequences revealed that the most prevalent phylum in the lung microbial community was Proteobacteria and variations of microbial composition at the genus level between individuals could be seen (Figure 2C, Supplementary Table S5).

We relied on LEfSe analysis to identify the major taxa that influenced the differences between the two groups and two-sided Welch's t-test. LEfSe analysis recognized 26 genera which had discrepant abundances between the two groups (LDA > 2.0, $P < 0.05$). In LC patients, an enrichment in Chloroflexi taxa was observed and *Lactobacillus, Massilia, Lactococcus*, Oscillospirales, Christensenellaceae were significantly more abundant in controls (Figure 2D). Additionally, the results of the two-sided Welch's t-test showed that *Lactobacillus delbrueckii subsp. bulgaricus, Massilia timonae, Lactobacillus reuteri* were more abundant in controls by species taxa ($P < 0.05$, Figure 2E).

Taken together, we identified some microbiota and metabolites that were different between two groups and their changes may be correlated. Therefore, a heat map showed the association between 20 differential genera and 40 differential metabolites closely related with the progression of LC (Supplementary Figure S2E).

Then, we used metagenomics analysis to predict gene functions of the lung microbiota and a total of 12 samples met the criteria after quality inspection. Based on the construction of non-redundant gene sets, we obtained 12, 671 genes with a total sequence length of 6, 216,664 (bp) and an average sequence length of 490.62. 179 different COG functional categories were identified ($P < 0.05$, Supplementary Table S6), and there were 5 functional categories had significant differences ($P < 0.01$, Figure 2F), including K^+-transporting ATPase, DNA polymerase III, PAS domain, membrane-associated protease RseP and predicted flavoprotein YhiN. Linear regression analysis of the relationship between the similarity in the functional attributes of the community and community composition indicated that there is a prominent correlation between the two parts ($R^2 > 0.8$, $P < 0.01$, Supplementary Figure S2F).

Microbial interaction networks in non-lung cancer and lung cancer patients

To identify the interactions of the lung microbiota in patients with or without lung cancer, we constructed the correlation networks of genus taxa. The networks showed different bacterial interactions in the two groups, especially the network of LC patients was more complex than that of the controls. Given the distinct microbial composition between two groups, we compared the topology of the networks in each group. The number of mean degree and transitivity were higher in the LC patients (mean degree, 4.9; transitivity, 0.64) compared with the controls (mean degree, 3.6; transitivity, 0.58), suggesting that LC patients-enriched genera had a stronger correlation with each other than controls. The results indicated that patients-enriched species affected the host by interacting and exerting similar effects. Furthermore, degree centrality (DC), closeness centrality (CC) and betweenness centrality (BC) were used to screen the influential microbiota in each network (DC > 0.1, CC > 0.2, BC > 0.1). In LC patients, the roles of *Campylobacter, Atopobium, Haemophilus* and *Streptococcus* were several network-hubs and they were important to the lung microbial community alteration of the LC patients (Figure 3A). Results in controls showed that *Bacillus, Fusobacterium, Alloprevotella, Klebsiella* and *Kroppenstedtia* contributed to more importance (Figure 3B).

We constructed a correlation network combining the significantly different genera, which were obtained by 16S rRNA amplicon sequencing. *Lactobacillus, Brevundimonas, Massilia*, *Christensenellaceae R-7 group* were positively correlated with each other, which were enriched in controls, and they were negatively correlated with *Veillonella, Atopobium, Haemophilus*, *Fusobacterium*, which were enriched in LC patients (Figure 3C). Based on the measurement indexes characterizing the properties of the networks (DC > 0.1, CC > 0.2, BC > 0.1), *Brevundimonas, Bacillus, Veillonella, Klebsiella* and *Pseudomonas* were identified.

Identifying biomarkers in LC patients

Due to the function of evaluating the predictive ability of models, ROC curve has been in widespread use. ROC curve was used to assess representative differential features for the diagnosis of LC in this study. As indicated by the results, the AUC of Cysteinyl-Valine, 3-Chlorobenzoic acid and 3,4-Dihydroxyphenyl ethanol were 0.8692, 0.859 and 0.8103 (Supplementary Table S7), which might be useful in identifying patients with LC (Figures 4A, B). In order to improve the accuracy of biomarkers, the three metabolites were combined for ROC analysis, which showed more strikingly capability of the diagnosis for LC (AUC:0.91, Figure 4C). LEfSe analysis based on 16S rRNA amplicon sequencing revealed 14 significantly different species (LDA > 2.0, $P < 0.01$), from which nine species were screened by LASSO (Supplementary Figure S3, Supplementary Table S8). ROC analysis on the basis of the combination of the 9 species, demonstrating that LC could be assessed by representative differential lung microbiota (Figure 4D).

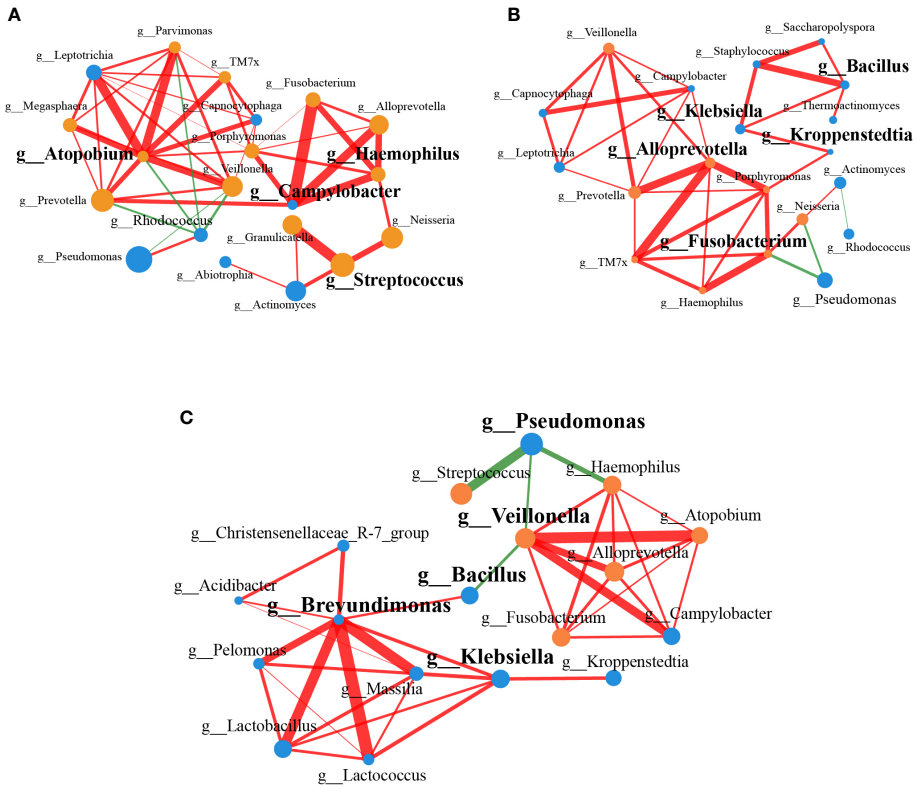


FIGURE 3
Correlation networks of genus associated with LC. **(A)** The correlation network between genera in LC patients (n=20). **(B)** The correlation network between genera in controls (n=20). **(C)** The correlation network based on the significantly different genera found by 16S rRNA amplicon sequencing (n=18, $P < 0.05$). The color of the nodes represented different groups, with nodes of the same color belonging to the same group. The yellow nodes represented LC group, and the blue nodes represented controls. The size of the nodes indicated the mean abundance of the genera in samples within the group. Edges between nodes represented the correlation between the abundance of two genera. Red lines represented positive correlations, while green lines represented negative correlations. The widths of the edges were proportional to the correlation strength, and wide line indicated stronger correlation.

Discussion

LC tumor microenvironment is colonized by microbiota, which can interact with the host, and new studies have indicated that this might be a potential factor affecting LC. Generally speaking, the normal tissue microenvironment protects the lungs, while the tumor microenvironment promotes cancer progression. Therefore, we used the samples which could characterize the changes of lung microenvironment to explore the effects of the lung microbiota and metabolites on the progression of LC. This study suggested that the altered microbiota and metabolites between the patients with or without lung cancer might play pivotal roles in LC pathogenesis.

In the metabolomics analysis of flushing fluid samples, multiple fatty acyls were significantly upregulated in LC patients and glycerophospholipids accounted for the largest proportion in controls, which indicated that lipid metabolism changed in LC patients. Increasing evidences suggested that lipid metabolism could be assisted in determining tumor metastasis, improving therapeutic efficacy and developing new therapeutic targets (18). Lipids are components of cell membranes that are involved not only in energy storage but also as messengers in signaling. In addition, the disorder of lipid metabolism in cancer cells will affect cell proliferation and differentiation and other processes (19). As the main components of pulmonary surfactant, which is a complex of phospholipids (85% phosphatidylcholine) and surfactant proteins, lipids have been shown to play essential roles in the pathogenesis of LC (20, 21). Pulmonary surfactant was synthesized and secreted by alveolar type II cell, a type of lung stem cell and it could transform into monoclonal lung tumor during active KRAS mutation in previous studies (22). Various studies have shown that the destruction of pulmonary surfactant and the changes of alveolar type II cell homeostasis were connected with the pathogenesis of LC (23).

In particular, we found that metabolites of sphingosine enriched in sphingolipid signaling pathway, significantly decreased in LC patients. Sphingolipids are bioactive

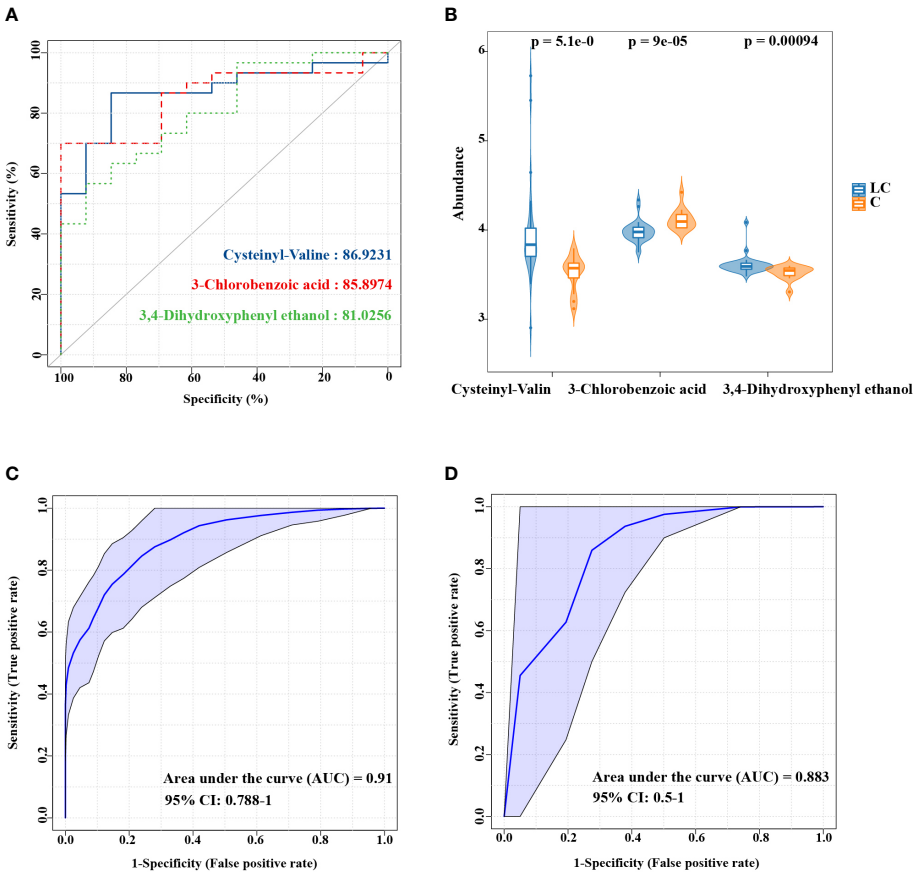


FIGURE 4
ROC curve analysis of the candidate biomarkers for LC. **(A)** Individual ROC curves and peak areas for Cysteinyl-Valine, 3-Chlorobenzoic acid and 3,4-Dihydroxyphenyl ethanol. **(B)** The differences in the abundance of three metabolic biomarkers between two groups. **(C)** ROC analysis based on the combination of three metabolites. **(D)** ROC analysis for the combination of 9 species screened by LASSO.

membrane lipids that act as first or second messengers (24, 25). In particular, the first sphingolipid detected was sphingosine, which could regulate various physiological processes such as cell cycle, apoptosis (26). Sphingosine, as a regulator that inhibits cell proliferation, can affect cell growth and apoptosis (27). Particularly, sphingosine is an important substance that helps protect the respiratory tract against bacterial pathogens (28). Sphingosine has been found to inhibit multiple pathogens, including *Staphylococcus aureus, Acinetobacter baumannii, Haemophilus influenzae, Escherichia coli, Fusobacterium nucleatum, Streptococcus sanguinis* (29). As the heat map showed, the bactericidal effect of sphingosine could have something to do with the downregulation of *Haemophilus* and *Streptococcus* in controls of our study.

Moreover, the pathway of ABC transporters, protein digestion and absorption and central carbon metabolism in patients were changed. Decreased level of ABC transporters was found in LC patients, containing betaine, L-Arginine and taurine. Betaine is widely regarded as an anti-oxidant and it has beneficial actions in several human diseases, such as obesity, diabetes and cancer (30). Tang et al. reported that choline-betaine pathway was conducive to hyperosmotic stress and lethal stress resistance in *Pseudomonas protegens SN15-2* (31), and this could have something to do with the enrichment of *Pseudomonas* in controls of our study. Arginine is present in the precursors of various organic compounds such as nitric oxide (NO), ornithine and myosine, which have huge impacts on immune cell biology, especially macrophage, dendritic cell and T cell immunobiology (32, 33). Kim et al. reported that arginine-induced changes in gut microbiota enhanced host lung immunity to nontuberculous mycobacterial infection, and that indicated that arginine might plays a protective role in lungs (34). Taurine, as conditionally essential amino acid of human, has multiple physiological functions, including the regulation of neural conduction, participating in endocrine activities, immunity enhancement, and strengthening the antioxidant capacity of cytomembrane (35). Taurine was found to inhibit the proliferation of lung cancer cells, significantly boosted the

apoptosis rate, and reduced the expression of migration factors matrix metallopeptidase 9 (MMP-9) and vascular endothelial growth factor (VEGF) (36, 37). Previous studies have shown that taurine ABC transporter protein has been identified in *Lactobacillus*, and this could have something to do with the enrichment of *Lactobacillus* in controls of our study (38). The upregulation of betaine, arginine and taurine in controls might contribute to the immunity enhancement and the boost of the antioxidant capacity of cells. The pathway of protein digestion and absorption and central carbon metabolism in cancer contained a variety of amino acids such as L-Tryptophan with decreased relative abundance in LC patients. Tryptophan is an essential amino acid and plays essential roles in various physiological processes. Down-regulated tryptophan concentration have been detected in patients with colorectal cancer, malignant melanoma and LC, and studies showed that tryptophan metabolites could drive the motility and migration of cancer cells (39). In addition, the pathway of Linoleic acid metabolism involved in metabolites of 9, 10-DHOME was up-regulated in LC. The level of 9, 10-DHOME, which was the epoxide hydrolase metabolite of the leukotoxin 9,10-EpOME, was found to be increased in disease. 9, 10-DHOME activates the NF-κB and AP-1 transcription factors of endotheliocyte to mediate inflammatory responses (40). Moreover, many studies showed that DiHOMEs might be part of the inflammatory response to environmental insults in lungs (41).

In this research, we explored the microbial changes in BALF samples using 16S rRNA amplicon sequencing. Results showed that the microbiota constitution in LC patients was different from that of controls and the microbiota differed in terms of beta-diversity. The microbial dysbiosis of LC patients was represented by decreasing microbial diversity, and increasing *Streptococcus*, *Prevotella*, *Veillonella* and *Haemophilus*, which were in accordance with the existing results (9, 42). Elevated abundance of *Streptococcus*, *Prevotella* and *Veillonella* were found in tumor tissues from LC patients previously, and the changes of these genera were related to the up-regulation of ERK and PI3K signaling pathways in LC patients (9). We also found that *Fusobacterium* was up-regulated in LC patients. The promotion effect of *Fusobacterium* on tumor cells is mainly achieved by inhibiting host immunity and inducing proinflammatory microenvironment (43). The available studies demonstrated that *Fusobacterium* acts as an inducer in various cancers, such as breast, colon and oral cancer (44–46). Several studies on the mechanism of *Fusobacterium* in promoting tumor development had provided different results. High levels of *Fusobacterium* promoted the activity of NF-κB and various pro-inflammatory factors, and the FadA virulence factor in *Fusobacterium* affected cell growth by regulating the β-catenin signaling pathway (47, 48).

LEfSe analysis showed that potential probiotics, including *Lactobacillus*, *Lactococcus*, Oscillospirales and Christensenellaceae, were down-regulated in LC patients. Probiotics were found to have the ability to achieve anticancer effects by promoting apoptosis of cancer cells and improving resistance to oxidative stress (49, 50). Multiple common microorganisms in the human gut have probiotic effects such as *Bifidobacterium*, *Lactobacillus*, *Lactococcus*. In particular, many lactic acid bacteria (LAB) have essential conducive impacts on the host such as anti-oxidation and anti-inflammation (51). The antioxidant capacity of LAB is based on the high catalase and α, α-diphenyl-β-picrylhydrazyl (DPPH) free radical scavenging activity, the anti-inflammatory property is achieved by the promotion of anti-inflammatory cytokines (IL-10) as well as the decrease of proinflammatory cytokines (IL-6) (43). Oscillospirales was believed to produce short-chain fatty acids, and the level of it was also found to be decreased in disease (52). The Christensenellaceae has been found in human bodies, which plays an important role in human health (53).

The correlation networks showed that multiple oral bacteria were enriched in the lungs, and there was a strong correlation between them such as *Veillonella, TM7x, Capnocytophaga, Parvimonas, Granulicatella*. There has been an increasing interest in detecting the connection between oral microbiota and the occurrence of respiratory tract infections. Associations between oral microbiota and several respiratory infections have been reported previously (54). A previous study found that oral commensal microbiota was enriched in the lower airway of LC patients, and the connections between the lower airway microbiota and host immunity in healthy subjects have also been explored (9). Previous studies confirmed that distinguishing oral commensal microbiotas were detected to have changes during the development of cancers such as pancreatic cancer, breast cancer or LC (55–57). However, none of them clearly elucidated the relationships between oral commensal microbiota and the pathogenesis of multiple cancers.

However, our results still have some shortcomings, and did not consider the tumor stage, the histological subtype and clinical validation. Many studies have found differences in the characteristics of microbiota between different tumor stages and histological subtypes in other cancers (58, 59). In the subsequent studies, we will expand sample size, and evaluate the potential marker in a larger cohort. We hope to verify the diagnostic value of the biomarkers and explore the molecular mechanisms by which lung microbiota and metabolites affect LC. Moreover, the relationship between lung microbiota and metabolites in different tumor stages and histological subtypes will be considered.

Conclusions

In this study, the differences in lung microbiota and metabolites between LC patients and patients with non-lung cancer were explored by 16S rRNA amplicon sequencing,

metagenomics and metabolomics. The results suggested that lung microbiota and metabolites might play critical roles in the progression of LC. The composition of the lung metabolites was significantly different between the LC patients and controls, which indicated that lipid metabolism, especially sphingolipid signaling pathway, changed in LC patients. The microbiota in LC patients were different from those in controls, with multiple probiotics were down-regulated in LC patients. Moreover, we found that oral commensal microbiota might be related to the development and progression of LC. Finally, we found 3 metabolites and 9 species, which have significantly differences, and they might have the potential to be clinical diagnostic markers and therapeutic targets associated with LC.

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://www.ncbi.nlm.nih.gov/, PRJNA858501, https://www.ncbi.nlm.nih.gov/, PRJNA858534.

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee of Zibo Municipal Hospital (20201102). The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

MS received the research grant, conceived and designed this study and revised the manuscript. BL and YL analyzed clinical and microbiome data and wrote the manuscript. LS enrolled the study participants, and revised the manuscript. WZ, RW, HC, JL, XY, SX and WW performed the experiments. LD and SL supervised and provided continuous guidance for the experiments. All authors approved the final version of the manuscript before submission.

Funding

This work was supported by the National Key Research and Development Program of China (grant no. 2021YFA0717002), Taishan Scholars Program of Shandong Province (grant no. tsqn202103196), and the Fundamental Research Funds for the Central Universities (grant no. 202172002).

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.2022.1058436/full#supplementary-material

References

1. International Agency for Research on Cancer. *Global cancer observatory: cancer today* (2020). Available at: https://gco.iarc.fr/today (Accessed Jan 19, 2020).

2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca-Cancer J Clin* (2018) 68:394–424. doi: 10.3322/caac.21492

3. Chansky K, Detterbeck FC, Nicholson AG, Rusch VW, Vallieres E, Groome P, et al. The IASLC lung cancer staging project: External validation of the revision of the TNM stage groupings in the eighth edition of the TNM classification of lung cancer. *J Thorac Oncol* (2017) 12:1109–21. doi: 10.1016/j.jtho.2017.04.011

4. Shende P, Augustine S, Prabhakar B, Gaud RS. Advanced multimodal diagnostic approaches for detection of lung cancer. *Expert Rev Mol Diagn* (2019) 19:409–17. doi: 10.1080/14737159.2019.1607299

5. Zamay TN, Zamay GS, Kolovskaya OS, Zukov RA, Petrova MM, Gargaun A, et al. Current and prospective protein biomarkers of lung cancer. *Cancers* (2017) 9:155. doi: 10.3390/cancers9110155

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

7. Dickson RP, Erb-Downward JR, Martinez FJ, Huffnagle GB. The microbiome and the respiratory tract. *Annu Rev Physiol* (2016) 78:481–504. doi: 10.1146/annurev-physiol-021115-105238

8. Liu YH, O'Brien JL, Ajami NJ, Scheurer ME, Amirian ES, Armstrong G, et al. Lung tissue microbial profile in lung cancer is distinct from emphysema. *Am J Cancer Res* (2018) 8:1775–87.

9. Tsay JCJ, Wu BG, Badri MH, Clemente JC, Shen N, Meyn P, et al. Airway microbiota is associated with upregulation of the PI3K pathway in lung cancer. *Am J Resp Crit Care* (2018) 198:1188–98. doi: 10.1164/rccm.201710-2118OC

10. Greathouse KL, White JR, Vargas AJ, Bliskovsky VV, Beck JA, von Muhlinen N, et al. Interaction between the microbiome and TP53 in human lung cancer. *Genome Biol* (2018) 19:123. doi: 10.1186/s13059-018-1501-6

11. Gollwitzer ES, Marsland BJ. Microbiota abnormalities in inflammatory airway diseases potential for therapy. *Pharmacol Therapeut* (2014) 141:32–9. doi: 10.1016/j.pharmthera.2013.08.002

12. Etzel CJ, Kachroo S, Liu M, D'Amelio A, Dong Q, Cote ML, et al. Development and validation of a lung cancer risk prediction model for African-americans. *Cancer Prev Res* (2008) 1:255–65. doi: 10.1158/1940-6207.Capr-08-0082

13. Marsland BJ, Gollwitzer ES. Host-microorganism interactions in lung diseases. *Nat Rev Immunol* (2014) 14:827–35. doi: 10.1038/nri3769

14. Gao B, Gallagher T, Zhang Y, Elbadawi-Sidhu M, Lai ZJ, Fiehn O, et al. Tracking polymicrobial metabolism in cystic fibrosis airways: *Pseudomonas aeruginosa* metabolism and physiology are influenced by *Rothia mucilaginosa*-derived metabolites. *Msphere* (2018) 3:e00151–18. doi: 10.1128/mSphere.00151-18

15. Steed AL, Christophi GP, Kaiko GE, Sun LL, Goodwin VM, Jain U, et al. The microbial metabolite desaminotyrosine protects from influenza through type I interferon. *Science* (2017) 357:498–502. doi: 10.1126/science

16. Saccenti E, Hoefsloot HCJ, Smilde AK, Westerhuis JA, Hendriks MMWB. Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics* (2014) 10:361–74. doi: 10.1007/s11306-013-0598-6

17. Chong J, Wishart DS, Xia J. Using MetaboAnalyst 4.0 for comprehensive and integrative metabolomics data analysis. *Curr Protoc Bioinf* (2019) 68:e86. doi: 10.1002/cpbi.86

18. Corbet C, Feron O. Emerging roles of lipid metabolism in cancer progression. *Curr Opin Clin Nutr* (2017) 20:254–60. doi: 10.1007/s12272-020-01273-8

19. Zhang JY, Huang YR, Tang XJ, Wei YP, Luo FM. Roles of lipid profiles in human non-small cell lung cancer. *Technol Cancer Res T* (2021) 20:1533033 8211041472. doi: 10.1177/15330338211041472

20. Salvador MM, de Cedron MG, Rubio JM, Martinez SF, Martinez RS, Casado E, et al. Lipid metabolism and lung cancer. *Crit Rev Oncol Hemat* (2017) 112:31–40. doi: 10.1016/j.critrevonc.2017.02.001

21. Whitsett JA, Wert SE, Weaver TE. Alveolar surfactant homeostasis and the pathogenesis of pulmonary disease. *Annu Rev Med* (2010) 61:105–19. doi: 10.1146/annurev.med.60.041807.123500

22. Desai TJ, Brownfield DG, Krasnow MA. Alveolar progenitor and stem cells in lung development, renewal and cancer. *Nature* (2014) 507:190–4. doi: 10.1038/nature12930

23. Lesko J, Triebl A, Stacher-Priehse E, Fink-Neubock N, Lindenmann J, Smolle-Juttner FM, et al. Phospholipid dynamics in ex vivo lung cancer and normal lung explants. *Exp Mol Med* (2021) 53:81–90. doi: 10.1038/s12276-020-00547-x

24. Bataller M, Sanchez-Garcia A, Garcia-Mayea Y, Mir C, Rodriguez I, LLeonart ME. The role of sphingolipids metabolism in cancer drug resistance. *Front Oncol* (2021) 11. doi: 10.3390/ijms22126492

25. Futerman AH, Hannun YA. The complex life of simple sphingolipids. *EMBO Rep* (2004) 5:777–82. doi: 10.1038/sj.embor.7400208

26. Smith ER, Merrill AH, Obeid LM, Hannun YA. Effects of sphingosine and other sphingolipids on protein kinase c. *Method Enzymol* (2000) 312:361–73. doi: 10.1016/s0076-6879(00)12921-0

27. Kohama T, Olivera A, Edsall L, Nagiec MM, Dickson R, Spiegel S. Molecular cloning and functional characterization of murine sphingosine kinase. *J Biol Chem* (1998) 273:23722–8. doi: 10.1074/jbc.273.37.23722

28. Wu YQ, Liu YJ, Gulbins E, Grassme H. The anti-infectious role of sphingosine in microbial diseases. *Cells-Basel* (2021) 10:1105. doi: 10.3390/cells10051105

29. Verhaegh R, Becker KA, Edwards MJ, Gulbins E. Sphingosine kills bacteria by binding to cardiolipin. *J Biol Chem* (2020) 295:7686–96. doi: 10.1074/jbc.RA119.012325

30. Zhao GF, He F, Wu CL, Li P, Li NZ, Deng JP, et al. Betaine in inflammation: Mechanistic aspects and applications. *Front Immunol* (2018) 9:1070. doi: 10.3389/fimmu.2018.01070

31. Tang DY, Wang XB, Wang J, Wang MF, Wang YP, Wang W. Choline-betaine pathway contributes to hyperosmotic stress and subsequent lethal stress resistance in *Pseudomonas protegens* SN15-2. *J Biosci* (2020) 45:85. doi: 10.1007/s12038-020-00060-3

32. Morris SM. Arginine: Master and Commander in innate immune responses. *Sci Signal* (2010) 3:pe27. doi: 10.1126/scisignal.3135pe27

33. Serbina NV, Salazar-Mather TP, Biron CA, Kuziel WA, Pamer EG. TN/iNOS-producing dendritic cells mediate innate immune defense against bacterial infection. *Immunity* (2003) 19:59–70. doi: 10.1016/s1074-7613(03)00171-7

34. Kim YJ, Lee JY, Lee JJ, Jeon SM, Silwal P, Kim IS, et al. Arginine-mediated gut microbiome remodeling promotes host pulmonary immune defense against nontuberculous mycobacterial infection. *Gut Microbes* (2022) 14:2073132. doi: 10.1080/19490976.2022.2073132

35. Jong CJ, Sandal P, Schaffer SW. The role of taurine in mitochondria health: More than just an antioxidant. *Molecules* (2021) 26:4913. doi: 10.3390/molecules26164913

36. Tu S, Zhang XL, Wan HF, Xia YQ, Liu ZQ, Yang XH, et al. Effect of taurine on cell proliferation and apoptosis human lung cancer A549 cells. *Oncol Lett* (2018) 15:5473–80. doi: 10.3892/ol.2018.8036

37. Deng YY, Li HJ, Tang YJ. The effect of suppression taurine on relocation and epithelial-mesenchymal transition in mankind lung cancer cells. *J Healthc Eng* (2021) 2021:6656080. doi: 10.1155/2021/6656080

38. Giri R, Sharma RK. Analysis of protein association networks regulating the neuroactive metabolites production in *Lactobacillus* species. *Enzyme Microb Tech* (2022) 154:109978. doi: 10.1016/j.enzmictec.2021.109978

39. Platten M, Nollen EAA, Rohrig UF, Fallarino F, Opitz CA. Tryptophan metabolism as a common therapeutic target in cancer, neurodegeneration and beyond. *Nat Rev Drug Discov* (2019) 18:379–401. doi: 10.1038/s41573-019-0016-5

40. Viswanathan S, Hammock BD, Newman JW, Meerarani P, Toborek M, Hennig B. Involvement of CYP2C9 in mediating the proinflammatory effects of linoleic acid in vascular endothelial cells. *J Am Coll Nutr* (2003) 22:502–10. doi: 10.1080/07315724.2003.10719328

41. Hildreth K, Kodani SD, Hammock BD, Zhao L. Cytochrome P450-derived linoleic acid metabolites EpOMEs and DiHOMEs: a review of recent studies. *J Nutr Biochem* (2020) 86:108484. doi: 10.1016/j.jnutbio.2020.108484

42. Shen WT, Tang DR, Deng YL, Li HL, Wang T, Wan P, et al. Association of gut microbiomes with lung and esophageal cancer: a pilot study. *World J Microb Biot* (2021) 37:128. doi: 10.1007/s11274-021-03086-3

43. Badgeley A, Anwar H, Modi K, Murphy P, Lakshmikuttyamma A. Effect of probiotics and gut microbiota on anti-cancer drugs: Mechanistic perspectives. *Bba-Rev Cancer* (2021) 1875:188494. doi: 10.1016/j.bbcan.2020.188494

44. Yu TC, Guo FF, Yu YN, Sun TT, Ma D, Han JX, et al. *Fusobacterium nucleatum* promotes chemoresistance to colorectal cancer by modulating autophagy. *Cell* (2017) 170:548–63. doi: 10.1016/j.cell.2017.07.008

45. Parhi L, Alon-Maimon T, Sol A, Nejman D, Shhadeh A, Fainsod-Levi T, et al. Breast cancer colonization by *Fusobacterium nucleatum* accelerates tumor growth and metastatic progression. *Nat Commun* (2020) 11:3259. doi: 10.1038/s41467-020-16967-2

46. McIlvanna E, Linden GJ, Craig SG, Lundy FT, James JA. *Fusobacterium nucleatum* and oral cancer: a critical review. *BMC Cancer* (2021) 21:1212. doi: 10.1186/s12885-021-08903-4

47. Kostic AD, Chun EY, Robertson L, Glickman JN, Gallini CA, Michaud M, et al. *Fusobacterium nucleatum* potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe* (2013) 14:207–15. doi: 10.1016/j.chom.2013.07.007

48. Rubinstein MR, Wang XW, Liu WD, Hao YJ, Cai GF, Han YPW. *Fusobacterium nucleatum* promotes colorectal carcinogenesis by modulating e-Cadherin/beta-Catenin signaling *via* its FadA adhesin. *Cell Host Microbe* (2013) 14:195–206. doi: 10.1016/j.chom.2013.07.012

49. Wollowski I, Rechkemmer G, Pool-Zobel BL. Protective role of probiotics and prebiotics in colon cancer. *Am J Clin Nutr* (2001) 73:451s–5s. doi: 10.1093/ajcn/73.2.451s

50. Nistal E, Fernandez-Fernandez N, Vivas S, Olcoz JL. Factors determining colorectal cancer: the role of the intestinal microbiota. *Front Oncol* (2015) 5:220. doi: 10.3389/fonc.2015.00220

51. Ayyanna R, Ankaiah D, Arul V. Anti-inflammatory and antioxidant properties of probiotic bacterium *Lactobacillus mucosae* AN1 and *Lactobacillus fermentum* SNR1 wistar albino rats. *Front Microbiol* (2018) 9:3063. doi: 10.3389/fmicb.2018.03063

52. Ruuskanen MO, Aberg F, Mannisto V, Havulinna AS, Meric G, Liu Y, et al. Links between gut microbiome composition and fatty liver disease in a large population sample. *Gut Microbes* (2021) 13:1–22. doi: 10.1080/19490976.2021.1888673

53. Huang WY, Yau Y, Zhu JR, Wang YJ, Dai ZP, Gan HJ, et al. Effect of electroacupuncture at zusanli (ST36) on intestinal microbiota in rats with chronic atrophic gastritis. *Front Genet* (2022) 13:824739. doi: 10.3389/fgene.2022.824739

54. Scannapieco FA. Role of oral bacteria in respiratory infection. *J Periodontol* (1999) 70:793–802. doi: 10.1902/jop.1999.70.7.793

55. Zhang L, Farrell JJ, Zhou H, Elashoff D, Akin D, Park NH, et al. Salivary transcriptomic biomarkers for detection of resectable pancreatic cancer. *Gastroenterology* (2010) 138:949–U194. doi: 10.1053/j.gastro.2009.11.010

56. Agha-Hosseini F, Mirzaii-Dizgah I, Rahimi A. Correlation of serum and salivary CA15-3 levels in patients with breast cancer. *Med Oral Patol Oral* (2009) 14:E521–m4. doi: 10.4317/medoral.14.e521

57. Li XZ, Yang TY, Lin JX. Spectral analysis of human saliva for detection of lung cancer using surface-enhanced raman spectroscopy. *J BioMed Opt* (2012) 17:037003. doi: 10.1117/1.Jbo.17.3.037003

58. Yachida S, Mizutani S, Shiroma H, Shiba S, Nakajima T, Sakamoto T, et al. Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nat Med* (2019) 25:968–76. doi: 10.1038/s41591-019-0458-7

59. Sarhadi V, Mathew B, Kokkola A, Karla T, Tikkanen M, Rautelin H, et al. Gut microbiota of patients with different subtypes of gastric cancer and gastrointestinal stromal tumors. *Gut Pathog* (2021) 13:11. doi: 10.1186/s13099-021-00403-x

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 30 November 2022
DOI 10.3389/fonc.2022.1047019

Check for updates

OPEN ACCESS

EDITED BY
Hongda Liu,
Nanjing Medical University, China

REVIEWED BY
Jing Yang,
Chongqing Technology and Business University, China
Qun Wang,
The First Affiliated Hospital of Soochow University, China
Mingran Xie,
USTC Life Sciences and Medicine, China

*CORRESPONDENCE
Jian Hu
dr_hujian@zju.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

RECEIVED 17 September 2022
ACCEPTED 04 November 2022
PUBLISHED 30 November 2022

CITATION
He T, Wu Z, Xia P, Wang W, Sun H, Yu L, Lv W and Hu J (2022) The combination of a seven-autoantibody panel with computed tomography scanning can enhance the diagnostic efficiency of non-small cell lung cancer.
Front. Oncol. 12:1047019.
doi: 10.3389/fonc.2022.1047019

COPYRIGHT
© 2022 He, Wu, Xia, Wang, Sun, Yu, Lv and Hu. 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 combination of a seven-autoantibody panel with computed tomography scanning can enhance the diagnostic efficiency of non-small cell lung cancer

Tianyu He†, Ziheng Wu†, Pinghui Xia, Weidong Wang, Hua Sun, Li Yu, Wang Lv and Jian Hu*

Department of Thoracic Surgery, the First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

Background: Non-small cell lung cancer (NSCLC) is still of concern in differentiating it from benign disease. This study aims to validate the diagnostic efficacy of a novel seven-autoantibody (7-AAB) panel for the diagnosis of NSCLC.

Methods: We retrospectively enrolled 2650 patients who underwent both the 7-AAB panel test and CT scanning. We compared the sensitivity, specificity, and PPV of 7-AAB, CT, and PET-CT in the diagnosis of NSCLC in different subgroups. Then, we established a nomogram based on CT image features and the 7-AAB panel to further improve diagnostic efficiency. Moreover, we compared the pathological and molecular results of NSCLC patients in the 7-AABs positive group and the negative group to verify the prognostic value of the 7-AAB panel.

Results: The strategy of a "both-positive rule" combination of 7-AABs and CT had a specificity of 95.4% and a positive predictive value (PPV) of 95.8%, significantly higher than those of CT or PET-CT used alone ($P<0.05$). The nomogram we established has passed the calibration test ($P=0.987>0.05$) with an AUC of 0.791. Interestingly, it was found that the 7-AABs positive group was associated with higher proportion of EGFR mutations ($P<0.001$), lower pathological differentiation degrees ($P=0.018$), more advanced pathological stages ($P=0.040$) and higher Ki-67 indexes ($P=0.011$) in patients with adenocarcinoma.

Conclusion: This study shows that combination of a 7-AAB panel with CT has can significantly enhance the diagnostic efficiency of lung cancer. Moreover, the 7-AAB panel also has potential prognostic value and has reference significance for the formulation of the treatment plan.

KEYWORDS

NSCLC, nomogram, diagnosis, autoantibodies, EGFR mutation

Introduction

According to the latest statistics of cancer incidence and mortality produced by the International Agency for Research on Cancer, lung cancer remained the leading cause of cancer death, with an estimated 1.8 million deaths (18%) worldwide (1). Non–small-cell lung cancer (NSCLC) is the most common pathological pattern of lung cancer, which accounts for over 80% of cases. For NSCLC, the five-year survival rate differs from 92% for patients with stage IA1 disease to only 6% for patients with stage IV disease (2). Therefore, detection of NSCLC at an early stage is critical to improve the overall survival. Regrettably, a large number of patients had been diagnosed with advanced diseases in the past.

Notably, two large randomized controlled trials provided evidence of statistically significant reductions in lung cancer mortality, which benefited from the widespread use of low-dose computerized tomography (LDCT) screening in high-risk populations (3, 4). Although LDCT demonstrates a sensitivity of more than 90% for detecting pulmonary nodules, it still lacks sufficient accuracy to distinguish benign nodules from early-stage lung cancer. So, it leads to high rate of reports of false-positive nodules, as well as unnecessary following-up or surgical procedures. At present, it is urgent to develop a novel examination method to assist CT in improving the diagnostic efficiency of NSCLC.

Due to its non-invasive and reproducible characteristics, serum marker detection can be a good supplement to LDCT. Traditional serum lung cancer biomarkers, such as carcinoembryonic antigen (CEA), squamous cell carcinoma antigen (SCCA) and neuron-specific enolase (NSE), have been clinically used for years. However, they are of limited value in detecting early-stage lung cancer. Tumor-associated antigens (TAAs) exist in most types of cells and are involved in biological processes such as proliferation and differentiation. However, when malignant cells deviate from their normal state, TAAs may be detected by the immune system and generate corresponding autoantibodies (AABs) (5–7). Interestingly, AABs can be detected in an early stage of tumor, especially before significant clinical symptoms appear (8–11). A previous study on a test panel consisting of six serum AABs (p53, NY-ESO-1, CAGE, GBU4-5, Annexin 1 and SOX2) demonstrated a sensitivity/specificity of 36%-39%/89%-91% (12). A prospective study evaluating the EarlyCDT-Lung blood test for differentiating pulmonary nodules between benign and malignant demonstrated that using the "both-positive rule" combination of EarlyCDT-Lung and CT can significantly improve diagnostic specificity (>92%) and positive prediction value (>70%) (13).

In this study, we retrospectively evaluated the diagnostic performance of a 7-AAB panel (p53, PGP9.5, SOX2, GAGE7, GBU4-5, MAGEA1 and CAGE) in lung cancer and compared it with CT and PET-CT for pulmonary nodules with different diameters, stages, imaging features, and pathological types in the same population. At the same time, we summarized the imaging, pathological and molecular characteristics of 7-AAB positive and negative groups in lung cancer patients, suggesting that 7-AAB may have a unique value in the prognostic prediction of lung cancer patients.

Materials and methods

Patients

This study included 2824 patients who underwent 7-AAB panel tests in The First Affiliated Hospital, Zhejiang University from January 2020 to April 2022. At the same time, basic clinical information, CT & PET-CT scanning reports (within 1 month before and after the 7-AAB panel test), pathological data, and Next Generation Sequencing (NGS) results were collected. The inclusion and exclusion procedure of patients is shown in Figure 1. The CT images are interpreted by at least two experienced radiologists and the final diagnosis is made, as is PET-CT. Patients with pathologically confirmed NSCLC were recommended to undergo NGS detection for their surgically resected tumor specimens according to the clinician's judgment and the patient's wishes. NGS testing includes at least the following ten lung cancer-related gene mutations: EGFR (19-DEL, L858R, T790M), KRAS, BRAF, ERBB2, ALK, MET, RET, ROS1, PIK3CA, TP53.

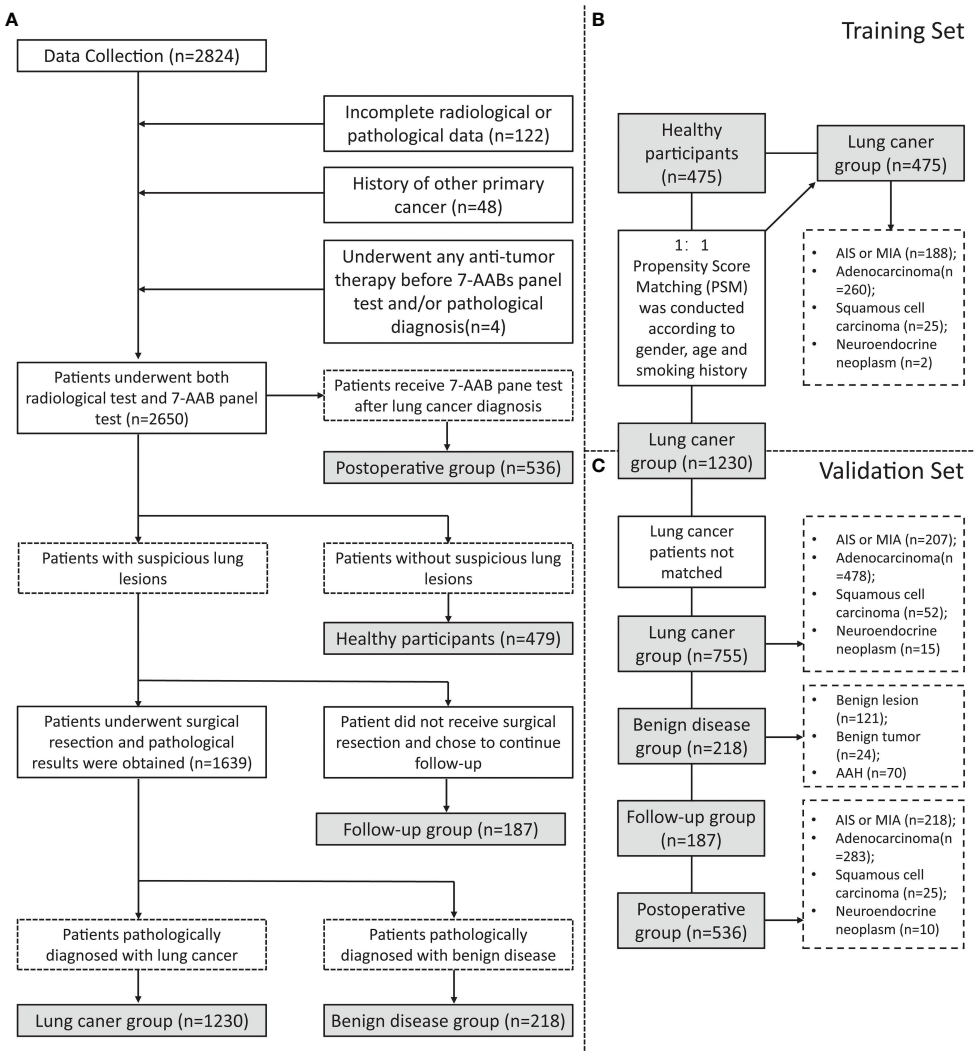


FIGURE 1
Flowchart of inclusion and grouping. **(A)** A flowchart of the inclusion and exclusion of participants in this study; **(B)** A flowchart of participants in training set matching by Propensity Score Matching (PSM); **(C)** A flowchart of participants inclusion in validation set.

This study was approved by the Medical Ethics Committee and institutional review board of The First Affiliated Hospital, Zhejiang University (Ethical number: IIT20220699A).

Quantitative detection of serum 7-AABs

The concentrations of serum autoantibodies were quantified by indirect enzyme-linked immunosorbent assay (ELISA) using a detection kit (Cancer Probe Biological Technology Co., Ltd., Hangzhou, China). The procedure of autoantibodies detection was illustrated in Figure S1. Briefly, 5ml of venous blood was collected from the patients (using a procoagulant tube), and the upper serum was obtained after centrifugation. The serum samples were diluted with phosphate-buffered saline (PBS) in a ratio of 1:109 and added into antigen-coated microwell plates (50μL/well). The microplate was oscillated and incubated at room temperature for binding of the autoantibodies (primary antibodies) in serum to the pre-incubated TAAs. After washing off the free antibodies with washing buffer, diluted horseradish peroxidase (HRP)-conjugated anti-human IgG (secondary antibody) was added to bind them to the primary antibody. Plates were washed once again using a buffer. Then add chromogenic substrate and incubate with oscillation at room temperature under dark conditions. After the reaction was terminated, the OD450 value was measured with a spectrophotometer immediately.

Calibration wells and quality control wells were set to draw standard curves. The optimal cutoff values of the 7-AABs were

defined as mean OD value ± 3 standard deviations (SDs) of the healthy group in the training set. These cutoff values are also used to verify whether the 7-AAB panel is positive or not in the validation set.

Statistical analysis

Continuous variables were represented by mean ± SD and analyzed by t-test, while categorical variables were expressed as Counts (Percentages) and analyzed by chi-square ($\chi 2$) test or Fisher exact test. Sensitivity, specificity, and positive diagnostic value (PPV) were calculated, and the diagnostic value of each diagnostic method was evaluated by area under the curve (AUC) and standard error (SE) for the respective receiver operating characteristic (ROC) curve. A Nomogram was established using logistic regression model and the diagnostic value of this model was evaluated by the calibration curve and ROC curve. For all the analyses, a 2-sided P-value of <0.05 were considered statistically significant. All analyses were performed using SPSS 26.0 software (IBM, Armonk, NY), GraphPad Prism 8.0 Software (GraphPad software, La Jolla, CA), and R 4.1.3 (The R Foundation for Statistical Computing, Vienna, Austria) with the rms statistical package.

Results

Study population

To investigate the diagnostic efficiency of the 7-AAB panel, a total of 2650 participants were finally enrolled in this study (Figure 1A). They comprised a training set (n=950, Figure 1B) and a validation set (n=1698, Figure 1C). The baseline clinical characteristics of the patients are summarized in Tables 1, 2. In this study, 475 healthy participants and 475 patients with pathologically-confirmed NSCLC were 1:1 matched (gender, age, and smoking history) using Propensity Score Matching (PSM) in the training set. Meanwhile, 755 patients with NSCLC, 218 patients with benign disease, 187 patients with suspicious lung lesions and continuing follow-up, and 536 postoperative patients (patients receive 7-AAB pane test 6 months to 1 year after surgery) are included in the validation set. Most were females (63.0%), non-smokers (79.9%) and with stage-I disease (80.5%).

The serum concentration of 7-AABs in the training set

The serum concentration of the 7 AABs (p53, PGP9.5, SOX2, GAGE7, GBU4-5, MAGEA1 and CAGE) were quantitated by using indirect ELISA. The results showed that the serum concentrations of all 7 AABs in patients with pathologically-confirmed NSCLC were higher than those in healthy participants with statistically significant difference ($P<0.05$) (Figure 2 & Table S1).

TABLE 1 Baseline clinical characteristics of patients in the training set.

Training set			
	Healthy participants (n=475)	**Lung cancer (n=475)**	**P value**
Gender, n (%)			
Female	316 (66.5%)	299 (62.9%)	0.248
Male	159 (33.5%)	176 (37.1%)	
Age, years (median, range)	51.2 ± 12.1	51.3 ± 12.3	0.847
Smoking history, n (%)			
Never	392 (82.5%)	374 (78.7%)	0.139
Current/Ever	83 (17.5%)	101 (21.3%)	
Lesion location, n (%)			
Upper left	0	116 (24.4%)	–
Lower left	0	70 (14.7%)	
Upper right	0	157 (33.1%)	
Middle right	0	39 (8.2%)	
Lower right	0	93 (19.6%)	
Pathologic type, n (%)			
AIS or MIA	0	188 (39.6%)	–
IAC	0	260 (54.7%)	
SCC	0	25 (5.3%)	
Neuroendocrine neoplasm	0	2 (0.4%)	
Pathological stage, n (%)			
IA	0	264 (55.6%)	–
IB	0	118 (24.8%)	
IIA	0	19 (4.0%)	
IIB	0	38 (8.0%)	
IIIA	0	8 (1.7%)	
IIIB	0	12 (2.5%)	
IVA	0	10 (2.1%)	
Group of diameters, n (%)			
≤8mm	0	171 (35.7%)	–
8<GGO ≤ 20mm	0	227 (47.4%)	
≥20mm	0	81 (16.9%)	
Number of nodules, n (%)			
Single	0	174 (36.3%)	–
Multiple	0	303 (63.3%)	
Composition, n (%)			
Pure GGO	0	49 (10.2%)	–
Mix GGO	0	305 (63.7%)	
Solid nodule	0	123 (25.7%)	

Determination of the cut-off values of the 7-AABs

The optimal cutoff values of the 7-AABs were defined as mean OD value ± 3 standard deviations (SDs) of the healthy participant's group as previous study described (12). According to the calculation, the cutoff values of p53, PGP9.5, SOX2, GAGE7, GBU4-5, MAGEA1 and CAGE were 17.44 u/ml, 8.16

10.3389/fonc.2022.1047019

TABLE 2 Baseline clinical characteristics of patients with surgically resected lung nodules.

Validation set

	Lung cancer (n=755)	Benign disease (n=218)	Follow-up group (n=187)	Post-operative group (n=536)
Gender, n (%)				
Female	485 (64.2%)	144 (66.1%)	107 (57.2%)	315 (58.8%)
Male	270 (35.8%)	74 (33.9%)	80 (42.8%)	221 (41.2%)
Age, years (median, range)	60 ± 11	54 ± 12	52 ± 15	59.0 ± 11.3
Smoking history, n (%)				
Never	606 (80.3%)	184 (84.4%)	149 (79.7%)	411 (76.7%)
Current/Ever	149 (19.7%)	34 (15.6%)	38 (20.3%)	125 (23.3%)
Lesion location, n (%)				
Upper left	190 (25.2%)	45 (20.6%)	38 (21.6%)	129 (24.1%)
Lower left	125 (16.6%)	39 (17.9%)	15 (8.5%)	85 (15.9%)
Upper right	244 (32.3%)	72 (33%)	64 (36.4%)	160 (29.9%)
Middle right	53 (7%)	18 (8.3%)	13 (7.4%)	48 (9.0%)
Lower right	143 (18.9%)	44 (20.2%)	46 (26.1%)	114 (21.3%)
Pathologic type, n (%)				
Benign lesion	0	121 (56.3%)	0	0
Benign tumor	0	24 (11.2%)	0	0
AAH	0	70 (32.6%)	0	0
AIS or MIA	207 (27.5%)	0	0	179 (33.4%)
IAC	478 (63.6%)	0	0	305 (56.9%)
SCC	52 (6.9%)	0	0	40 (7.5%)
Neuroendocrine neoplasm	15 (2%)	0	0	12 (2.2%)
Pathological stage, n (%)				
IA	586 (77.6%)	0	0	423 (78.9%)
IB	51 (6.8%)	0	0	42 (7.8%)
IIA	31 (4.1%)	0	0	17 (3.2%)
IIB	41 (5.4%)	0	0	23 (4.3%)
IIIA	26 (3.4%)	0	0	20 (3.7%)
IIIB	11 (1.5%)	0	0	8 (1.5%)
IVA	9 (1.2%)	0	0	3 (0.6%)
Group of diameters, n (%)				
≤8mm	182 (24.1%)	98 (45%)	99 (52.9%)	161 (30.0%)
8<GGO ≤ 20mm	364 (48.2%)	88 (40.4%)	72 (38.5%)	256 (47.8%)
≥20mm	209 (27.7%)	32 (14.7%)	16 (8.6%)	119 (22.2%)
Number of nodules, n (%)				
Single	493 (65.3%)	159 (73.3%)	100 (53.5%)	360 (67.2%)
Multiple	262 (34.7%)	58 (26.7%)	86 (46.0%)	176 (32.8%)
Composition, n (%)				
Pure GGO	49 (6.5%)	27 (12.4%)	27 (14.4%)	51 (9.5%)
Mix GGO	419 (55.5%)	86 (39.6%)	114 (61.0%)	318 (59.3%)
Solid nodule	287 (38%)	104 (47.9%)	44 (23.5%)	167 (31.2%)
Spiculation sign, n (%)	459 (60.8%)	88 (40.4%)	61 (32.6%)	295 (55.0%)
Vessels sign, n (%)	560 (74.2%)	97 (44.5%)	74 (39.6%)	374 (69.8%)
Vocule sign, n (%)	302 (40%)	53 (24.4%)	41 (21.9%)	210 (39.2%)
Air bronchogram, n (%)	188 (24.9%)	25 (11.5%)	12 (6.4%)	111 (20.7%)
Pleural indentation, n (%)	231 (30.6%)	30 (13.8%)	16 (8.6%)	142 (26.5%)

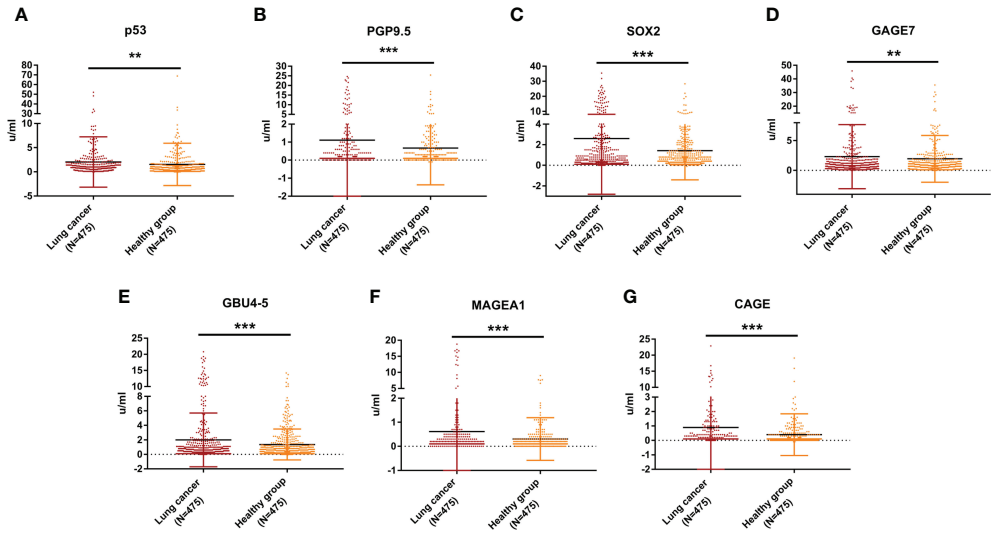


FIGURE 2
Expression levels of 7-AABs in the training set. **(A–G)** The serum concentration levels of p53, PGP9.5, SOX2,GAGE7, GBU4_5, MAGEA1 and CAGE in lung cancer group and healthy group in the training set, respectively. ** represents p<0.01; *** represents p<0.001.

u/ml, 11.32 u/ml, 15.72 u/ml, 8.52 u/ml, 3.28 u/ml and 5.76, respectively.

Determination of the diagnostic efficiency of the 7-AAB panel compared with CT scanning or PET-CT in the validation set

The baseline clinical characteristics of patients in the validation set were shown in Table 2. The serum concentration of p53, PGP9.5, SOX2, GAGE7, and CAGE were significantly higher in patients with malignant disease than those with benign disease (*P<0.05*, Table S2). Different from previous studies, our study observed a significant decrease in serum concentration of p53, PGP9.5, SOX2, GAGE7 and GBU4_5 between the postoperative group compared with the lung cancer group (*P<0.05*, Table S2). Therefore, we did not include the postoperative group in the subsequent diagnostic model analysis. In addition, serum concentrations of all 7-AABs except MAGEA1 were found significantly higher in patients with high-risk nodules who were still under radiographic follow-up but had not undergone surgery than in patients with benign disease (*P<0.05*, Table S2).

We compared the diagnostic efficiency of the single-use 7-AAB panel with CT diagnosis, PET-CT diagnosis, and a "both-positive rule" combination of 7-AABs and CT (shown in Figure 3). The result demonstrated the specificity of the 7-AAB panel was higher than both single-used CT diagnosis or PET-CT diagnosis (88.5% vs 28.0% vs 50.0% [*P<0.01*]). The positive predictive value (PPV) of the 7-AAB panel were higher than single-used CT diagnosis and similar to single-used PET-CT diagnosis (90.5% vs 56.8% [*P<0.01*], 90.5% vs 83.3% [*P=0.681*]). Moreover, when combined 7-AAB panel with CT diagnosis, the specificity and PPV reached a higher level (95.4% and 95.8%, respectively). However, the sensitivity of the 7-AAB panel were lower than CT diagnosis and PET-CT diagnosis (41.7% vs 65.7% vs 97.3%, [*P<0.05*]).

In addition, we conducted subgroup analyses to investigate the diagnostic efficiency of the 7-AAB panel in patients with different diameters of lung lesions, composition of radiographic nodule and histological types. In patients with different diameters of lung lesion (≤8mm, 8mm<φ≤20mm, >20mm), a similar trend was observed in all subgroups that the specificity and PPV of the 7-AAB panel were higher than both single-used CT diagnosis or PET-CT diagnosis (shown in Figure 3 & Table S3).

A nomogram for predicting the probability of malignant disease when combining the 7-AAB panel with patient clinical features in patients with radiological nodules

To further optimize diagnostic performance, a nomogram was established based on gender, age, smoking history, 7-AAB panel, and CT imaging characteristics (diameter, composition, spiculation sign, vessels sign and pleural indentation) of patients for predicting malignant disease (Figure 4A). Each factor shows a score according to the axis, and the scores of each factor can be added up to obtain an overall score, based on which the

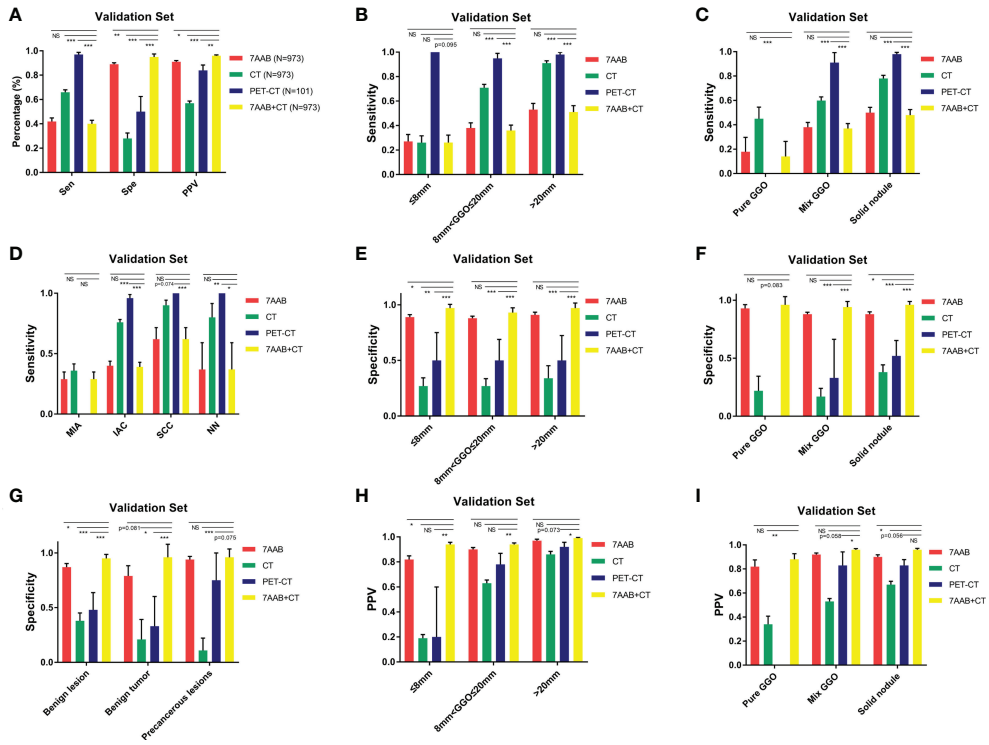


FIGURE 3
diagnostic efficiency of the 7-AAB panel comparing with CT scanning or PET-CT in different subgroups. **(A)** The sensitivity, specificity and PPV of four different diagnostic methods for diagnosing malignant disease in the validation set; **(B, E, H)** The sensitivity, specificity and PPV of four different diagnostic methods for diagnosing malignant disease in different lesion diameter subgroups, respectively; **(C, F, I)** The sensitivity, specificity and PPV of four different diagnostic methods for diagnosing malignant disease in different lesion composition subgroups, respectively; **(D)** The sensitivity of four different diagnostic methods for diagnosing malignant disease in different pathological type subgroups; **(G)** The specificity of four different diagnostic methods for excluding benign disease in different pathological type subgroups; NS represents no significant difference;* represents p<0.05; ** represents p<0.01; *** represents p<0.001.

probability of malignant disease can be predicted. The corresponding calibration curve of the nomogram is shown in Figure 4B and the corresponding ROC curve is shown in Figure 4C. The result indicated that this model has passed the calibration test (P=0.987>0.05) and the malignant disease probabilities predicted by the nomogram accorded well with the actual probability, with an AUC of 0.791, showing good discrimination ability. The diagnostic model achieved a sensitivity of 70.1% and specificity of 72.6%, significantly higher than either method used alone.

The predictive value of the 7-AAB panel for pathological and molecular characteristics of invasive adenocarcinoma and Squamous Cell Carcinoma patients

To investigate the predictive value of the 7-AAB panel in the evaluation of prognosis and subsequent treatment planning, we collected pathological and NGS testing results of pathologically confirmed IAC and SCC patients (shown in Table 3). Interestingly, we found that among IAC patients, 7-AAB panel-positive group showed higher proportion of EGFR mutations (81.1% vs 57.6%, P<0.001), compared with lower proportion of ERBB2 mutations (also known as EGFR2, 0.8% vs 6.2%, P=0.020) and ROS1 fusion mutations (0% vs 3.0%, P=0.052). At the same time, the 7-AAB positive group showed lower pathological differentiation degrees (P=0.018), more cases with specific pathological subtype (P=0.067), more advanced pathological stages (P=0.040) and higher Ki-67 indexes (39.15% ± 15.14% vs 26.83% ± 17.58%, P=0.011). To our surprise, a similar trend was not observed in patients with SCC.

Discussion

In this retrospective study, we investigated a 7-AAB panel for the diagnosis of NSCLC in a large scale of Chinese population (N=2824), 7-AAB panel test turns out to show

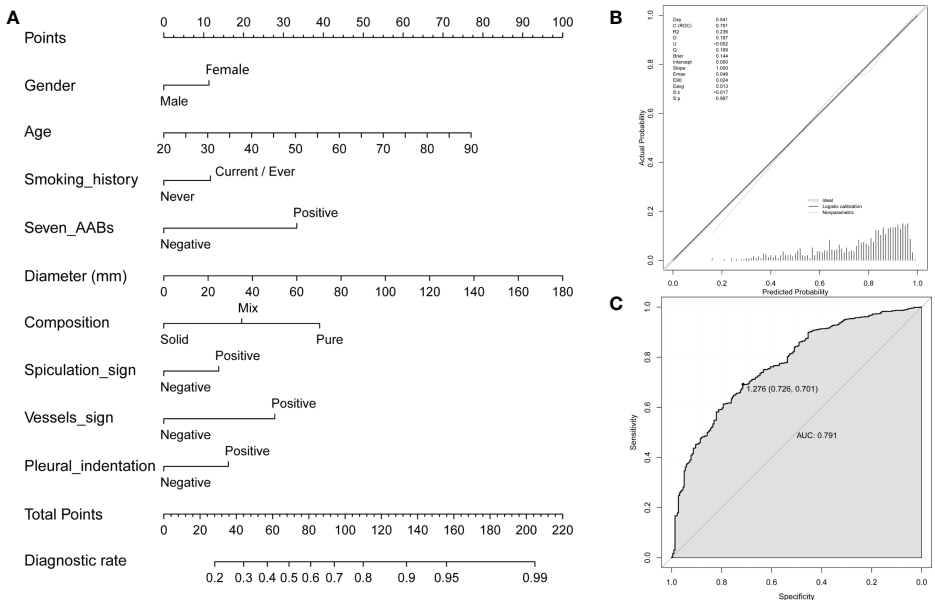


FIGURE 4
A nomogram for predicting probability of malignant disease. **(A)** A nomogram for predicting probability of malignant disease combining the 7-AAB panel with patient clinical features; **(B)** The calibration curve of the nomogram; **(C)** Receiver operating characteristic (ROC) curve and the corresponding area under the curve (AUC) using the nomogram to predict NSCLC in the training group.

superior specificity (88.5%) and PPV (90.5%) to CT scanning and PET-CT scanning for detecting NSCLC. In addition, diagnostic specificity and PPV were further improved when using the "both-positive rule" combination of 7-AABs and CT. This demonstrates the significant value of 7-AABs combined with CT scanning in the diagnosis of NSCLC. And this value has also been verified in lung lesions with different diameters and imaging compositions. Subsequently, we established a nomogram to predict the possibility of malignant disease based on basal information, 7-AAB panel test results, and CT scanning features of patients. Through ROC curve analysis, we proved that it has good predictive performance.

Although the popularization of LDCT screening makes more lung cancer patients diagnosed at an early stage, the low specificity of CT scanning also makes it difficult to distinguish between benign and malignant lesions, especially in patients with pure GGOs and solid pulmonary nodules (14, 15). This may lead to a large number of false-positive reports, which on the one hand occupies more medical resources, and on the other hand, multiple CT examinations also bring potential radiation exposure risks to patients. Therefore, there is an urgent need for a novel diagnostic method to supplement CT scanning in the detection of early lung cancer. Various biomarkers have been used to detect lung cancer, such as circulating tumor cells (CTCs) (16, 17), circulating tumor DNA (ctDNA) (18–20), microbial DNA (mbDNA) (21, 22), microRNA (23, 24), DNA methylation (25) and tumor autoantibodies (12, 26).

In several European studies of EarlyCDT-Lung, the seven-autoantibody panel (p53, NY-ESO-1, GBU4-5, CAGE, SOX2, HuD, and MAGE A4) were confirmed to have 88%-91% specificity and 11%-39% sensitivity (12, 27–29). At present, a number of studies on the use of a 7-AAB panels (p53, PGP9.5, SOX2, GAGE7, GBU4-5, MAGEA1 and CAGE) in the diagnosis of early lung cancer have also been carried out in China. In a prospective study, the sensitivity and specificity of the 7-AAB panel were 61% and 90%, respectively, which were considerably higher than for traditional biomarkers (including CEA, NSE, and CYFRA21-1) (30). In our previous small-scale retrospective study, the 7-AAB panel had a specificity of 90.2% and a PPV of 92.7%, significantly higher than CT scanning. In addition, the utility of the "both-positive rule" combination of 7-AABs and CT has the potential to avoid unnecessary follow-up (31).

In this study, we further increased the sample size and included multiple groups, including the lung cancer group, healthy participants, benign disease follow-up group, and post-operative group. We evaluated the diagnostic performance of the 7-AAB panel compared with CT/PET-CT in multiple dimensions. Compared with our previous study, we got similar results. Since more sample sizes were included in this study and a control group of healthy participants was set, it has a higher reference value. In addition, we redelimited the cutoff value in the training set of a larger sample, because the cutoff value previously used was determined by a researcher in a smaller sample (30).

10.3389/fonc.2022.1047019

TABLE 3 Baseline clinical characteristics of patients with positive or negative 7-AAB test results.

	IAC			SCC		
	7-AABs Positive (n=217)	7-AABs Negative (n=554)	P Value	7-AABs Positive (n=35)	7-AABs Negative (n=42)	P Value
EGFR mutation, n (%)	99 (81.1%)	152 (57.6%)	**<0.001**	0	0	–
19 exon Del	39 (39.4%)	69 (45.4%)		0	0	
21 exon L858R	57 (57.6%)	78 (51.3%)		0	0	
20 exon T790M	4 (4.0%)	5 (3.3%)		0	0	
Others	2 (2.0%)	3 (2.0%)		0	0	
ALK fusion, n (%)	1 (0.8%)	6 (2.3%)	0.320	0	0	–
ROS1 fusion, n (%)	0	8 (3%)	0.052	1 (20%)	1 (7.1%)	0.811
KRAS mutation, n (%)	8 (6.6%)	18 (6.8%)	0.925	0	0	–
BRAF mutation, n (%)	2 (1.6%)	5 (1.9%)	0.862	0	0	–
ERBB2 mutation, n (%)	1 (0.8%)	16 (6.1%)	**0.020**	0	0	–
MET mutation, n (%)	1 (0.8%)	6 (2.3%)	0.318	0	0	–
RET fusion, n (%)	2 (1.6%)	3 (1.1%)	0.684	0	1 (9.1%)	0.381
TP53 mutation, n (%)	17 (13.9%)	40 (15.2%)	0.756	6 (75%)	8 (72.7%)	0.912
PIK3CA mutation, n (%)	7 (5.7%)	8 (3%)	0.200	1 (12.5%)	4 (36.4%)	0.243
Differentiation degree, n (%)			**0.018**			0.661
High-middle	112 (48.1%)	271 (53.7%)		1 (2.8%)	2 (4.9%)	
Middle	74 (31.8%)	170 (33.7%)		9 (25%)	14 (34.1%)	
Low-middle	26 (11.2%)	45 (8.9%)		16 (44.4%)	13 (31.7%)	
Low	21 (9.0%)	19 (3.8%)		10 (27.8%)	12 (29.3%)	
Pleural invasion, n (%)	16 (6.9%)	30 (5.9%)	0.629	2 (5.6%)	4 (9.8%)	0.493
Vascular cancer embolus, n (%)	12 (5.2%)	12 (2.4%)	**0.048**	13 (36.1%)	8 (19.5%)	0.103
Spread through air spaces, n (%)	6 (2.6%)	16 (3.2%)	0.660	12 (33.3%)	8 (19.5%)	0.168
Specific pathological subtype, n (%)	56 (24.0%)	92 (18.2%)	0.067	0	0	–
Micropapillary predominant	24 (42.9%)	41 (44.6%)		0	0	
Solid predominant	25 (44.6%)	42 (45.7%)		0	0	
Cribriform acinar predominant	5 (8.9%)	7 (7.6%)		0	0	
Partial neuroendocrine differentiation	2 (3.6%)	2 (2.2%)		0	0	
Pathological stage, n (%)			**0.040**			0.330
IA	167 (77.0%)	439 (79.2%)		5 (13.9%)	11 (26.8%)	
IB	22 (10.1%)	51 (9.2%)		3 (8.3%)	6 (14.6%)	
IIA	8 (3.7%)	23 (4.2%)		1 (2.8%)	3 (7.3%)	
IIB	4 (1.8%)	21 (3.8%)		13 (36.1%)	15 (36.6%)	
IIIA	10 (4.6%)	13 (2.3%)		9 (25%)	3 (7.3%)	
IIIB	2 (0.9%)	5 (0.9%)		4 (11.1%)	3 (7.3%)	
IVA	4 (1.8%)	2 (0.4%)		1 (2.8%)	0	
Ki-67 positive (%)	39.15 ± 15.14	26.83 ± 17.58	**0.011**	64.17 ± 26.44	55.00 ± 23.21	0.158

Values in bold represent P values of less than 0.05, indicating statistical significance.

The 7-AAB panel has similar sensitivity and specificity used alone and has higher specificity and PPV when combining with CT scanning compared to the European Study of EarlyCDT-Lung. Even for solid nodules that were difficult to distinguish, we observed a specificity of 96.0% and a PPV of 96.0% when using the "both-positive rule". Moreover, a nomogram we established based on basal information, 7-AAB panel test results, and CT scanning features of patients can further improve the accuracy of a lung cancer diagnosis. The diagnostic model achieved a sensitivity of 70.1% and

specificity of 72.6%, significantly higher than either method used alone.

Interestingly, by comparing the 7-AABs positive group with the 7-AABs negative group, we found that the 7-AABs positive group was associated with a higher proportion of EGFR mutations and a lower proportion of ROS1 fusion and ERBB2 mutations in IAC patients. This suggests that 7-AABs testing may help screen out potential EGFR mutation patients, that is, patients who may potentially benefit from TKIs treatment and significantly improve their prognosis. One possible explanation is that the activation of EGFR downstream pathway caused by EGFR mutation may lead to the increased expression of some TAAs and the production of detectable autoantibodies. And many studies have confirmed the expression correlation or cascade reaction mechanism between EGFR and SOX2 (32), p53 (33), MAGE-A (34), CAGE (35). The 7-AABs positive group also had lower pathological differentiation degrees, more advanced pathological stages, and higher Ki-67 indexes, which indicate that the tumors in the 7-AABs positive group may have a higher degree of malignancy and a faster proliferation rate. As is known to all, lung cancer patients with EGFR mutation, pathologically low differentiation, advanced pathological stage, and high Ki-67 index have a worse prognosis (36–39), which also suggests the potential prognostic value of the 7-AAB panel test to some extent. A previous study showed that high expression of autoantibodies in lung cancer patients was positively associated with lymph node metastasis and distant metastasis (40). A follow-up study of 264 post-operative patients with NSCLC found that the autoantibody expression level was an independent predictor of poor prognosis (41). Another study of 157 patients with NSCLC showed a five-year survival rate of 62% for the overall population, compared with only 7.6% for those with positive autoantibodies (42).

Unfortunately, due to the short follow-up period, we do not collect enough prognostic data at present. In addition, surprisingly, we did not observe a similar phenomenon in patients with SCC, which may be due to tumor specificity of different pathologic types. Therefore, we will continue to follow up with patients in a long-term study in the future to obtain more prognostic data, to further investigate the relationship between 7AABs and prognosis. In addition, we will include more different pathological types of lung cancer, to better explore the diagnostic and prognostic value of 7-AAB for different types of lung cancer.

Conclusion

In this study, we verified the diagnostic strategy of a "both-positive rule" combination of 7-AABs and CT scanning in NSCLC that achieved a satisfactory specificity and PPV. In addition, our study developed and validated a novel nomogram based on the 7-AAB panel and CT signature for predicting the risk of NSCLC. Moreover, we revealed that lung adenocarcinoma patients with positive 7-AABs test had a higher ratio of EGFR mutation and worse pathologic features. Taken together, the 7-AABs panel test and our nomogram exhibited robust potential for the diagnosis of NSCLC in clinical practice.

Data availability statement

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

Author contributions

Conception and design: TH, ZW, and JH; Data collection: TH, ZW, PX, WW, HS; Analysis and interpretation of data: TH, ZW, PX, LY; Drafting of the manuscript: TH, ZW; Final approval of the manuscript: All authors. All authors contributed to the article and approved the submitted version.

Funding

This study was funded by Major science and technology special project of Zhejiang Province (2020C03058); Diagnosis and treatment technology research center of pulmonary neoplasm in Zhejiang Province (JBZX-202007) and A Project Supported by Scientific Research Fund of Zhejiang Provincial Education Department (Y202250800).

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.2022.1047019/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Schematic diagram of 7-AABs detection. (A) Procedure for 7-AABs detection by indirect ELISA assay; (B) Setting of patient's serum samples, loading control samples, calibration samples and quality control samples during 7-AAB panel testing.

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. Chansky K, Detterbeck FC, Nicholson AG, Rusch VW, Vallières E, Groome P, et al. IASLC Staging and Prognostic Factors Committee, Advisory Boards, and Participating Institutions. The IASLC Lung Cancer Staging Project: External Validation of the Revision of the TNM Stage Groupings in the Eighth Edition of the TNM Classification of Lung Cancer. *J Thorac Oncol* (2017) 12(7):1109–21. doi: 10.1016/j.jtho.2017.04.011

3. Oudkerk M, Liu S, Heuvelmans MA, Walter JE, Field JK. Lung cancer LDCT screening and mortality reduction - evidence, pitfalls and future perspectives. *Nat Rev Clin Oncol* (2021) 18(3):135–51. doi: 10.1038/s41571-020-00432-6

4. de Koning HJ, van der Aalst CM, de Jong PA, Scholten ET, Nackaerts K, Heuvelmans MA, et al. Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial. *N Engl J Med* (2020) 382(6):503–13. doi: 10.1056/NEJMoa1911793

5. Tan EM, Zhang J. Autoantibodies to tumor-associated antigens: reporters from the immune system. *Immunol Rev* (2008) 222:328–40. doi: 10.1111/j.1600-065X.2008.00611.x

6. Ehx G, Perreault C. Discovery and characterization of actionable tumor antigens. *Genome Med* (2019) 11(1):29. doi: 10.1186/s13073-019-0642-x

7. Finn OJ. Human tumor antigens yesterday, today, and tomorrow. *Cancer Immunol Res* (2017) 5(5):347–54. doi: 10.1158/2326-6066.CIR-17-0112

8. Li Y, Karjalainen A, Koskinen H, Hemminki K, Vainio H, Shnaidman M, et al. p53 autoantibodies predict subsequent development of cancer. *Int J Cancer* (2005) 114(1):157–60. doi: 10.1002/ijc.20715

9. Wang W, Zhong W, Chen C, Meng Q, Wei J. Circulating antibodies to linearpeptide antigens derived from ANXA1 and FOXP3 in lung cancer. *Anticancer Res* (2017) 37(6):3151–5. doi: 10.21873/anticanres.11673

10. Myšíková D, Adkins I, Hradilová N, Palata O, Šimonek J, Pozniak J, et al. Case-Control Study: Smoking History Affects the Production of Tumor Antigen-Specific Antibodies NY-ESO-1 in Patients with Lung Cancer in Comparison with Cancer Disease-Free Group. *J Thorac Oncol* (2017) 12(2):249–57. doi: 10.1016/j.jtho.2016.09.136

11. Yao X, Jiang H, Zhang C, Wang H, Yang L, Yu Y, et al. Dickkopf-1 autoantibody is a novel serological biomarker for non-small cell lung cancer. *Biomarkers* (2010) 15(2):128–34. doi: 10.3109/13547500903325662

12. Boyle P, Chapman CJ, Holdenrieder S, Murray A, Robertson C, Wood WC, et al. Clinical validation of an autoantibody test for lung cancer. *Ann Oncol* (2011) 22(2):383–9. doi: 10.1093/annonc/mdq361

13. Massion PP, Healey GF, Peek LJ, Fredericks L, Sewell HF, Murray A, et al. Autoantibody Signature Enhances the Positive Predictive Power of Computed Tomography and Nodule-Based Risk Models for Detection of Lung Cancer. *J Thorac Oncol* (2017) 12(3):578–84. doi: 10.1016/j.jtho.2016.08.143

14. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* (2011) 365(5):395–409. doi: 10.1056/NEJMoa1102873

15. Boiselle PM. Computed tomography screening for lung cancer. *JAMA* (2013) 309(11):1163–70. doi: 10.1001/jama.2012.216988

16. He Y, Shi J, Schmidt B, Liu Q, Shi G, Xu X, et al. Circulating Tumor Cells as a Biomarker to Assist Molecular Diagnosis for Early-Stage Non-Small Cell Lung Cancer. *Cancer Manag Res* (2020) 12:841–54. doi: 10.2147/CMAR.S240773

17. Lou J, Ben S, Yang G, Liang X, Wang X, Ni S, et al. Quantification of rare circulating tumor cells in non-small cell lung cancer by ligand-targeted PCR. *PloS One* (2013) 8(12):e80458. doi: 10.1371/journal.pone.0080458

18. Santarpia M, Liguori A, D'Aveni A, Karachaliou N, Gonzalez-Cao M, Daffinà MG, et al. Liquid biopsy for lung cancer early detection. *J Thorac Dis* (2018) 10(Suppl 7):S882–97. doi: 10.21037/jtd.2018.03.81

19. Vallée A, Marcq M, Bizieux A, Kouri CE, Lacroix H, Bennouna J, et al. Plasma is a better source of tumor-derived circulating cell-free DNA than serum for the detection of EGFR alterations in lung tumor patients. *Lung Cancer.* (2013) 82(2):373–4. doi: 10.1016/j.lungcan.2013.08.014

20. Moding EJ, Diehn M, Wakelee HA. Circulating tumor DNA testing in advanced non-small cell lung cancer. *Lung Cancer* (2018) 119:42–7. doi: 10.1016/j.lungcan.2018.02.019

21. Jin C, Lagoudas GK, Zhao C, Bullman S, Bhutkar A, Hu B, et al. Commensal Microbiota Promote Lung Cancer Development *via* γ δ T Cells. *Cell* (2019) 176 (5):998–1013. doi: 10.1016/j.cell.2018.12.040

22. Poore GD, Kopylova E, Zhu Q, Carpenter C, Fraraccio S, Wandro S, et al. Microbiome analyses of blood and tissues suggest cancer diagnostic approach. *Nature* (2020) 579(7800):567–74. doi: 10.1038/s41586-020-2095-1

23. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, et al. Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci U S A.* (2008) 105(30):10513–8. doi: 10.1073/pnas.0804549105

24. Ulivi P, Zoli W. miRNAs as non-invasive biomarkers for lung cancer diagnosis. *Molecules* (2014) 19(6):8220–37. doi: 10.3390/molecules19068220

25. Chen Z, Xiong S, Li J, Ou L, Li C, Tao J, et al. DNA methylation markers that correlate with occult lymph node metastases of non-small cell lung cancer and a preliminary prediction model. *Transl Lung Cancer Res* (2020) 9(2):280–7. doi: 10.21037/tlcr.2020.03.13

26. Murray A, Chapman CJ, Healey G, Peek LJ, Parsons G, Baldwin D, et al. Technical validation of an autoantibody test for lung cancer. *Ann Oncol* (2010) 21 (8):1687–93. doi: 10.1093/annonc/mdp606

27. González Maldonado S, Johnson T, Motsch E, Delorme S, Kaaks R. Can autoantibody tests enhance lung cancer screening?-an evaluation of EarlyCDT®-lung in context of the German lung cancer screening intervention trial (LUSI). *Transl Lung Cancer Res* (2021) 10(1):233–42. doi: 10.21037/tlcr-20-727

28. Sullivan FM, Mair FS, Anderson W, Armory P, Briggs A, Chew C, et al. Early Diagnosis of Lung Cancer Scotland (ECLS) Team. Earlier diagnosis of lung cancer in a randomised trial of an autoantibody blood test followed by imaging. *Eur Respir J* (2021) 57(1):2000670. doi: 10.1183/13993003.00670-2020

29. Borg M, Wen SWC, Nederby L, Hansen TF, Jakobsen A, Andersen RF, et al. Performance of the EarlyCDT® Lung test in detection of lung cancer and pulmonary metastases in a high-risk cohort. *Lung Cancer.* (2021) 158:85–90. doi: 10.1016/j.lungcan.2021.06.010

30. Ren S, Zhang S, Jiang T, He Y, Ma Z, Cai H, et al. Early detection of lung cancer by using an autoantibody panel in Chinese population. *Oncoimmunology* (2017) 7(2):e1384108. doi: 10.1080/2162402X.2017.1384108

31. Wang W, Zhuang R, Ma H, Fang L, Wang Z, Lv W, et al. The diagnostic value of a seven-autoantibody panel and a nomogram with a scoring table for predicting the risk of non-small-cell lung cancer. *Cancer Sci* (2020) 111(5):1699–710. doi: 10.1111/cas.14371

32. Jiang Y, Jiang YY, Xie JJ, Mayakonda A, Hazawa M, Chen L, et al. Co-activation of super-enhancer-driven CCAT1 by TP63 and SOX2 promotes squamous cancer progression. *Nat Commun* (2018) 9(1):3619. doi: 10.1038/s41467-018-06081-9

33. Jiao XD, Qin BD, You P, Cai J, Zang YS. The prognostic value of TP53 and its correlation with EGFR mutation in advanced non-small cell lung cancer, an analysis based on cBioPortal data base. *Lung Cancer* (2018) 123:70–5. doi: 10.1016/j.lungcan.2018.07.003

34. Sang M, Gu L, Yin D, Liu F, Lian Y, Zhang X, et al. MAGE-A family expression is correlated with poor survival of patients with lung adenocarcinoma: a retrospective clinical study based on tissue microarray. *J Clin Pathol* (2017) 70 (6):533–40. doi: 10.1136/jclinpath-2016-203718

35. Kim H, Kim Y, Goh H, Jeoung D. Histone deacetylase-3/CAGE axis targets EGFR signaling and regulates the response to anti-cancer drugs. *Mol Cells* (2016) 39(3):229–41. doi: 10.14348/molcells.2016.2244

36. Paez JG, Jänne PA, Lee JC, Tracy S, Greulich H, Gabriel S, et al. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* (2004) 304(5676):1497–500. doi: 10.1126/science.1099314

37. Cheung WK, Nguyen DX. Lineage factors and differentiation states in lung cancer progression. *Oncogene* (2015) 34(47):5771–80. doi: 10.1038/onc.2015.85

38. Jakobsen JN, Sørensen JB. Clinical impact of ki-67 labeling index in nonsmall cell lung cancer. *Lung Cancer* (2013) 79(1):1–7. doi: 10.1016/j.lungcan.2012.10.008

39. Xu J, Liu P, Da J, Hao J, Peng W, Sun G. Prognostic value of ki-67 in stage I non-small-cell lung cancer: A meta-analysis involving 1931 patients. *Pathol Res Pract* (2019) 215(5):855–60. doi: 10.1016/j.prp.2019.02.020

40. Li S, Ma Y, Xiong Y, Zhang P, Wang X, Wang Y, et al. Five tumor-associated autoantibodies expression levels in serum predict lung cancer and associate with poor outcome. *Transl Cancer Res* (2019) 8(4):1364–73. doi: 10.21037/tcr.2019.07.25

41. Ma L, Yue W, Teng Y, Zhang L, Gu M, Wang Y. Serum anti-CCNY autoantibody is an independent prognosis indicator for postoperative patients with early-stage nonsmall-cell lung carcinoma. *Dis Markers* (2013) 35(5):317–25. doi: 10.1155/2013/935943

42. Patel AJ, Tan TM, Richter AG, Naidu B, Blackburn JM, Middleton GW. A highly predictive autoantibody-based biomarker panel for prognosis in early-stage NSCLC with potential therapeutic implications. *Br J Cancer* (2022) 126(2):238–46. doi: 10.1038/s41416-021-01572-x

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 15 December 2022
DOI 10.3389/fonc.2022.1096449

Check for updates

OPEN ACCESS

EDITED BY
Hongda Liu,
Nanjing Medical University, China

REVIEWED BY
Jincheng Wang,
Hokkaido University, Japan
Dejun Wu,
Shanghai Pudong Hospital, China

*CORRESPONDENCE
Yuewei Chen
chenyueweicyw@163.com
Lei Zhou
150283304@qq.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

RECEIVED 12 November 2022
ACCEPTED 30 November 2022
PUBLISHED 15 December 2022

CITATION
Zeng Z, Yu J, Yang Z, Du K, Chen Y and Zhou L (2022) Investigation of M2 macrophage-related gene affecting patients prognosis and drug sensitivity in non-small cell lung cancer: Evidence from bioinformatic and experiments.
Front. Oncol. 12:1096449.
doi: 10.3389/fonc.2022.1096449

COPYRIGHT
© 2022 Zeng, Yu, Yang, Du, Chen 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.

Investigation of M2 macrophage-related gene affecting patients prognosis and drug sensitivity in non-small cell lung cancer: Evidence from bioinformatic and experiments

Zhen Zeng[1†], Jiachen Yu[2†], Zhuo Yang[3†], Kangming Du[4], Yuewei Chen[4*] and Lei Zhou[5*]

[1]Department of Emergency, Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, Sichuan, China, [2]Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing, China, [3]Department of Cardiology, Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, Sichuan, China, [4]Department of Cardiothoracic Surgery, Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, Sichuan, China, [5]Pain Management, Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu, Sichuan, China

Background: The progression process of lung cancer can be accelerated by M2 macrophages. However, genes that affect M2 macrophage polarization remain unidentified.

Methods: The Cancer Genome Atlas, Gene Expression Omnibus, and Arrayexpress databases were used to obtain open-access data. The analysis of public data was mostly performed with R studio. The RNA levels of specific genes were detected using quantitative real-time PCR. The proliferation ability of the cells was assessed by CCK8, colony formation, and EdU assays.

Results: Based on the multiple datasets, we noticed a poor prognosis in patients with high M2 macrophage infiltration. There were 114 genes differentially expressed between high and low M2 macrophages infiltrated samples, regarded as M2 macrophage-related genes. Subsequently, a prognosis prediction signature consisting of ABHD5, HS3ST2, TM6SF1, CAPZA2, LEPROT, HNMT, and MRO was identified and presented a satisfactory performance. The pathway enrichment results revealed a positive correlation between riskscore and enrichment scores for most immunotherapy-related positive terms. Also, there might be an increase in genomic instability among patients at high risk. Interestingly, low risk patients are most likely to benefit from PD-1 therapy, while high risk patients may benefit from CTLA-4 therapy. Meanwhile, the estimated IC50 of seven drugs differs significantly between two risk groups, including Cisplatin, Docetaxel, Doxorubicin, Gefitinib, Paclitaxel, Sunitinib and Vinorelbine.

Moreover, further experiments indicated that HNMT was overexpressed and can enhance the proliferation ability in lung cancer cells.

Conclusions: In summary, our study identified the molecules significantly affecting M2 macrophage infiltration and identified a prognosis signature that robustly indicated patients prognosis. Moreover, we validated the cancer-promoting effect of HNMT using *in vitro* experiments.

KEYWORDS

M2 macrophages, HNMT, model, immunotherapy, drug

Introduction

The world over, lung cancer is responsible for a disproportionate number of cancer related-deaths (1). Among the pathological types of lung cancer, non-small cell lung cancer (NSCLC) is the most prevalent, which consists of lung squamous cell carcinomas (LUSC) and lung adenocarcinomas (LUAD) (2). Lung cancer prevalence is often multifactorial, which brings difficulties to its prevention and treatment (3). For lung cancer patients with early-stage, surgical resection combined with chemoradiation may provide a better prognosis than chemotherapy alone, but their efficacy is still limited for the metastatic stage (4). Consequently, identifying effective molecular targets for diseases is imperative (5).

As research progresses, it is gradually understood that tumor occurrence, growth, and development are strongly influenced by the tumor microenvironment (6). Research has shown that tumor-infiltrating immune cells make up the majority of the microenvironment (7). Among these, macrophages might play a non-negligible role. Macrophages inside tumors have been defined as tumor-associated macrophages (TAMs). Generally, M1 macrophages hamper tumor development, whereas M2 macrophages contribute to tumor progression (8). Recently, more and more research is being devoted to understanding how TAMs work in specific tumors. In lung cancer, Xu and their colleagues demonstrated that the growth of lung cancer and metastasis can be inhibited by astragaloside IV through its modulation of macrophage M2 polarization through AMPK signaling (9). Wu and their colleagues showed that the succinate derived from cancer cells could contribute to macrophage polarization, further enhancing tumor metastasis through the succinate receptor (10). In the tumor microenvironment, macrophage M2 polarization was influenced by multiple factors. Meanwhile, the abnormal expression of specific genes could affect local tissue recruitment of TAMs, especially M2 macrophages. Exploration of the factors associated with macrophage M2 polarization could help us get an improved understanding of tumor progression and metastasis, allowing the identification of new targets for clinical therapy.

Bioinformatics can enhance people's understanding of diseases (11, 12). In our study, infiltration of M2 macrophages was observed as a cancer-promoting effect of NSCLC in several independent cohorts. Meanwhile, Next, we established a prognosis prediction signature based on seven M2 macrophage-related genes ABHD5, HS3ST2, TM6SF1, CAPZA2, LEPROT, HNMT, and MRO, which showed great prediction efficiency. Further, the potential difference in different risk groups was investigated, including pathway enrichment, and genomic mutation exploration. Interestingly, low risk patients are most likely to benefit from PD-1 therapy, while high risk patients may benefit from CTLA-4 therapy. Also, the estimated IC50 of seven drugs differs significantly between two risk groups, including Cisplatin, Docetaxel, Doxorubicin, Gefitinib, Paclitaxel, Sunitinib and Vinorelbine. Moreover, further experiments indicated that HNMT was overexpressed and can enhance the proliferation ability in lung cancer cells. Our study can improve the understanding of researchers on M2 macrophages in lung cancer.

Methods

Open-accessed data acquisition

NSCLC datasets with complete expression matrix data and clinical annotations have been searched exhaustively in public databases. Finally, six independent NSCLC cohorts were identified in our analysis, including TCGA-LUAD, TCGA-LUSC, E-GEOD-30219, GSE37745, GSE50081 and GSE68465. The expression profile was transcripts per kilo-base million (TPM) type. Using the reference file Homo_sapiens.GRCh38, probe annotation was conducted. GSE68465 (GPL96), GSE50081 (GPL570) (13), and GSE37745 (GPL570) were identified from the GSE database (14). E-GEOD-30219 (GPL570) was identified from the Arrayexpress database. Considering the same platforms of GSE50081, GSE37745, and E-GEOD-30219, the intra-batch and inter-batch effects of these

were corrected using the sva package. A standardization procedure was followed before data analysis (15).

Prognosis model establishment and validation

Following the identification of M2 macrophage-related DEGs, we screened for prognosis-related genes sequentially using univariate Cox analysis, LASSO regression, and multivariate Cox analysis. Finally, the prognosis model was established with the following formula: Riskscore = Σcoef*Exp(genes).

Nomogram plot, calibration curve and decision curve

Clinical features and riskscore of patients were combined to establish a nomogram. Meanwhile, the evaluation of the accuracy of the nomogram was conducted using the calibration curve and decision curve analysis (DCA).

Immune infiltration quantification and pathway enrichment

Based on a CIBERSORT algorithm, an evaluation of the microenvironment surrounding NSCLC tumors revealed 22 types of infiltrating immune cells (16). Biological investigation in different groups was conducted using the Gene Set enrichment analysis (GSEA) algorithms based on Hallmark, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) gene sets (17). Single sample gene set enrichment analysis (ssGSEA) was used to calculate the correlation between riskscore and specific pathway score (18).

Genomic instability analysis

Tumor mutation burden (TMB) measures how many base mutations are found in a 1Mb region of DNA. TMB was calculated and compared in different groups based on the data from TCGA. Using previously sorted data, the microsatellite instability (MSI) of NSCLC patients was assessed (19). The R package maftools were utilized to identify the mutated genes in different groups with statistically significant (20).

Sensitivity analysis of immunotherapy and chemotherapy

The sensibility of immunotherapy was quantified with the Tumor Immune Dysfunction and Exclusion (TIDE) algorithm. Also, the Subclass mapping algorithm was used to assess the genomic similarity between different risk patients and 47 immunotherapy-responding patients (21). Drug sensitivity analysis was performed using the Genomics of Drug Sensitivity in Cancer (GDSC) database (22).

Cell culture and quantitative real-time PCR

The BEAS-2B, H838, A549, H441 and H1299 were routine storage in the laboratory and cultured under standard conditions. A total RNA extraction kit was applied for RNA extraction. Processes of qRT-PCR were completed using the standard procedures. Primers used were as follows: HNMT, forward, 5'-GTTTGCTTGGCATAAGGAGACA-3', reverse, 5'-TGATCCGTACTTTTTCCACAGC-3', GAPDH, forward, 5'-GCAAATTCCATGGCACCGT-3', reverse, 5'-TCGCCCCACTTGATTTTGG-3'.

Cell proliferation assay

Evaluation of cell proliferation ability was conducted using the CCK8, colony formation and EdU assay according to the standard procedures (23).

Statistical analysis

All the statistical analysis was conducted using R software (version 4.0.0), SPSS (version 23.0) and GraphPad Prism 8. Briefly, the significance of the difference is determined by the p-value < 0.05. For continuous variables with normal distribution, the Student T test is used. Data that were not normally distributed were compared using Mann-Whitney U tests.

Results

Quantification of TAMs in NSCLC

Firstly, we quantified the immune microenvironment of the NSCLC tissue microenvironment, including TAMs (Figure 1A). Univariate Cox regression analysis indicated that in multiple independent NSCLC cohorts, M2 macrophages might exert a risk factor of patient OS, but M0 macrophages not (Figure 1B; M2 macrophages, E-GEOD, HR = 1.30, 95% CI = 1.02-1.78; GSE37745, HR = 1.36, 95% CI = 0.97-1.91; GSE50081, HR = 1.60, 95% CI = 1.02-2.52; GSE68465, HR = 1.25, 95% CI = 0.96-1.64; TCGA-LUAD, HR = 1.33, 95% CI = 0.99-1.77; TCGA-LUSC, HR = 1.48, 95% CI = 1.12-1.98, M0 macrophages, E-GEOD, HR = 1.23, 95% CI = 0.90-1.67; GSE37745, HR = 0.93,

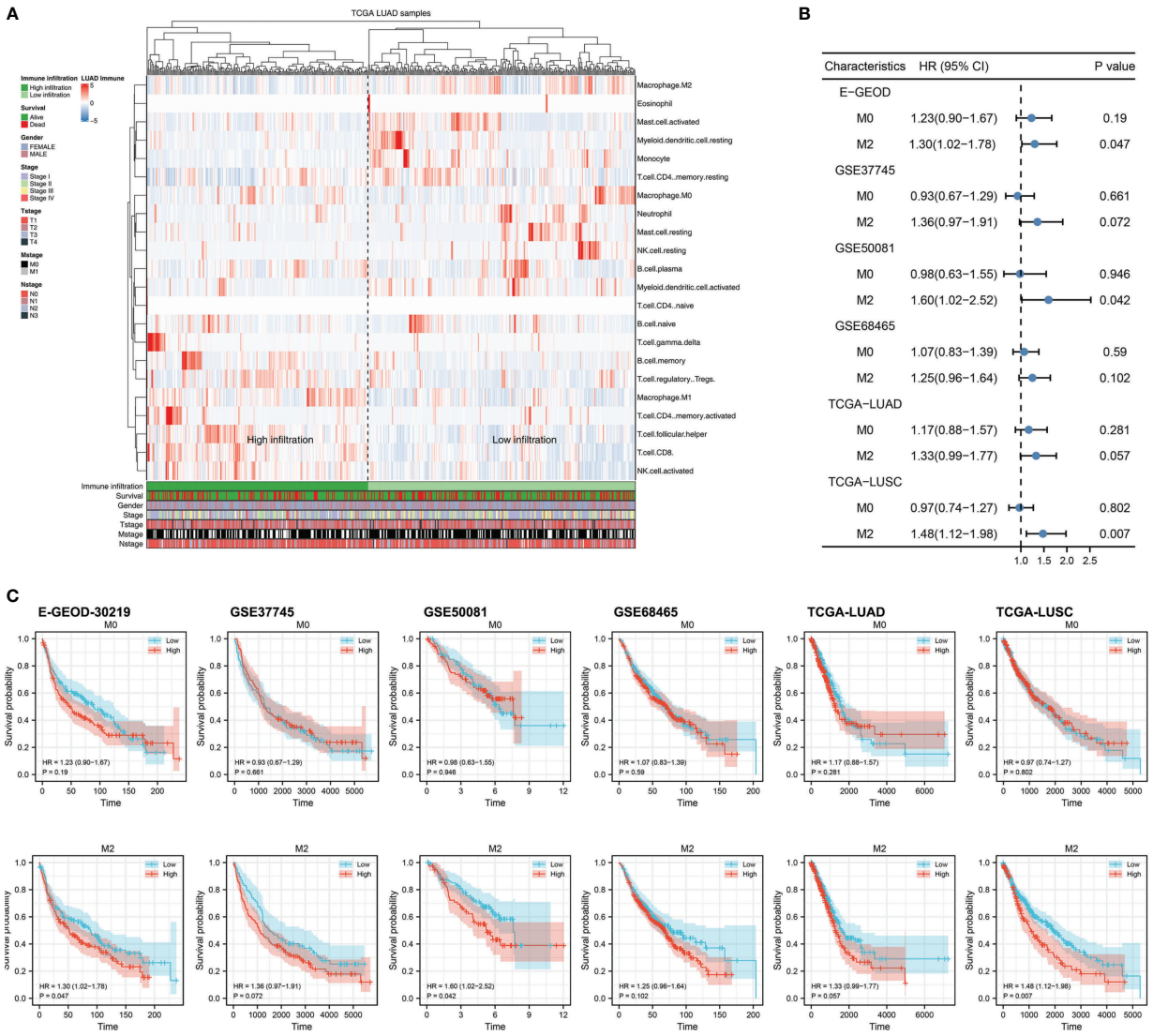


FIGURE 1
Exploration of M2 macrophage in lung cancer **(A)** The infiltration level of M2 macrophage was quantified using the CIBERSORT algorithm; **(B)** The prognosis correlation of M0 and M2 macrophage in multiple independent lung cancer cohorts; **(C)** KM survival curves of M0 and M2 macrophage in multiple independent lung cancer cohorts.

95% Cl = 0.67-1.29; GSE50081, HR = 0.98, 95% Cl = 0.63-1.55; GSE68465, HR = 1.07, 95% Cl = 0.83-1.39; TCGA-LUAD, HR = 1.17, 95% Cl = 0.88-1.57; TCGA-LUSC, HR = 0.97, 95% Cl = 0.74-1.27). The same trends of Kaplan-Meier (KM) survival curves was shown in Figure 1C.

Biological pathway effect of M2 macrophages in NSCLC

Figure 2A illustrated that M2 macrophages were positively correlated with monocyte and activated mast, but negatively correlated with Tregs, plasma B cells, CD8+ T cells, memory B cells, activated NK cells and follicular helper T cells. We then tried to combine the E-GEOD-30219, GSE37749, and GSE50081 into a large population cohort for the same platform. Batch differences between these cohorts were significant (Figure 2B). Using sva package, the batch effect of these three NSCLC cohorts was remarkably decreased (Figure 2C). The GSEA analysis revealed that in LUAD patients with high M2 macrophages infiltration, pathways of TGF-β signaling, apoptosis, P53 signaling, the epithelial-mesenchymal transition were significantly activated, yet the PI3K/AKT/mTOR signaling, G2M checkpoint, E2F target was downregulated (Figure 2D).

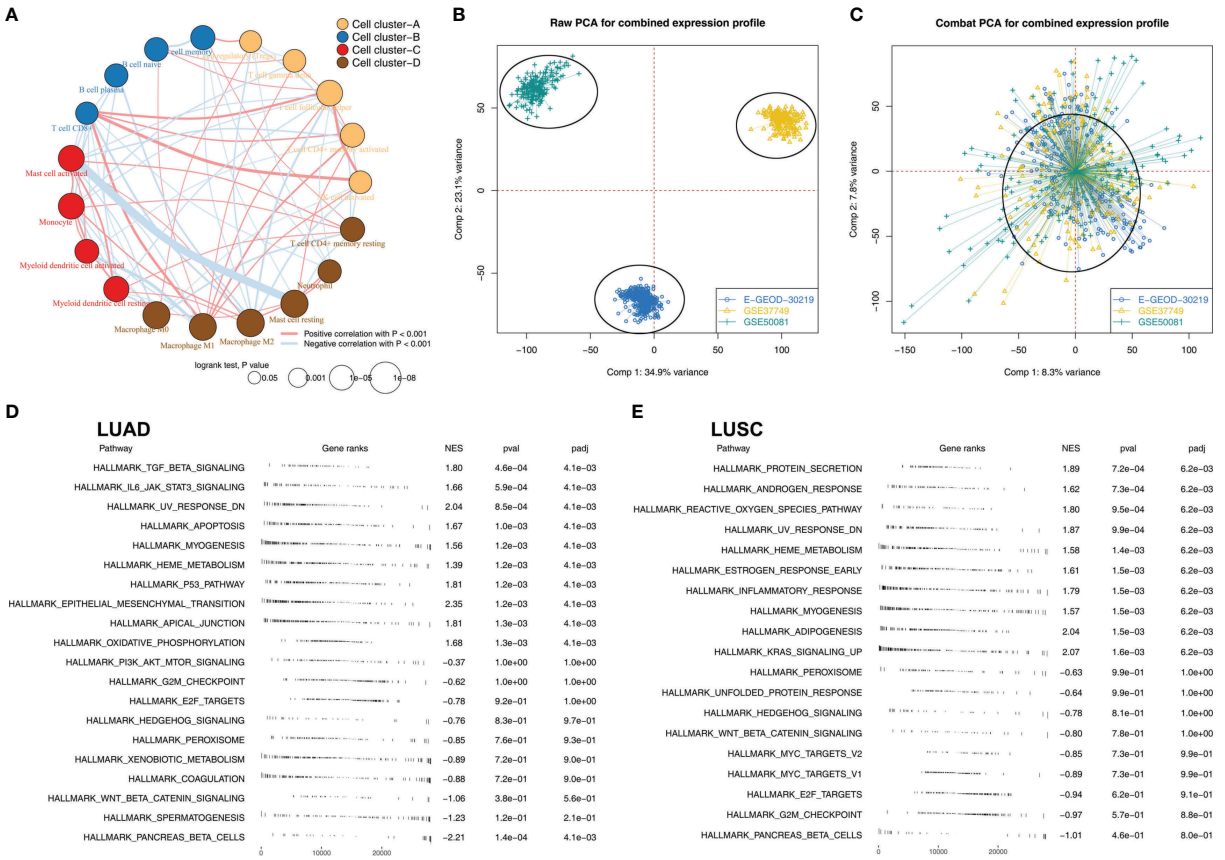


FIGURE 2
The biological role of M2 macrophages in lung cancer **(A)** Correlation analysis of the quantified 22 immune cells; **(B)** Significant batch effect was observed in E-GEOD-30219, GSE37749 and GSE50081; **(C)** Sva package was used for data combination; **(D)** GSEA analysis was performed to explore the biological pathway differences between high and low M2 macrophages LUAD samples; **(E)** GSEA analysis was performed to explore the biological pathway differences between high and low M2 macrophages LUSC samples.

In LUSC patients with high M2 macrophage levels, pathways of protein secretion, androgen response, and reactive oxygen species were significantly upregulated, yet the peroxisome, unfolded protein response, and hedgehog signaling was downregulated (Figure 2E). For GO analysis, in LUAD patients with high M2 macrophage infiltration, the terms of white fat cell differentiation, abnormal cardiac exercise stress test and sialic acid binding were activated (Figure S1A); in LUSC patients with M2 macrophage infiltration, the terms of neurotransmitter gated ion channel clustering, regulation of systemic arterial blood pressure by circulatory renin angiotensin and sialic acid binding were activated (Figure S1B). For KEGG analysis, in LUAD patients with high M2 macrophage infiltration, the terms of melanoma, renal cell carcinoma and leishmania infection were activated (Figure S1A); in LUSC patients with high M2 macrophage infiltration, the terms of leishmania infection, lysosome and cell adhesion molecules cams (Figure S2B).

Identification of M2 macrophages-related genes associated with patients prognosis

There were 114 genes differentially expressed between high and low M2 macrophages infiltrated samples, regarded as M2 macrophage-related genes (Figure 3A). Next, we aimed to identify a prognosis signature based on the M2 macrophage-related gene to robustly predict the patients OS. The TCGA-LUAD cohort was selected as the training cohort, and TCGA-LUSC and combined cohort (E-GEOD-30219 + GSE37749 + GSE50081) were used for validation. Univariate Cox regression analysis was firstly conducted to identify the molecules associated with patients OS with $P < 0.05$. A dimensionality reduction analysis was then conducted using the Lasso regression, and cross-validation was conducted 10 times (Figures 3B, C). Furthermore, multivariate Cox regression analysis indicated that ABHD5, HS3ST2, TM6SF1, CAPZA2, LEPROT, HNMT, and MRO were prominently associated with the risk of

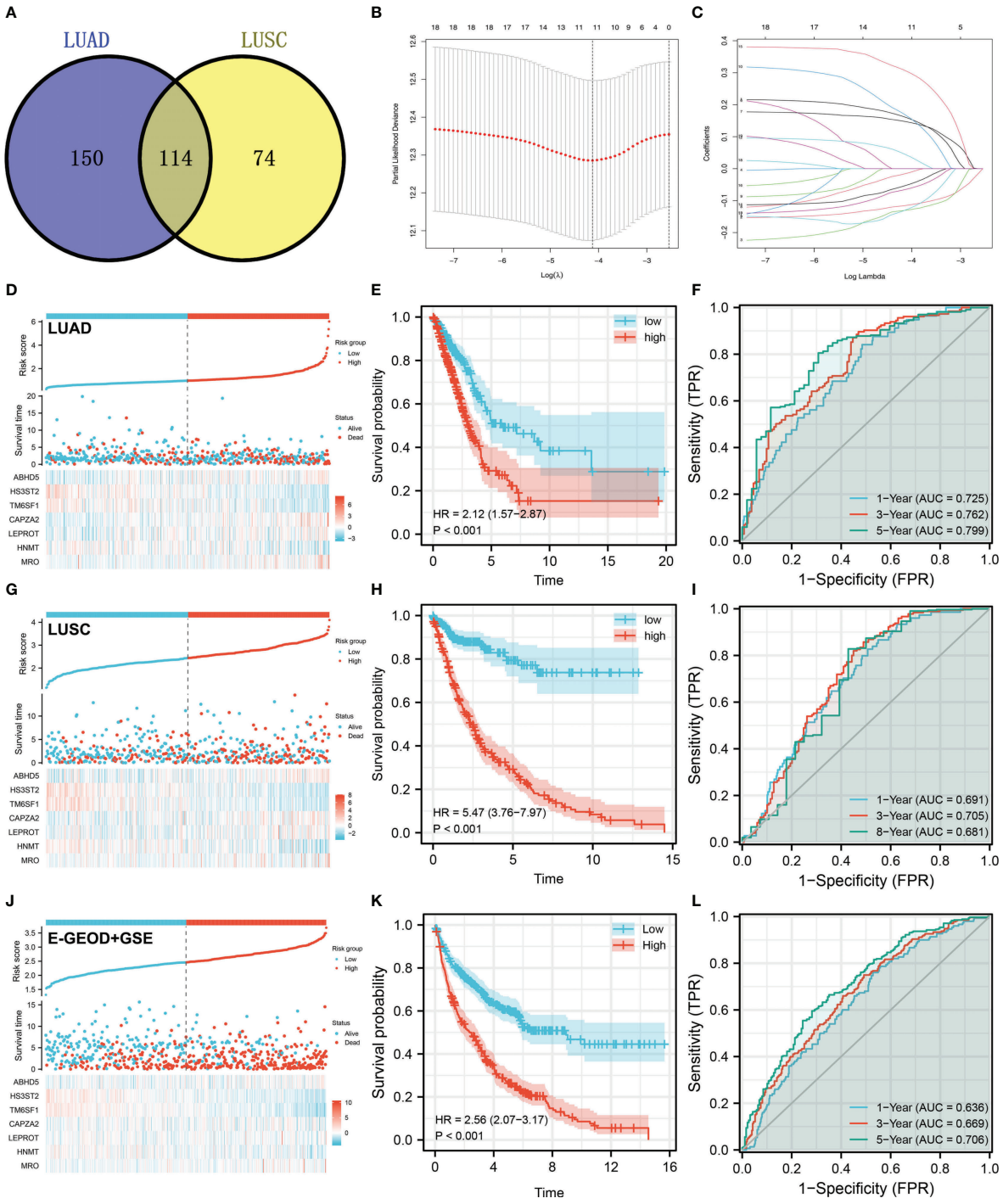


FIGURE 3
Prognosis model construction based on the M2 macrophage-related genes **(A)** A total of 114 genes were identified as M2 macrophage-related genes with the intersection of LUAD and LUSC data; **(B, C)** LASSO regression analysis was used for dimensionality reduction; **(D)** The overview of our prognosis model in the LUAD cohort; **(E)** KM survival curve was performed to explore the prognosis differences between high and low risk patients in LUAD cohort; **(F)** ROC curve was performed to evaluate the prediction efficiency of our model in LUAD cohort; **(G–I)** Model validation in LUSC cohort; **(J–L)** Model validation in E-GEOD + GSE cohort.

patients survival. These seven genes have been used to predict the survival rate of NSCLC patients with "riskscore = ABHD5 * 0.2406 + HS3ST2 * -0.1412 + TM6SF1 * -0.3109 + CAPZA2 * 0.1895 + LEPROT * 0.2919 + HNMT * -0.1554 + MRO * 0.3260". Training cohorts with high risk were observed to have a higher proportion of dead cases (Figure 3D). KM survival curves revealed that patients with higher riskscores might have a worse outcome (Figure 3E). According to the ROC curve, our model was highly effective at predicting patients outcomes (Figure 3F, AUC of 1-, 3- and 5-year were 0.725, 0.762 and 0.799). In the TCGA-LUSC cohort, the prognosis prediction efficacy is still good (Figures 3H, I, HR = 5.47, P < 0.001; AUC of 1-, 3- and 5-year were 0.691, 0.705 and 0.681). The same conclusion was also found in the E-GEOD + GSE cohort (E-GEOD-30219 + GSE37749 + GSE50081) (Figures 3J–L, HR = 2.56, P < 0.01; AUC of 1-, 3- and 5-year were 0.636, 0.669 and 0.706).

Development of a prognostic nomogram

Cox regression analysis was performed to further determine if our model could be a prognosis factor independent of traditional clinical features. Results of univariate Cox regression analysis demonstrated that some clinical features and riskscore were distinctively linked with patients survival (Figure 4A). Nevertheless, only riskscore remained an independent prognostic factor following multivariate Cox regression analysis (Figure 4B). Moreover, a nomogram was established that included five clinical variables and riskscore (Figure 4C). Based on the calibration plot, the data indicated good agreement between the real survival observation and the prediction for 1-, 3-, and 8 years (Figures 4D–F), and DCA analysis showed that the model with clinical features and riskscore had the best benefit to a treatment decision (Figure 4G). In addition, we assessed the significance of riskscore and seven model genes on correlations with clinicopathological parameters (Figures 4H–K). The result showed that ABHD5, CAPZA2, LEPROT and riskscore might be associated with worse clinical stage; ABHD5 and riskscore might be associated with worse T-classification, yet HS3ST2 was contrary; ABHD5, CAPZA2, LEPROT and riskscore might be associated with more progressive N-classification.

Comparative genomic analyses of the model

Riskscore was positively correlated with most immunotherapy-related terms, including mismatch repair, cell cycle and DNA replication (Figure 5A). Additionally, we examined the correlation between riskscores and Hallmark gene pathway scores, from which a strong linear correlation can be observed between riskscore and multiple oncogenetic pathways, including G2M checkpoint, glycolysis, E2F targets, DNA repair, mTORC1 signaling and PI3K/AKT/mTOR signaling (Figure 5A). KM analysis was conducted on patients with different levels of the seven model genes in TCGA-LUDA, TCGA-LUSC and E-GEOD+GSE cohorts (Figures 5B–D). The results showed that ABHD5, CAPZA2, LEPROT and MRO might be the risk factor of NSCLC, while HS3ST2, TM6SF1, and HNMT might be the protective factors. In addition, as shown in Figure 6A, we identified a relatively big number of nonsynonymous somatic mutations in both LUAD patients and LUSC patients. Then TMB score and MSI score for each patient were calculated, and we found that riskscore was positively correlated with the TMB score in both LUAD patients and LUSC patients (Figures 6B–E). Somatic mutation data of LUAD patients and LUSC patients were also analyzed, and we found that a higher somatic mutation including non-synonymous, synonymous mutations was enriched in high risk patients (Figures 6F–I). After maftools analysis, differential mutated genes with p < 0.05 were identified. TP53, PAPPA2, DNAH11, UBR4, POM121L12, TNR, and LRRIQ1 mutated more often in high risk LUAD group (Figure 6J), while ZBBX, TNN, CACNA1E, USH2A, DNAH5, BRINP3, DNAH10, PCDH15 and PRDM9 mutated more often in high risk LUSC group (Figure 6K).

Therapy prediction and potential drug identification

Immune checkpoints exert an important role in cancer immunotherapy. Results indicated that patients with high and low risk exhibited significant differences in immune checkpoint expression, indicating the underlying difference of immunotherapy response rate (Figure 7A). Meanwhile, TIDE score was calculated to predict the likelihood of response to immunotherapy (Figure 7B). Results revealed patients with low riskscore may respond better to immunotherapy (Figure 7C). Besides, the subclass mapping algorithm was also applied to investigate the genomic similarity between patients in two risk groups and the patients that responded to immunotherapies. Interestingly, low risk patients are most likely to benefit from PD-1 therapy, while high risk patients may benefit from CTLA-4 therapy (Figure 7D). Furthermore, the GDSC database was employed in our analysis to estimate the IC50 of twelve commonly used drugs between two risk groups. Finally, the estimated IC50 of seven drugs differs significantly between two risk groups, including Cisplatin, Docetaxel, Doxorubicin, Gefitinib, Paclitaxel, Sunitinib and Vinorelbine (Figure 7E).

HNMT enhances the proliferation ability of lung cancer

HNMT was identified for further investigation. The qRT-PCR of cell lines indicated that HNMT was overexpressed in

10.3389/fonc.2022.1096449

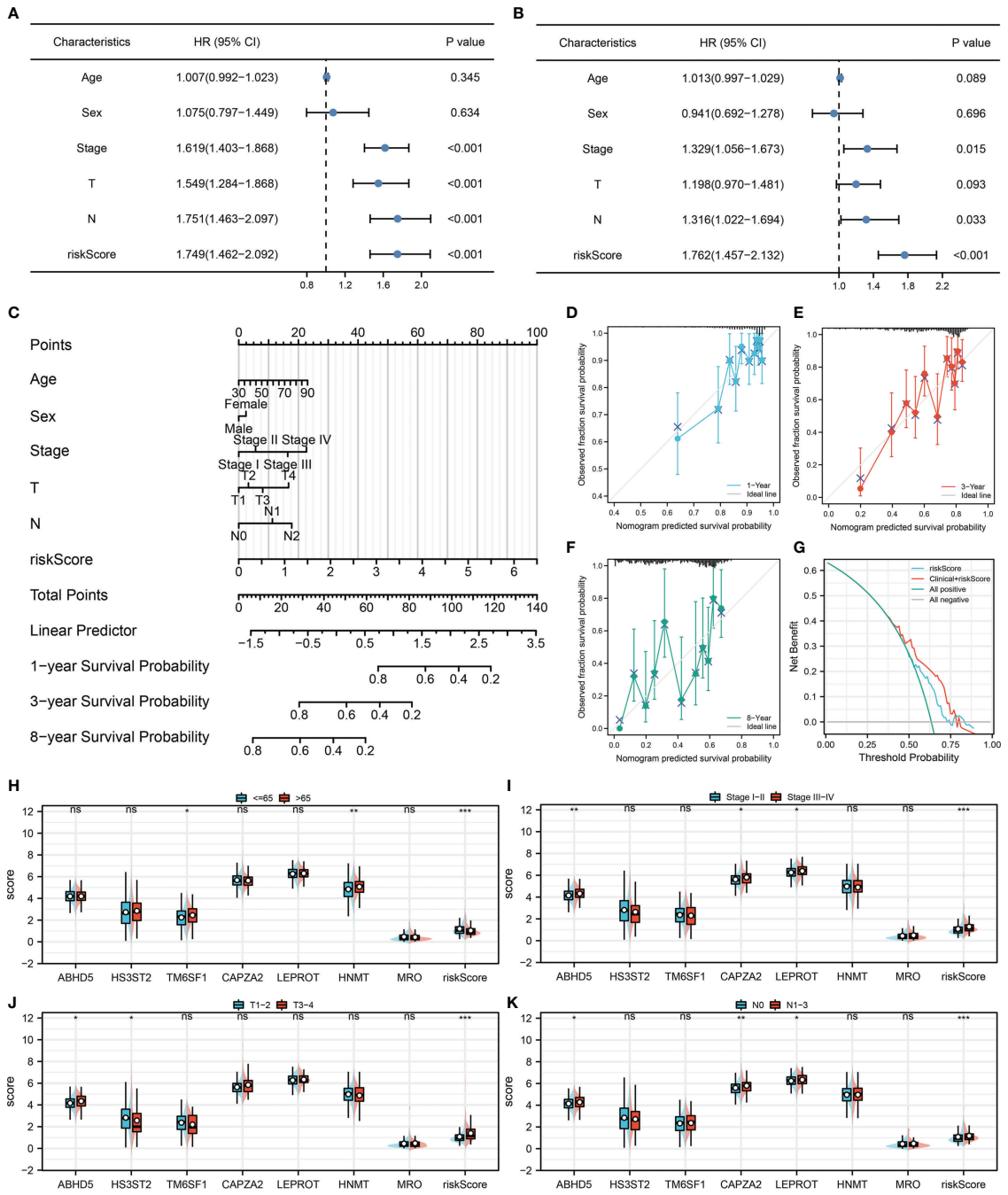


FIGURE 4
Nomogram and clinical correlation **(A, B)** Univariate and multivariate Cox regression analysis were performed to evaluate the independence of prognostic models; **(C)** Nomogram was constructed by combining the riskscore and clinical features; **(D–F)** Calibration curves of the nomogram; **(G)** DCA curve of the nomogram; **(H–K)** Clinical correlation of the model genes and riskscore.

lung cancer cells (Figure 8A). A satisfactory knockdown efficiency was presented in Figure 8B. CCK8 and colony formation assay revealed that the knockdown of HNMT can remarkably weaken the cell proliferation ability of lung cancer cells (Figures 8C–E). Moreover, a lower number of EdU-positive cells was observed in the cell with HNMT knockdown (Figure 8F). These results indicated that the HNMT can promote lung cancer proliferation.

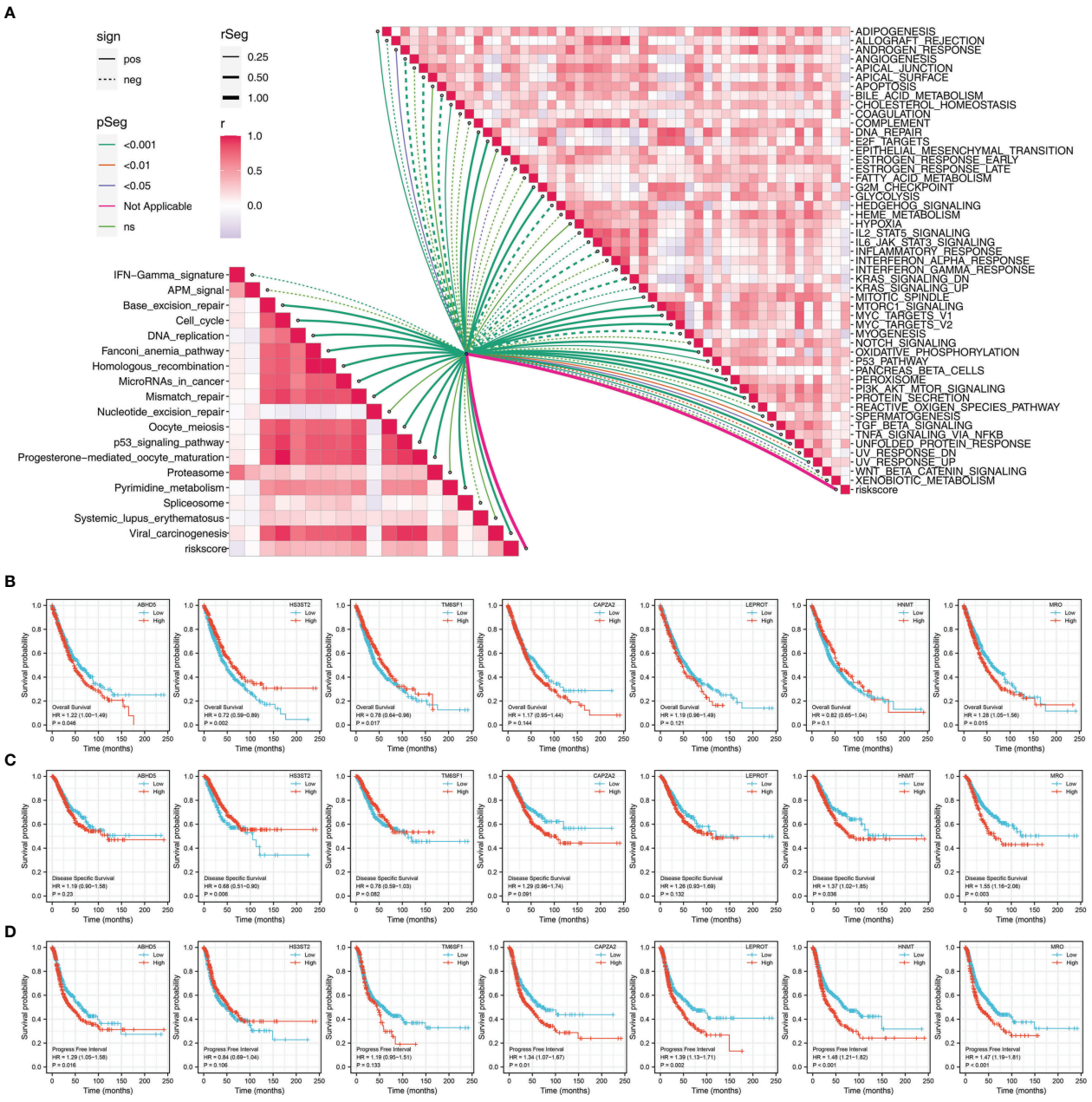


FIGURE 5
Pathway enrichment of the riskscore **(A)** Correlation of the riskscore and immune and metabolism pathways; **(B)** KM survival curves in OS of model genes; **(C)** KM survival curves in DSS of model genes; **(D)** KM survival curves in PFS of model genes.

Discussion

As of now, lung cancer continues to pose a major threat to global health. For most lung cancer patients at an early stage, surgery is the mainstay of treatment, and lobectomy is the preferred operation (24). However, it is noteworthy that most lung cancer patients are in an advanced stage when they receive their first diagnosis. Meanwhile, there remains controversy over the benefits of surgical therapy for lung cancer patients with advanced stage (25). Thus, exploration of the intrinsic mechanisms of NSCLC could help us identify novel diagnostic and therapeutic targets.

Macrophages could greatly affect cancer development and metastasis (26). The macrophage is both an antigen-presenting and immune cell. Macrophages are widely distributed and can specifically bind to tumors (27). In general, M2 macrophages

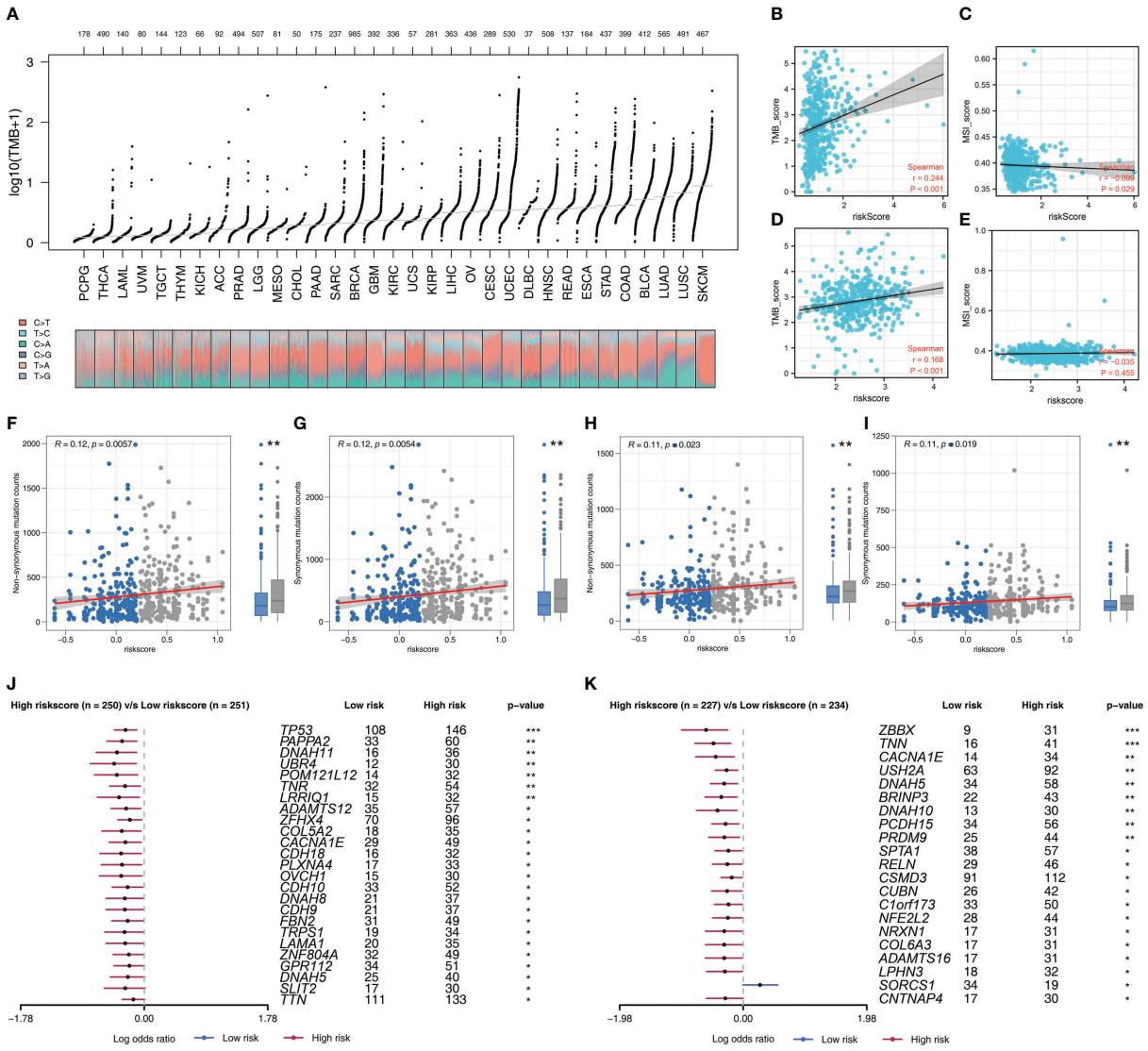


FIGURE 6
Genomic instability analysis **(A)** The overview of tumor mutation in pan-cancer; **(B, C)** The correlation of TMB and MSI score with riskscore in LUAD cohort; **(D, E)** The correlation of TMB and MSI score with riskscore in LUSC cohort; **(F, G)** The correlation of non-synonymous mutation counts and synonymous mutation counts with riskscore in LUAD cohort; **(H, I)** The correlation of non-synonymous mutation counts and synonymous mutation counts with riskscore in LUSC cohort; **(J)** The top mutated genes differentially existed in high and low risk LUAD patients; **(K)** The top mutated genes differentially existed in high and low risk LUSC patients.

play a cancer-promoting role in most malignancies, which could cause an immunosuppressive tumor microenvironment and are actively involved in cancer metastasis (28).

Here, we firstly explored the prognosis effect of M2 macrophages in NSCLC. Same with originally conceived, in multiple independent NSCLC cohorts, we noticed a poor prognosis in patients with high M2 macrophage infiltration. Next, we identified 114 M2 macrophage-related genes and established a prognosis model to predict patients OS based on seven genes, including ABHD5, HS3ST2, TM6SF1, CAPZA2, LEPROT, HNMT and MRO. KM survival curves and ROC survival curves revealed that our signature was reliable. Furthermore, results of biological enrichment showed that the pathway of DNA repair, G2M checkpoint, E2F targets, glycolysis, mTORC1 signaling and PI3K/AKT/mTOR signaling were aberrantly activated in the high risk group. Signaling pathways PI3K/AKT/mTOR, which is a classic pathway with a wide investigation, play a vital role in the proliferation and differentiation of cells (29). The G2/M checkpoint is the second checkpoint of the cell cycle and its abnormality of it might result in cycle disturbance (30). One of the main causes of cancer outbreaks is the change in DNA repair pathways. Meanwhile, compared with normal cells, tumor cells are

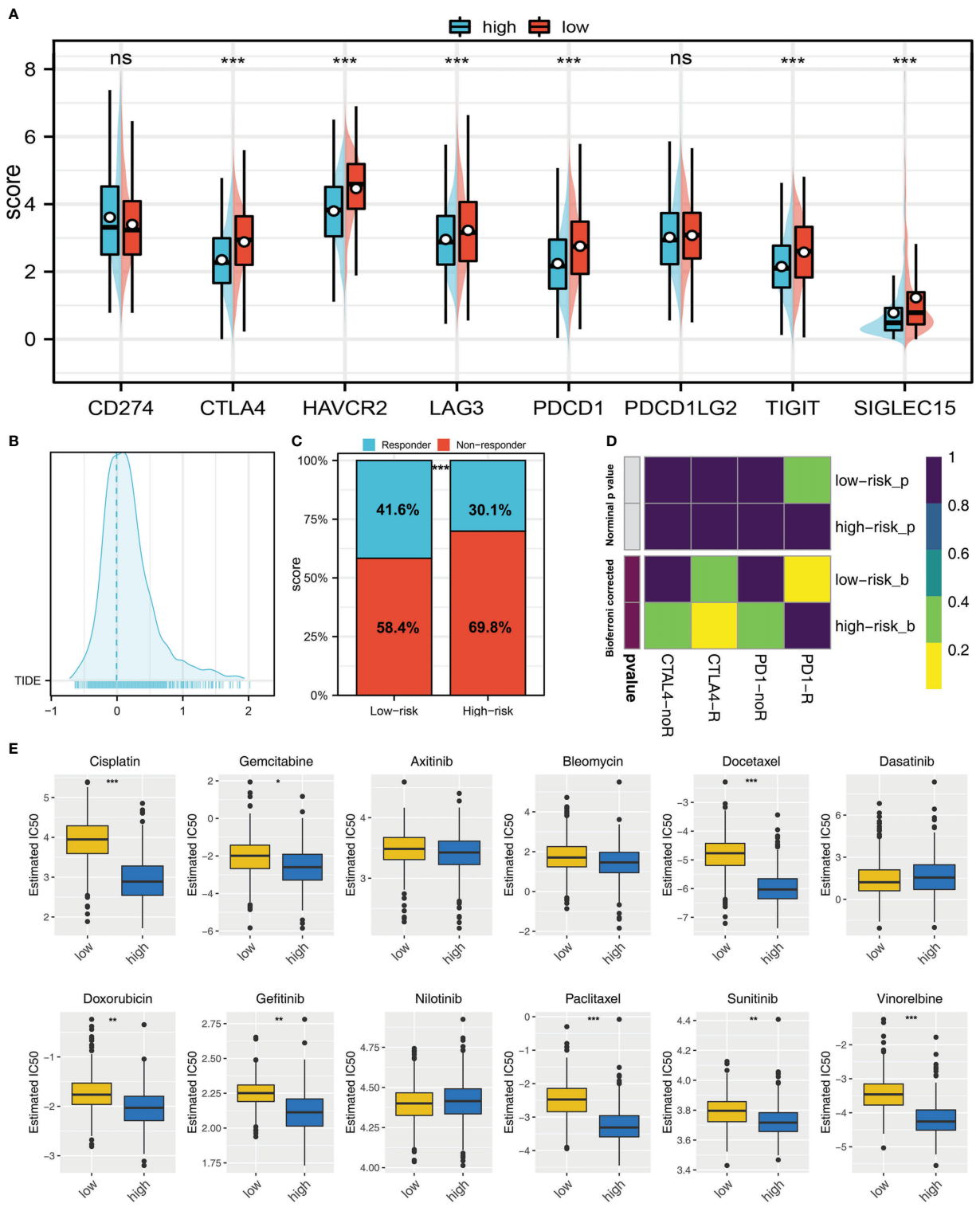


FIGURE 7
Immunotherapy and drug sensitivity **(A)** Important immune checkpoint expression in high and low risk patients; **(B)** The patients with TIDE score > 0 was regarded as non-responders and < 0 was regarded as responders; **(C)** A higher percentage of responders was observed in low risk group; **(D)** Submap algorithm showed that the low risk patients might be more sensitive to PD-1 therapy, while high risk patients might be more sensitive to CTLA-4 therapy; **(E)** The difference of chemotherapy sensitivity in high and low risk patients. ns = P > 0.05, * = P < 0.05, ** = P < 0.01, *** = P < 0.001.

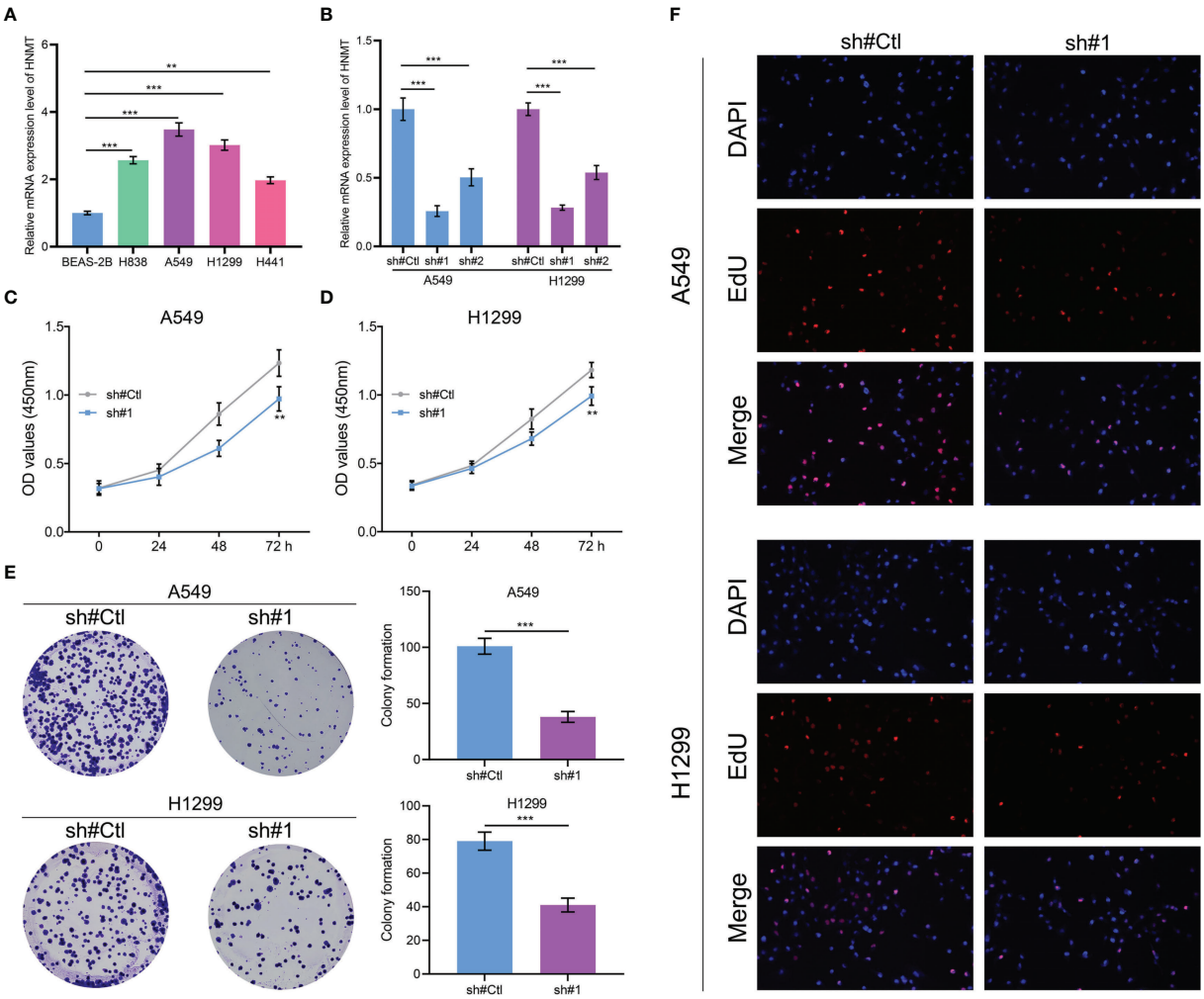


FIGURE 8
Role of HNMT in lung cancer **(A)** qRT-PCR was utilized to detect the RNA level of HNMT in lung cancer cells; **(B)** The knockdown efficiency of HNMT in lung cancer cells; **(C, D)** CCK8 assay was performed in control and HNMT knockdown cell; **(E)** Colony formation assay was performed in control and HNMT knockdown cell; **(F)** EdU assay was performed in control and HNMT knockdown cell. ** = P < 0.01, *** = P < 0.001.

more susceptible to DNA damage (31). These results showed that the aberrant activation of these oncogenic pathways might may result in a worse prognosis.

Our prognosis based on the ABHD5, HS3ST2, TM6SF1, CAPZA2, LEPROT, HNMT and MRO showed great prediction efficiency on patients OS. Meanwhile, these seven model genes were associated with higher M2 macrophage infiltration in NSCLC tissue. Liang and their colleagues found that cancer-derived exosomal TRIM59 could physically bind with ABHD5, further regulating macrophage and lung cancer progression (32). Hwang and their colleagues indicated that HS3ST2 had a high methylation signature in NSCLC cells, which could significantly lung cancer development (33). Zhong and their colleagues revealed that the TM6SF1 was related to the NSCLC tumor Microenvironment (34). Kuo and their colleagues found that the upregulation of HNMT could induce tumor stemness in NSCLC (35). MRO and CAPZA2 have not been reported in NSCLC. Our result showed that these genes are associated with M2 macrophages and might be the potential biomarker of NSCLC.

Immunotherapy, including PD-1/PD-L1 checkpoint blockade immunotherapy, has initiated a novel era of cancer treatment. Recently, a new computational method referred to as TIDE has been developed to model tumor immune evasion, demonstrating strong clinical utility for immunotherapy research (36). In our research, we first performed differential expression analysis on multiple important immune checkpoints (SIGLEC15, TIGIT, HAVCR2, PDCD1, CTLA4, LAG3, and PDCD1LG2) between two risk groups, and we found these immune checkpoints were expressed ubiquitously with high expression in high risk groups. In addition, TIDE analysis revealed that patients who responded to

immunotherapy accounted for more in the low risk group. Collectively, the subclass mapping algorithm was developed to evaluate similarities of expression matrix in responding to immunotherapies (37), showing the same results that immunotherapy was more effective in patients with low riskscore. All results suggested that immunotherapy efficacy could be predicted by our model. Moreover, the GDSC database was employed in our study, and we found the estimated IC50 of seven drugs differs significantly between two risk groups, including Cisplatin, Docetaxel, Doxorubicin, Gefitinib, Paclitaxel, Sunitinib and Vinorelbine, aiding clinicians in helping tailor therapy accordingly.

On the whole, our study identified the molecules significantly affecting M2 macrophage infiltration and identified a prognosis signature that robustly indicated patients prognosis. Moreover, we validated the cancer-promoting effect of HNMT using *in vitro* experiments. However, there are still several limitations that should be noted. Firstly, though a comprehensive search for public databases including appropriate expression matrix and clinical information was performed, further validation of our findings should be conducted in other cohorts. Secondly, M2 macrophages and genes related to them need to be studied further.

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, YJ and YZ contributed equally to this work. ZZ, ZY and KD conducted all the analysis. ZZ and JY wrote the manuscript. YC and LZ designed this work. 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/fonc.2022.1096449/full#supplementary-material

SUPPLEMENTARY FIGURE 1
GO analysis of M2 macrophages in lung cancer patients **(A)** GO analysis of M2 macrophage in LUAD patients; **(B)** GO analysis of M2 macrophage in LUSC patients.

SUPPLEMENTARY FIGURE 2
KEGG analysis of M2 macrophages in lung cancer patients **(A)** KEGG analysis of M2 macrophage in LUAD patients; **(B)** KEGG analysis of M2 macrophage in LUSC patients.

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. Nasim F, Sabath BF, Eapen GA. Lung cancer. *Med Clinics North America* (2019) 103(3):463–73. doi: 10.1016/j.mcna.2018.12.006

3. Mao Y, Yang D, He J, Krasna MJ. Epidemiology of lung cancer. *Surg Oncol Clinics North America* (2016) 25(3):439–45. doi: 10.1016/j.soc.2016.02.001

4. Hirsch FR, Scagliotti GV, Mulshine JL, Kwon R, Curran WJJr., Wu YL, et al. Lung cancer: Current therapies and new targeted treatments. *Lancet (London England)* (2017) 389(10066):299–311. doi: 10.1016/s0140-6736(16)30958-8

5. Yin Z, Wu D, Shi J, Wei X, Jin N, Lu X, et al. Identification of Aldh3a2 as a novel prognostic biomarker in gastric adenocarcinoma using integrated bioinformatics analysis. *BMC Cancer* (2020) 20(1):1062. doi: 10.1186/s12885-020-07493-x

6. Gu X, Wei S, Li Z, Xu H. Machine learning reveals two heterogeneous subtypes to assist immune therapy based on lipid metabolism in lung adenocarcinoma. *Front Immunol* (2022) 13:1022149. doi: 10.3389/fimmu.2022.1022149

7. Gajewski TF, Schreiber H, Fu YX. Innate and adaptive immune cells in the tumor microenvironment. *Nat Immunol* (2013) 14(10):1014–22. doi: 10.1038/ni.2703

8. Wu K, Lin K, Li X, Yuan X, Xu P, Ni P, et al. Redefining tumor-associated macrophage subpopulations and functions in the tumor microenvironment. *Front Immunol* (2020) 11:1731. doi: 10.3389/fimmu.2020.01731

9. Xu F, Cui WQ, Wei Y, Cui J, Qiu J, Hu LL, et al. Astragaloside iv inhibits lung cancer progression and metastasis by modulating macrophage polarization through ampk signaling. *J Exp Clin Cancer Res CR* (2018) 37(1):207. doi: 10.1186/s13046-018-0878-0

10. Wu JY, Huang TW, Hsieh YT, Wang YF, Yen CC, Lee GL, et al. Cancer-derived succinate promotes macrophage polarization and cancer metastasis *Via* succinate receptor. *Mol Cell* (2020) 77(2):213–27.e5. doi: 10.1016/j.molcel.2019.10.023

11. Wei Y, Chen X, Ren X, Wang B, Zhang Q, Bu H, et al. Identification of Mx2 as a novel prognostic biomarker for sunitinib resistance in clear cell renal cell carcinoma. *Front Genet* (2021) 12:680369. doi: 10.3389/fgene.2021.680369

12. Wu D, Yin Z, Ji Y, Li L, Li Y, Meng F, et al. Identification of novel autophagy-related lncrnas associated with a poor prognosis of colon adenocarcinoma through bioinformatics analysis. *Sci Rep* (2021) 11(1):8069. doi: 10.1038/s41598-021-87540-0

13. Der SD, Sykes J, Pintilie M, Zhu CQ, Strumpf D, Liu N, et al. Validation of a histology-independent prognostic gene signature for early-stage, non-Small-Cell lung cancer including stage ia patients. *J Thorac Oncol Off Publ Int Assoc Study Lung Cancer* (2014) 9(1):59–64. doi: 10.1097/jto.0000000000000042

14. Botling J, Edlund K, Lohr M, Hellwig B, Holmberg L, Lambe M, et al. Biomarker discovery in non-small cell lung cancer: Integrating gene expression profiling, meta-analysis, and tissue microarray validation. *Clin Cancer Res an Off J Am Assoc Cancer Res* (2013) 19(1):194–204. doi: 10.1158/1078-0432.Ccr-12-1139

15. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinf (Oxford England)* (2012) 28(6):882–3. doi: 10.1093/bioinformatics/bts034

16. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* (2015) 12 (5):453–7. doi: 10.1038/nmeth.3337

17. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci United States America* (2005) 102(43):15545–50. doi: 10.1073/pnas.0506580102

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

19. Bonneville R, Krook MA, Kautto EA, Miya J, Wing MR, Chen HZ, et al. Landscape of microsatellite instability across 39 cancer types. *JCO Precis Oncol* (2017) 2017:PO.17.00073. doi: 10.1200/po.17.00073

20. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: Efficient and comprehensive analysis of somatic variants in cancer. *Genome Res* (2018) 28 (11):1747–56. doi: 10.1101/gr.239244.118

21. Lu X, Jiang L, Zhang L, Zhu Y, Hu W, Wang J, et al. Immune signature-based subtypes of cervical squamous cell carcinoma tightly associated with human papillomavirus type 16 expression, molecular features, and clinical outcome. *Neoplasia (New York NY)* (2019) 21(6):591–601. doi: 10.1016/j.neo.2019.04.003

22. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of drug sensitivity in cancer (Gdsc): A resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res* (2013) 41(Database issue): D955–61. doi: 10.1093/nar/gks1111

23. Hua Q, Jin M, Mi B, Xu F, Li T, Zhao L, et al. Linc01123, a c-Myc-Activated long non-coding rna, promotes proliferation and aerobic glycolysis of non-small cell lung cancer through mir-199a-5p/C-Myc axis. *J Hematol Oncol* (2019) 12 (1):91. doi: 10.1186/s13045-019-0773-y

24. Sihoe ADL. Video-assisted thoracoscopic surgery as the gold standard for lung cancer surgery. *Respirol (Carlton Vic)* (2020) 25 Suppl 2:49–60. doi: 10.1111/resp.13920

25. Shi Y, Sun Y, Yu J, Ding C, Wang Z, Wang C, et al. China Experts consensus on the diagnosis and treatment of advanced stage primary lung cancer (2016 version). *Asia-Pac J Clin Oncol* (2017) 13(1):87–103. doi: 10.1111/ajco.12608

26. Xia Y, Rao L, Yao H, Wang Z, Ning P, Chen X. Engineering macrophages for cancer immunotherapy and drug delivery. *Advanced materials (Deerfield Beach Fla)* (2020) 32(40):e2002054. doi: 10.1002/adma.202002054

27. Murray PJ, Wynn TA. Protective and pathogenic functions of macrophage subsets. *Nat Rev Immunol* (2011) 11(11):723–37. doi: 10.1038/nri3073

28. Lin Y, Xu J, Lan H. Tumor-associated macrophages in tumor metastasis: Biological roles and clinical therapeutic applications. *J Hematol Oncol* (2019) 12 (1):76. doi: 10.1186/s13045-019-0760-3

29. Polivka JJr., Janku F. Molecular targets for cancer therapy in the Pi3k/Akt/Mtor pathway. *Pharmacol Ther* (2014) 142(2):164–75. doi: 10.1016/j.pharmthera.2013.12.004

30. Smith HL, Southgate H, Tweddle DA, Curtin NJ. DNA Damage checkpoint kinases in cancer. *Expert Rev Mol Med* (2020) 22:e2. doi: 10.1017/erm.2020.3

31. Lodovichi S, Cervelli T, Pellicioli A, Galli A. Inhibition of DNA repair in cancer therapy: Toward a multi-target approach. *Int J Mol Sci* (2020) 21(18):6684. doi: 10.3390/ijms21186684

32. Liang M, Chen X, Wang L, Qin L, Wang H, Sun Z, et al. Cancer-derived exosomal Trim59 regulates macrophage Nlrp3 inflammasome activation to promote lung cancer progression. *J Exp Clin Cancer Res CR* (2020) 39(1):176. doi: 10.1186/s13046-020-01688-7

33. Hwang JA, Kim Y, Hong SH, Lee J, Cho YG, Han JY, et al. Epigenetic inactivation of heparan sulfate (Glucosamine) 3-O-Sulfotransferase 2 in lung cancer and its role in tumorigenesis. *PLos One* (2013) 8(11):e79634. doi: 10.1371/journal.pone.0079634

34. Zhong H, Wang J, Zhu Y, Shen Y. Comprehensive analysis of a nine-gene signature related to tumor microenvironment in lung adenocarcinoma. *Front Cell Dev Biol* (2021) 9:700607. doi: 10.3389/fcell.2021.700607

35. Kuo KT, Lin CH, Wang CH, Pikatan NW, Yadav VK, Fong IH, et al. Hnmt upregulation induces cancer stem cell formation and confers protection against oxidative stress through interaction with Her2 in non-Small-Cell lung cancer. *Int J Mol Sci* (2022) 23(3):1663. doi: 10.3390/ijms23031663

36. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med* (2018) 24(10):1550–8. doi: 10.1038/s41591-018-0136-1

37. Hoshida Y, Brunet JP, Tamayo P, Golub TR, Mesirov JP. Subclass mapping: Identifying common subtypes in independent disease data sets. *PLos One* (2007) 2 (11):e1195. doi: 10.1371/journal.pone.0001195

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 23 December 2022
DOI 10.3389/fonc.2022.1090544

Check for updates

OPEN ACCESS

EDITED BY
Hongda Liu,
Nanjing Medical University, China

REVIEWED BY
Jing Yang,
Chongqing Technology and Business University, China
Xiaoliang Shao,
First People's Hospital of Changzhou, China

*CORRESPONDENCE
Chunhua Xu
xuch2188@163.com
Li Li
lili_nj86@163.com

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

SPECIALTY SECTION
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

RECEIVED 05 November 2022
ACCEPTED 07 December 2022
PUBLISHED 23 December 2022

CITATION
Liu W, Zhang Q, Zhang T, Li L and Xu C (2022) Minor histological components predict the recurrence of patients with resected stage I acinar- or papillary-predominant lung adenocarcinoma.
Front. Oncol. 12:1090544.
doi: 10.3389/fonc.2022.1090544

COPYRIGHT
© 2022 Liu, Zhang, Zhang, Li 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.

Minor histological components predict the recurrence of patients with resected stage I acinar- or papillary-predominant lung adenocarcinoma

Wei Liu[1,2†], Qian Zhang[1,2†], Tiantian Zhang[1,2], Li Li[1,2*] and Chunhua Xu[1,2*]

[1]Department of Respiratory Medicine, Affiliated Nanjing Brain Hospital, Nanjing Medical University, Nanjing, Jiangsu, China, [2]Clinical Center of Nanjing Respiratory Diseases and Imaging, Nanjing chest hospital, Jiangsu, China

Objective: Invasive lung adenocarcinoma is composed of five different histological subgroups with diverse biological behavior and heterogeneous morphology, the acinar/papillary-predominant lung adenocarcinomas are the most common subgroups and recognized as an intermediate-grade group. In the real world, clinicians primarily consider predominant patterns and ignore the impact of minor components in the prognosis of lung adenocarcinoma. The study evaluated the clinicopathologic characteristics of the lepidic, solid, and micropapillary patterns as non-predominant components and whether the minimal patterns had prognostic value on acinar/papillary-predominant lung adenocarcinomas.

Methods: A total of 153 acinar/papillary-predominant lung adenocarcinoma patients with tumor size ≤4 cm were classified into four risk subgroups based on the presence of lepidic and micropapillary/solid components: MP/S$^-$Lep$^+$, MP/S$^+$Lep$^+$, MP/S$^-$Lep$^-$, and MP/S$^+$Lep$^-$ groups. The Cox-proportional hazard regression model was used to assess disease-free survival (DFS).

Results: The risk subgroups based on the non-predominant patterns were associated with differentiation (P = 0.001), lymphovascular invasion (P = 0.001), and recurrence (P = 0.003). In univariate analysis, DFS was correlated with non-predominant components (P = 0.014), lymphovascular invasion (P = 0.001), carcinoembryonic antigen (CEA) (P = 0.001), and platelet-to-lymphocyte ratio (PLR) (P = 0.012). In the multivariate analysis, non-predominant components (P = 0.043) and PLR (P = 0.032) were independent prognostic factors for DFS. The 5-year survival rates of MP/S$^-$Lep$^+$, MP/S$^+$Lep$^+$, MP/S$^-$Lep$^-$ and MP/S$^+$Lep$^-$ subgroups were 93.1%,92.9%,73.1%,61.9%, respectively. The MP/S$^-$Lep$^+$ subgroup had the favorable prognosis than MP/S$^+$Lep$^-$ subgroup with a statistically significant difference (P = 0.002). As minor components, the

lepidic patterns were a protective factor, and the solid and micropapillary components were poor factors. The recurrence was related to the presence of non-predominant patterns rather than their proportion. Adjuvant chemotherapy did not significantly improve the prognosis of the MP/$S^{+}Lep^{-}$ subgroup (P = 0.839).

Conclusions: Regardless of the proportion, the presence of micropapillary/solid components and the absence of lepidic patterns are aggressive factors of DFS in patients with resected stage I acinar- or papillary-predominant lung adenocarcinoma.

KEYWORDS

lung adenocarcinoma, prognosis, stage I, non-predominant patterns, histological subtype

1 Introduction

Non-small cell lung accounts for 85% of lung cancer and the most common histological type of which is lung adenocarcinoma. Histological classification of lung adenocarcinoma was established by the International Association for the Study of Lung Cancer, American Thoracic Society and European Respiratory Society (IASLC/ATS/ERS), which is an international standard for histologic subclassification of lung adenocarcinoma (1). Lung adenocarcinoma is comprised of five pathological subtypes including lepidic, acinar, papillary, micropapillary, and solid, of which the acinar-predominant component is the most common subtype accounting for about 40-50%. The lepidic pattern is categorized as low grade, acinar and papillary patterns as an intermediate grade, and micropapillary and solid patterns as high grade (2, 3). Many studies have reported a negative prognosis for solid- and micropapillary-predominant adenocarcinoma and a beneficial prognosis for patients with lepidic-predominant adenocarcinoma (4, 5). The DFS rate of lepidic predominant adenocarcinomas and the acinar/papillary predominant adenocarcinomas was 99% and 80.8-82.4% respectively, micropapillary and solid predominant lung adenocarcinomas have a poor prognosis with the 5-year DFS rate of approximately 33.3-73.6% (6).

Interestingly, recent studies have shown the minimal presence of lepidic components plays a protective role and micropapillary (7, 8) and solid components are risk factors in early-stage lung adenocarcinoma (9, 10), even if not as predominant patterns. Studies have shown that solid and micropapillary components account for 41.8% and 60.4% of early stage lung adenocarcinoma, and even in stage IA, solid and micropapillary components are still unfavorable factors even though they are not the predominant components (11–13). When the most predominant histologic pattern was intermediate-grade, the second most predominant pattern was high-grade, and recurrence risk increased by 4.2-fold compared with the low-grade group (14). However, in most cases, the composition of lung adenocarcinoma is complex and heterogeneous, with few lung adenocarcinomas having a pure component and most adenocarcinomas (80-90%) having a mixture of two or more growth patterns. In the real world, clinicians primarily consider the predominant pattern and pay little attention to the effect of minimal components in the prognosis of lung adenocarcinoma. Liu. et al. found that the recurrence hazard curve in early stage adenocarcinoma patients showed a typical “double-peaked” pattern. The first recurrence peak occurred 20–22 months after surgery and the second peak occurred 5–6 years after surgery (15). The DFS is more informative than over survival (OS) in predicting the biological behavior in early stage tumors. We explored the clinicopathological features of the non-predominant components and focused on the impact of lepidic, micropapillary, and solid as minimal components in the recurrence of intermediate-grade adenocarcinoma.

2 Patients and methods

2.1 Patients

We enrolled 153 patients with clinical stage T1-2aN0M0 stage I lung adenocarcinoma who underwent complete resection at Nanjing Chest Hospital from May 2014 to August 2017, and the follow-up deadline was August 2022. The inclusion criteria were as follows: 1. anatomical resection with standard mediastinal Lymph node dissection; 2. The pathological stages were determined

according to the WHO eighth edition classification criteria; 3. The predominant subtype is acinar/papillary; 4. The minimal patterns include lepidic, solid, and micropapillary; Exclusion criteria were variants adenocarcinomas, adenocarcinomas *in situ* and minimally invasive adenocarcinomas, preoperative neoadjuvant chemotherapy or radiotherapy, positive surgical margin, a history of infection or other malignant tumors.

This study was approved by the Ethics Committee of Nanjing Brain Hospital. Informed consent was obtained from all participants in the study.

2.2 Histopathological evaluation

All resected specimens were formalin-fixed and stained with hematoxylin and eosin. The histologic classification of all slides was evaluated by two experienced pathologists according to IASLC/ATS/ERS classification criteria, respectively. The discrepancies were discussed to reach a consensus. Each histologic component present was recorded semiquantitatively in 5% increments. The non-predominant pattern was greater than or equal to 5% of the tumor. When the percentage of patterns was less than 5%, which was considered absent (16, 17). The predominant histological subtype was the highest proportion of the tumor, not necessarily 50% or greater (18, 19).

A total of 153 lung adenocarcinoma patients were classified into four groups based on the presence or absence of minimal patterns: MP/$S^{-}Lep^{+}$ group (micropapillary and solid components were absent, and lepidic components were present), MP/$S^{+}Lep^{+}$ group (either micropapillary or solid components were present, and lepidic components were present), MP/$S^{-}Lep^{-}$ group (both micropapillary, solid and lepidic components were absent) MP/$S^{+}Lep^{-}$ group (either micropapillary or solid components were present, and lepidic components were absent).

2.3 Methods

The histopathologic and clinicopathologic features of clinical data are as follows: gender, age, smoking history, surgical resection, predominant subtype, the proportion of micropapillary/solid components, the presence of lepidic and micropapillary/solid components, stage, differentiation, chemotherapy, laterality, visceral pleural invasion, lymphovascular invasion, DFS, OS, CEA, and PLR. The medical records and telephone were used for follow-up. A total of 153 patients were evaluated every 6 months for the first 2 years after surgery, and then once a year for the next 3-5 years. The last follow-up was in August 2022. The primary outcome was disease-free survival (DFS). DFS was calculated from the date of surgery to the date of first recurrence or death from any cause, or last contact. The definition of overall survival (OS) was the time interval from surgery to the date of death or last contact. The normal value for CEA was < 5ug/l.

2.4 Statistical analysis

The categorical variables were summarized as frequencies and percentages, and the difference was analyzed by the chi-square test. The optimal cut-off value of PLR was determined by Youden's Index and implemented through receiver operating characteristic (ROC). DFS was estimated by the Kaplan–Meier method and compared using the log-rank test. The Cox proportional hazards model was used to perform univariate and multivariate analyses. $P < 0.05$ values were considered statistically significant. The data were performed with SPSS 25.0 statistical software.

3 Results

3.1 The basic characteristics

The basic characteristics of the included patients were presented in Table 1. There were 55 females (35.9%) and 98 males (64.1%) with a median age of 60 years (31–84 years). Smokers were observed in 32 of 153 cases (20.9%). 146 patients (95.4%) underwent lobectomy and the remaining (n=7) underwent sublobar resection. Out of 153 patients, 88(57.5%) received postoperative adjuvant chemotherapy. Chemotherapy comprised platinum-based regimens, the platinum double-drug chemotherapy programs were: pemetrexed in 73 cases (83.0%), docetaxel in 5 cases (5.7%), gemcitabine in 2 cases (2.3%), paclitaxel in 2 cases (2.3%), and unknown regimens in 6 cases (6.7%). The percentage of stage IA1, IA2, IA3 and IB patients was 3.9%, 44.4%, 30.1%, and 21.6%, respectively. 58 tumors (37.9%) were located in the left lung. The visceral pleural invasion was observed in 26 cases (17.0%) and lymphovascular invasion was seen in 60 cases (39.2%). Poor differentiation was observed in 74 cases (49.3%). CEA ≥5 ng/ml was found in 28 cases (19.0%).

For the predominant subtype, the frequency of papillary predominant adenocarcinoma was 61.4%, followed by acinar predominant adenocarcinoma (38.6%). As a non-predominant pattern, lepidic components were displayed in 43 cases (28.1%), solid and micropapillary patterns in 98 tumors (64.1%) according to the new IASLC histologic grading system. The proportion of high-grade patterns greater than 5% or equal to 5% and less than 20% was observed in 50 cases (32.7%). The proportion of high-grade patterns greater than 20% or equal to 20% and less than 50% was observed in 48 cases (31.4%). The percentage of MP/$S^{-}Lep^{+}$, MP/$S^{+}Lep^{+}$, MP/$S^{-}Lep^{-}$ and MP/$S^{+}Lep^{-}$ group was 19.0%, 9.2%, 17.0%, and 54.9%, respectively. The median time to follow up was 55.3 months (interquartile

TABLE 1 Clinicopathological characteristics of patients with stage I lung cancer.

Variable	N (%)
Gender	
Male	55 (35.9)
Female	98 (64.1)
Age at diagnosis(years)	
≥65	44 (28.8)
<65	109 (71.2)
Smoking history	
Smoker	32 (20.9)
Nonsmoker	121 (79.1)
Surgical resection	
Sublobar resection	7 (4.6)
Lobectomy	146 (95.4)
Predominant subtybe	
acinar	59 (38.6)
papillary	94 (61.4)
The proportion of MP/S+	
0%	55 (35.9)
5%-20%	50 (32.7)
20%-50%	48 (31.4)
Lepidic	
Absent	110 (71.9)
Present	43 (28.1)
Stage	
IA1	6 (3.9)
IA2	68 (44.4)
IA3	46 (30.1)
IB	33 (21.6)
Differentiation	
Well	76 (50.7)
Poorly	74 (49.3)
Chemotherapy	
Yes	88 (57.5)
No	65 (42.5)
Laterality	
Left	58 (37.9)
Right	95 (62.1)
Visceral pleural invasion	
Absent	127 (83.0)
Present	26 (17.0)
Lymphovascular invasion	
Absent	93 (60.0)
Present	60 (39.2)
Recurrence	
Yes	42 (27.5)
No	111 (72.5)
Survival	
Yes	7 (4.8)
No	140 (95.2)
CEA(ng/ml)	
<5.00	119 (81.0)
≥5.00	28 (19.0)
PLR	
<154.13	128 (84.2)
≥154.13	24 (15.8)

MP/S$^+$, micropapillary/solid component were present; CEA, carcinoembryonic antigen; PLR, platelet-to-lymphocyte ratio.

range, 56.5 to 67 months). In the follow-up period, 27.5% of patients experienced recurrence and 4.8% of patients occurred death.

3.2 Optimal cut-off value for ROC curve

A cut-off value of 154.1 was used to discriminate between patients with high and low preoperative PLR, with AUC of 0.561 (sensitivity: 0.286, specificity: 0.892). Patients were distributed into PLR high groups (n = 24) and low PLR groups (n =128). PLR and CEA were combined for diagnosis, and the cut off value was 0.298 (sensitivity: 0.488; specificity: 0.783; AUC: 0.612), which can slightly enhance the diagnostic performance (Figure 1).

3.3 Relationships between clinicopathological characteristics and the risk subgroups based on non-predominant components

The relationship between the risk subgroup divided according to the minor components and clinicopathological features was

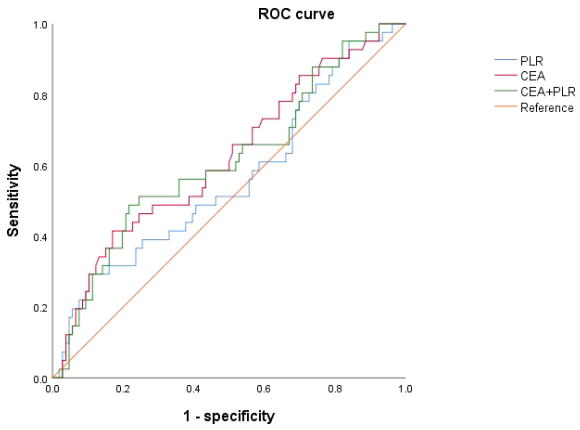


FIGURE 1
Receiver operating characteristic curves for analysis of using CEA and PLR in predicting disease-free survival.

shown in Table 2. The risk groups based on non-predominant components were associated with differentiation (P = 0.001). The MP/S$^+$Lep$^-$ group was vulnerable to poorly differentiation. The minor pattern was related to Lymphovascular invasion (P = 0.001). The MP/S$^+$Lep$^-$ group was more susceptible to Lymphovascular invasion. The minor pattern was significantly correlated with postoperative recurrence (P = 0.003). The 5-year DFS rates of MP/S$^-$Lep$^+$, MP/S$^+$Lep$^+$, MP/S$^-$Lep$^-$ and MP/S$^+$Lep$^-$ group were 93.1%, 92.9%, 73.1%, 61.9%, respectively. The survival time of MP/S$^-$Lep$^+$ group was longer than MP/S$^+$Lep$^-$ group with a statistically significant difference (P = 0.002) (Figure 2). It indicated the presence of micropapillary/solid components and absence of lepidic patterns are poor prognostic factors of DFS in patients with resected stage I acinar- or papillary-predominant lung adenocarcinoma.

In the patients without micropapillary/solid patterns, the lepidic components predicted a better prognosis compared with the non-lepidic components (P=0.042). The same result was seen in the patients with micropapillary/solid patterns (P = 0.031). The results suggested that lepidic components played a protective role in the early stage acinar- or papillary-predominant lung adenocarcinoma. In the patients with or without the lepidic components, the solid and micropapillary components showed no beneficial effect on prognosis (P = 0.976 and P = 0.276, respectively).

Nevertheless, we found the MP/S$^-$Lep$^-$ group had shorter DFS than MP/S$^+$Lep$^+$ group, but there was no statistically significant difference in the 5-year DFS rate between the two groups (P = 0.135), indicating there was no statistically significant difference in the rate of relapse when the high grade and low grade were both present or absent together. The minor components were related to aggressive factors such as poorly differentiation and lymphovascular invasion. The MP/S$^+$Lep$^-$ group was more vulnerable to recurrence compared with other subgroups.

3.4 The prognostic factors of disease-free survival in univariate analysis and multivariate analysis

Univariate analysis indicated that non-predominant patterns (P = 0.014), lymphovascular invasion (P = 0.001), elevated CEA level (P = 0.001), and elevated PLR level (P = 0.012) were the potential predictive factors for the stage I acinar- or papillary- predominant adenocarcinoma.

Multivariate Cox regression analysis further revealed that MP/S$^+$Lep$^-$ group (P = 0.043) and elevated PLR level (P = 0.032) were independent risk factors of DFS (Table 3).

3.5 Subgroup analysis of minimal components

We categorized acinar/papillary-predominant lung adenocarcinoma according to the presence of lepidic patterns. There were 28.1% (n = 43) patients in the group with the presence of lepidic patterns, the remaining was 71.9% (n =110) patients. The 5-year DFS rates of the group with the presence of lepidic patterns were 93.0% and the group with the absence of lepidic patterns was 64.5%. Therefore the lepidic pattern was a positive factor in stage I lung adenocarcinoma (Figure 3A). The non-predominant components were divided into two groups according to the presence of solid and micropapillary components: MP$^+$/S$^+$ subtype (either micropapillary or solid component was present) and MP$^-$&S$^-$ subtype (both micropapillary and solid components were absent). The MP$^+$/S$^+$ subtype was present in 63.3% (n = 97) of patients and MP$^-$&S$^-$ subtype was observed in 36.7% (n = 56) of patients. The 5-year DFS rates of MP$^+$/S$^+$ subtype and MP$^-$&S$^-$ subtype were 66.0% and 83.9%, respectively, and the difference was significant (P = 0.014). The solid and micropapillary patterns were associated with poor outcomes (Figure 3B).

The 2020 IASLC/ATS/ERS grading system proposed a cut-off of 20% for high-grade patterns as a risk factor (2). According to the proposal of the new IASA system, we classified the total proportion of solid and micropapillary components (TPSM) in tumors into two groups in our study: patients with 5-20% proportion of solid and micropapillary were defined as TPSM-low (n=50) and patients with 20%-50% proportion of solid and micropapillary were defined as TPSM-high (n = 48). The 5-year DFS rates in TPSM-low and TPSM-high were 68.0% and 64.6%, respectively. The result showed there was no significant difference between the two groups (P = 0.792) (Figure 3C). The presence of solid/micropapillary patterns rather than the

TABLE 2 Relationships between clinicopathological characteristics and the presence of lepidic and micropapillary/solid pathological patterns as minor components in patients with stage I lung adenocarcinoma.

				Risk Group			
Variables	N	MP/S^{-}Lep^{+} (%)	MP/S^{+}Lep^{+}(%)	MP/S^{-}Lep^{-} (%)	MP/S^{+}Lep^{-}(%)	χ^2	*P* value
Gender						2.994	0.393
Female	98	21 (21.4)	9 (9.2)	19 (19.4)	49 (50.0)		
Male	55	8 (14.5)	5 (9.1)	7 (12.7)	35 (63.3)		
Age at diagnosis (years)						4.813	0.186
≥65	44	8 (18.2)	6 (13.6)	10 (22.7)	20 (45.5)		
<65	107	21 (19.6)	6 (5.6)	16 (15.0)	64 (59.8)		
Smoking history						0.326	0.955
Smoker	32	5 (15.6)	3 (9.4)	6 (18.8)	18 (56.3)		
Nonsmoker	121	24 (19.8)	11 (9.1)	20 (16.5)	66 (54.5)		
Predominant subtype						7.002	0.072
acinar	59	15 (25.4)	7 (11.9)	5 (8.5)	32 (54.2)		
papillary	94	14 (14.9)	7 (7.4)	21 (22.3)	52 (55.3)		
Stage						7.406	0.060
IA	120	25 (20.8)	14 (11.7)	21 (17.5)	60 (50.0)		
IB	33	4 (12.1)	0 (0.0)	5 (15.2)	24 (72.7)		
Differentiation						38.241	**0.001**
Poorly	74	4 (5.4)	6(8.1)	5 (6.8)	59 (79.7)		
Well	76	23 (30.3)	8(10.5)	21 (27.6)	24 (31.6)		
Chemotherapy						3.953	0.267
Yes	88	13 (14.8)	7 (8.0)	14 (15.9)	54 (61.4)		
No	65	16 (24.6)	7 (10.8)	12 (18.5)	30 (46.2)		
Laterality						5.845	0.119
left	58	13 (22.4)	5 (8.6)	14 (24.1)	26 (44.8)		
Right	135	16 (11.9)	9 (6.7)	26 (19.3)	84 (62.2)		
Visceral pleural invasion						0.408	0.939
Absent	127	23 (18.1)	12 (9.4)	22 (17.3)	70 (55.1)		
Present	26	6 (23.1)	2 (7.7)	4 (15.4)	14 (53.8)		
Lymphovascular invasion						26.284	**0.001**
Absent	93	26 (28.0)	8 (8.6)	22 (23.7)	37 (39.8)		
Present	60	3 (5.0)	6 (10.0)	4 (6.7)	47 (78.3)		
Recurrence						13.834	**0.003**
Yes	42	2 (4.8)	1 (2.4)	7 (16.7)	32 (76.2)		
No	111	27 (24.3)	13 (11.7)	19 (17.1)	52 (46.8)		
Survival						3.020	0.389
Yes	7	0(0.0)	0 (0.0)	2 (28.6)	5 (71.4)		

(Continued)

TABLE 2 Continued

Variables	N	Risk Group				χ^2	*P* value
		MP/S$^-$Lep$^+$ (%)	MP/S$^+$Lep$^+$(%)	MP/S$^-$Lep$^-$ (%)	MP/S$^+$Lep$^-$(%)		
No	140	28 (20.0)	14 (10.0)	24 (17.1)	74 (52.9)		
CEA						5.078	0.166
Low	119	25 (21.0)	11 (9.2)	21 (17.6)	62 (52.1)		
High	28	3 (10.7)	2 (7.1)	2 (7.1)	21 (75.0)		
PLR						4.960	0.175
Low	128	23 (18.0)	14 (10.9)	23 (18.0)	68 (53.1)		
High	24	6 (25.0)	0 (0.0)	2 (8.3)	16 (66.7)		

CEA, carcinoembryonic antigen; PLR, platelet-to-lymphocyte ratio; MP/S$^-$Lep$^+$ (micropapillary/solid components were absent and lepidic components were present); MP/S$^+$Lep$^+$ (both micropapillary/solid component and lepidic component were present); MP/S$^-$Lep$^-$ (both micropapillary/solid component and lepidic component were absent); MP/S$^+$Lep$^-$ (micropapillary/solid component were present and lepidic component were absent). The bold values statistical significance at P < 0.05 level.

proportion of solid/micropapillary patterns affected the DFS in lung adenocarcinoma.

For Lung adenocarcinoma with MP$^+$/S$^+$ subtype, the DFS rates in patients with the presence of lepidic patterns and absence of those were 92.9% and 61.4%, respectively. The results demonstrated that even in patients with poor prognosis, the presence of lepidic patterns could improve survival significantly (P = 0.029) (Figure 3D). The published literature only took the solid and micropapillary components into account, or just considered the lepidic component alone. In the future, we should emphasize the inclusion of all three components simultaneously when assessing the prognosis of lung adenocarcinoma.

The MPS$^-$/Lep$^+$ group had the adverse outcome among those with acinar- or papillary-predominant lung adenocarcinoma, hence we explored whether these patients could benefit from adjuvant chemotherapy. In terms of the prognosis of patients in MPS$^-$/Lep$^+$ group, the DFS rate was no significant difference between postoperative chemotherapy and non-postoperative chemotherapy (P = 0.839) (Figure 4).

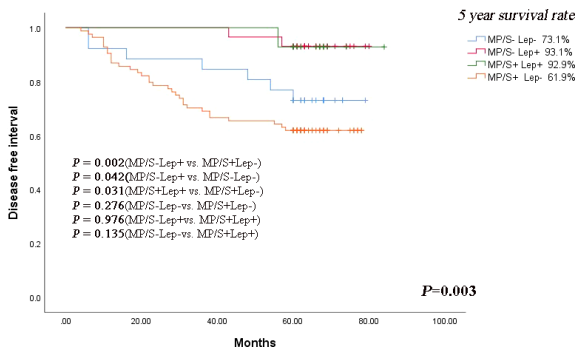


FIGURE 2
Kaplan-Meier curve of DFS according to lepidic, micropapillary/solid minor components in stage I lung adenocarcinoma patients. The comparison of DFS between 2 groups: MP/S$^-$Lep$^+$ vs. MP/S$^+$Lep$^-$ group (P = 0.002), MP/S$^-$Lep$^+$ vs. MP/S$^-$Lep$^-$ group (P=0.042), MP/S$^+$Lep$^+$ vs. MP/S$^+$Lep$^-$ group (P = 0.031), MP/S$^-$Lep$^-$ vs. MP/S$^+$Lep$^-$ group (P = 0.276), MP/S$^-$Lep$^+$ vs. MP/S$^+$Lep$^+$ group (P = 0.976), MP/S$^-$Lep$^-$ vs. MP/S$^+$Lep$^+$ group (P = 0.135).

4 Discussion

The MP/S$^+$Lep$^-$ and high PLR level were determined as poor risks in acinar/papillary-predominant lung adenocarcinoma patients with tumor size ≤4 cm. The MP/S$^+$Lep$^-$ group was prone to recurrence, which may be strongly associated with poorly differentiation and lymphovascular invasion. Adjuvant chemotherapy did not significantly improve DFS in the MP/S$^+$Lep$^-$ subgroup. The above results reflected that even for the same predominant histopathological subtype, the prognosis varied with the diverse composition of the minimal patterns. It is advised to concentrate on all growth patterns observed in tumors beyond predominant components. Zhao et al. also concluded not only predominant subtypes but also minor components had an important value in clinical outcomes (20). Mäkinen et al. reported that the non-predominant lepidic components were related to a favorable outcome in invasive adenocarcinoma (21). A study indicated solid minor components and solid predominant subtypes both predicted a worse prognosis compared with the absent solid pattern. Chemotherapy was beneficial for solid predominant components rather than solid as minor patterns (22). The MP/S$^+$Lep$^-$ was an independent factor for DFS and OS, and stage IA patients in the MP/S$^+$Lep$^-$ subgroup did not benefit from chemotherapy. Those findings are supported by our results (23).

We found that patients in the MP/S$^+$Lep$^-$ subgroup were closely connected with poorly differentiation and lymphovascular invasion. Perhaps this can be explained by the following theoretical mechanism. The solid/micropapillary

TABLE 3 Univariate and multivariate COX regression analysis for disease-free survival in stage I lung adenocarcinoma.

	Disease Free Survival			
	Univariate		Multivariate	
Variables	HR (95%CI)	***P*-value**	HR (95%CI)	***P*-value**
Age(≥65 vs.<65)	0.906 (0.456-1.804)	0.780		
Gender (male vs. female)	0.857 (0.460-1.598)	0.627		
Smoking history (cur/for vs. never)	1.550 (0.779-3.084)	0.212		
Predominant subtybe (acinar vs. papillary)	0.720 (0.392-1.321)	0.289		
Group				
$MP/S^{+}Lep^{-}$	Ref			
$MP/S^{-}Lep^{-}$	0.634 (0.280-1.437)	0.275	0.949 (0.375-2.400)	0.912
$MP/S^{-}Lep^{+}$	0.144 (0.034-0.599)	**0.008**	0.220 (0.051-0.955)	**0.043**
$MP/S^{+}Lep^{+}$	0.148 (0.020-1.087)	0.060	0.242 (0.032-1.821)	0.168
Stage (IB vs. IA)	3.262 (1.759-6.051)	**0.001**	1.564 (0.755-3.239)	0.229
Differentiation (Poorly vs. Well)	1.213 (0.657-2.242)	0.537		
Lymphovascular invasion (present vs. absent)	2.806 (1.513-5.203)	**0.001**	1.761 (0.883-3.514)	0.108
Visceralpleural invasion (absent vs. present)	0.802 (0.371-1.733)	0.575		
Surgical resection (Sublobar resection vs. Lobectomy)	0.462 (0.064-3.359)	0.446		
Laterality (Left vs. right)	0.590 (0.302-1.152)	0.122		
Chemotherapy (yes vs. no)	1.292 (0.693-2.408)	0.421		
CEA(high vs. low)	2.922 (1.529-5.586)	**0.001**	2.094 (0.983-4.464)	0.056
PLR (high vs. low)	2.365 (1.209-4.628)	**0.012**	2.179 (1.068-4.445)	**0.032**

CEA, carcinoembryonic antigen; PLR, platelet-to-lymphocyte ratio; $MP/S^{-}Lep^{+}$ (micropapillary/solid components were absent and lepidic components were present); $MP/S^{+}Lep^{+}$ (both micropapillary/solid component and lepidic component were present); $MP/S^{-}Lep^{-}$ (both micropapillary/solid component and lepidic component were absent); $MP/S^{+}Lep^{-}$ (micropapillary/solid component were present and lepidic component were absent). The bold values statistical significance at $P < 0.05$ level.

patterns include increased laminin-5 expression levels, which is an extracellular matrix protein that is crucial in cell migration, intercellular adhesion, and tumor invasion reflecting the biologically aggressive nature of tumors (24). The presence of lepidic patterns indicated lower cancer cell-specific expression levels of hypoxia markers and a smaller number of tumor-promoting stromal cells (25, 26). The presence of solid and micropapillary patterns were related to metastatic lymph nodes, inversely, the presence of lepidic patterns predicted nonmetastatic lymph nodes (27). In conclusion, the lung adenocarcinoma with the presence of solid/micropapillary components and the absence of lepidic pattern presented the worse DFS. To the best of our knowledge, this is the first study investigating the relationship between PLR and non-predominant components on survival.

We noticed that the presence of non-predominant patterns rather than their proportion was associated with recurrence in stage I acinar/papillary-predominant lung adenocarcinomas. Chen et al. reported the proportion of solid or micropapillary patterns was not related to recurrence in IA lung adenocarcinoma (28). As a result, identifying the presence of a solid or micropapillary component is more valuable than determining the percentage of either of these components. A prognostic nomogram according to a new classification of combined micropapillary and solid components revealed that patients with a total proportion of solid and micropapillary components ≥40% in stage IA patients had shorter DFS and OS compared with less than the total proportion (29). Our conclusions contradict the above findings, maybe due to our small sample size.

In our study, the $MP/S^{+}Lep^{-}$ group could not benefit from adjuvant chemotherapy. Cao et al. also reported that adjuvant chemotherapy is beneficial for solid predominant patterns of stage IB lung adenocarcinoma, while those with solid minor patterns will not (30). The low relapse rates in the non-predominant population especially of stage I tumors and

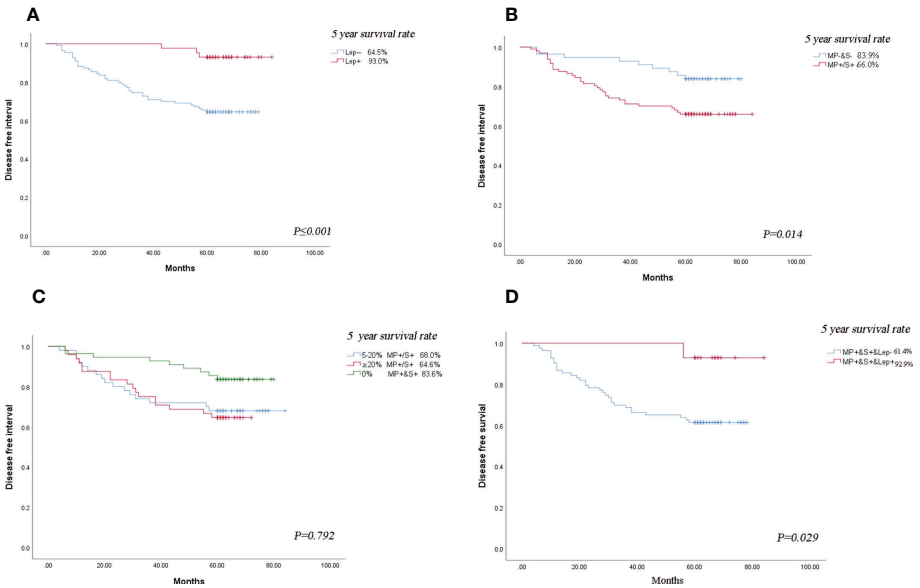


FIGURE 3
Subgroup analysis. **(A)** Comparison of DFS in patients with lung adenocarcinoma based on presence or absence of lepidic minor component. **(B)** Comparison of DFS in patients with lung adenocarcinoma based on the presence or absence of micropapillary/solid minor components component. **(C)** Comparison of DFS in patients with lung adenocarcinoma based on the proportion of micropapillary/solid minor components component. **(D)** Comparison of the effect of the presence or absence of lepidic component on DFS in patients with the micropapillary/solid minor component. The symbol of "&" means and.

limited sample size may explain the failure of adjuvant chemotherapy as a prognostic factor.

There are some limitations in this study. Firstly, owing to the short follow-up period plus the small number of participants, it was difficult to perform the analysis of OS, and we are working to enlarge the sample size. Secondly, the data collection of this study is retrospective, there is an imbalance in the distribution of histological grades, so it is urgent to conduct multi-institutional randomized clinical trials to verify the results. Additionally, we did not evaluate the status of EGFR and KRAS mutation and the PD-L1 expression to explore whether the immunotherapy or targeted therapy could benefit the MP/S^{+}Lep^{-} group. The results show the emphasis should be placed on the non-predominant histological subtype and the necessity of classifying acinar/papillary-predominant adenocarcinomas into different risk groups so as to select high-risk subgroups of patients to administer more intensive surveillance.

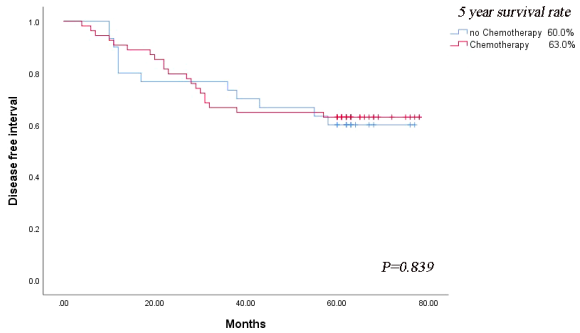


FIGURE 4
Survival outcomes regarding disease-free survival based on the strategy of adjuvant chemotherapy.

Data availability statement

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

Ethics statement

This study was approved by the Ethics Committee of Nanjing Brain Hospital. Informed consent was obtained from all participants in the study. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

WL and QZ searched the database, judged study eligibility, and extracted data. WL and TZ analyzed data and wrote the paper. CX and LLdesigned the study and revised this paper. All authors contributed to the article and approved the submitted version.

Funding

The study was supported by the Major Program of Nanjing Medical Science and Technique Development Foundation (ZKX17044). The funders had no role in the study design, data collection, analysis, or preparation of the 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.

The handling editor HL declared a shared parent affiliation with the authors at the time of review.

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. Travis WD, Brambilla E, Noguchi M, Nicholson A, Geisinger K, Yatabe Y, et al. International association for the study of lung Cancer/American thoracic Society/European respiratory society international multidisciplinary classification of lung adenocarcinoma. *J Thorac Oncol* (2011) 6:244–85. doi: 10.1097/JTO.0b013e318206a221

2. Moreira AL, Ocampo PS, Xia Y, Zhong H, Russell PA, Minami Y, et al. A grading system for invasive pulmonary adenocarcinoma: a proposal from the IASLC pathology committee. *J Thorac Oncol* (2020) 15(10):1599–610. doi: 10.1016/j.jtho.2020.06.001

3. Kim M, Chung YS, Kim KA, Shim HS. Prognostic factors of acinar-or papillary-predominant adenocarcinoma of the lung. *Lung Cancer* (2019) 137:129–35. doi: 10.1016/j.lungcan.2019.09.026

4. Watanabe K, Sakamaki K, Ito H, Yokose T, Yamada K, Nakayama H, et al. Impact of the micropapillary component on the timing of recurrence in patients with resected lung adenocarcinoma. *Eur J Cardiothorac Surg* (2020) 58(5):1010–8. doi: 10.1093/ejcts/ezaa138

5. Ma M, She Y, Ren Y, Dai C, Zhang L, Xie H, et al. Micropapillary or solid pattern predicts recurrence free survival benefit from adjuvant chemotherapy in patients with stage IB lung adenocarcinoma. *J Thorac Dis* (2018) 10(9):5384–93. doi: 10.21037/jtd.2018.08.64

6. Murakami S, Ito H, Tsubokawa N, Mimae T, Sasada S, Yoshiya T, et al. Prognostic value of the new IASLC/ATS/ERS classification of clinical stage IA lung adenocarcinoma. *Lung Cancer* (2015) 90(2):199–204. doi: 10.1016/j.lungcan.2015.06.022

7. Zhu E, Dai C, Xie H, Su H, Hu X, Li M, et al. Lepidic component identifies a subgroup of lung adenocarcinoma with a distinctive prognosis: a multicenter propensity-matched analysis. *Ther Adv Med Oncol* (2020) 12:1758835920982845. doi: 10.1177/1758835920982845

8. Okubo Y, Kashima J, Teishikata T, Muraoka Y, Yotsukura M, Yoshidav Y, et al. Prognostic impact of the histological lepidic component in pStage IA adenocarcinoma. *J Thorac Oncol* (2022) 17(1):67–75. doi: 10.1016/j.jtho.2021.09.006

9. Wang W, Hu Z, Zhao J, Huang Y, Rao S, Yang J, et al. Both the presence of a micropapillary component and the micropapillary predominant subtype predict poor prognosis after lung adenocarcinoma resection: a meta-analysis. *J Cardiothorac Surg* (2020) 15(1):154. doi: 10.1186/s13019-020-01199-8

10. Ninomiya H, Inamura K, Mun M, Nishio M, Ishikawa Y. Relationship between PathologicT1 categories and pathologic factors affecting prognosis in pulmonary adenocarcinoma. *JTO Clin Res Rep* (2022) 3:100293. doi: 10.1016/j.jtocrr.2022.100293

11. Peng B, Li G, Guo Y. Prognostic significance of micropapillary and solid patterns in stage IA lung adenocarcinoma. *Am J Transl Res* (2021) 13(9):10562–9.

12. Zhao Z-R, To KF, Mok TSK, Ng C. Is there significance in identification of non-predominant micropapillary or solid components in early-stage lung adenocarcinoma? *Interact Cardiovasc Thorac Surg* (2017) 24(1):121–5. doi: 10.1093/icvts/ivw283

13. Choi SH, Jeong JY, Lee SY, Shin KM, Jeong SY, Park TI, et al. Clinical implication of minimal presence of solid or micropapillary subtype in early-stage lung adenocarcinoma. *Thorac Cancer* (2021) 12(2):235–44. doi: 10.1111/1759-7714.13754

14. Choi Y, Kim J, Park H, Kim H, Kim J, Jeong J, et al. Rethinking a non-predominant pattern in invasive lung adenocarcinoma: Prognostic dissection focusing on a high-grade pattern. *Cancers (Basel)* (2021) 13(11):2785. doi: 10.3390/cancers13112785

15. Liu X, Sun K, Yang F, et al. Different pathologic types of early stage lung adenocarcinoma have different post-operative recurrence patterns. *Thorac Cancer* (2021) 12:2205–13. doi: 10.1111/1759-7714.14049

16. Qu Y, Lin H, Zhang C, Sui X, Jiang G, Wang J, et al. Cribriform pattern in lung invasive adenocarcinoma correlates with poor prognosis in a Chinese cohort. *Pathol Res Pract* (2019) 215(2):347–53. doi: 10.1016/j.prp.2018.12.014

17. Yanagawa N, Shiono S, Abiko M, Katahira M, Osakabe M, Ogata S, et al. The clinical impact of solid and micropapillary patterns in resected lung adenocarcinoma. *J Thorac Oncol* (2016) 11(11):1976–83. doi: 10.1016/j.jtho.2016.06.014

18. Yoshizawa A, Motoi N, Riely GJ, Sima CS, Gerald WL, Kris MG, et al. Impact of proposed IASLC/ATS/ERS classification of lung adenocarcinoma: prognostic subgroups and implications for further revision of staging based on analysis of 514 stage I cases. *Mod Pathol* (2011) 24(5):653–64. doi: 10.1038/modpathol.2010.232

19. Yao J, Zhu E, Li M, Liu J, Zhang L, Ke H, et al. Prognostic impact of micropapillary component in patients with node-negative subcentimeter lung adenocarcinoma: A Chinese cohort study. *Thorac Cancer* (2020) 11(12):3566–75. doi: 10.1111/1759-7714.13702

20. Zhao Y, Wang R, Shen X, Pan Y, Cheng C, Li Y, et al. Minor components of micropapillary and solid subtypes in lung adenocarcinoma are predictors of lymph node metastasis and poor prognosis. *Ann Surg Oncol* (2016) 23(6):2099–105. doi: 10.1245/s10434-015-5043-9

21. Mäkinen JM, Laitakari K, Johnson S, Mäkitaro R, Bloigu R, Lappi-Blanco E, et al. Nonpredominant lepidic pattern correlates with better outcome in invasive lung adenocarcinoma. *Lung Cancer* (2015) 90(3):568–74. doi: 10.1016/j.lungcan.2015.10.014

22. Chen T, Luo J, Gu H, Gu Y, Huang Q, Wang Y, et al. Impact of solid minor histologic subtype in postsurgical prognosis of stage I lung adenocarcinoma. *Ann Thorac Surg* (2018) 105(1):302–8. doi: 10.1016/j.athoracsur.2017.08.018

23. Hou Y, Song W, Chen M, Zhang J, Luo Q, Um S, et al. The presence of lepidic and micropapillary/solid pathological patterns as minor components has prognostic value in patients with intermediate-grade invasive lung adenocarcinoma. *Transl Lung Cancer Res* (2022) 11(1):64–74. doi: 10.21037/tlcr-21-934

24. Takuwa T, Ishii G, Nagai K, Yoshida J, Nishimura M, Hishida T, et al. Characteristic immunophenotype of solid subtype component in lung adenocarcinoma. *Ann Surg Oncol* (2012) 19(12):3943–52. doi: 10.1245/s10434-012-2428-x

25. Chong Y, Wu Y, Liu J, Han C, Gong L, Liu X, et al. Clinicopathological models for predicting lymph node metastasis in patients with early-stage lung adenocarcinoma: the application of machine learning algorithms. *J Thorac Dis* (2021) 13(7):4033–42. doi: 10.21037/jtd-21-98

26. Katsumata S, Aokage K, Miyoshi T, Tane K, Nakamura H, Sugano M, et al. Differences of tumor microenvironment between stage I lepidic-positive and lepidic-negative lung adenocarcinomas. *J Thorac Cardiovasc Surg* (2018) 156 (4):1679–1688.e2. doi: 10.1016/j.jtcvs.2018.05.053

27. Mäkinen JM, Laitakari K, Johnson S, Mäkitaro R, Bloigu R, Pääkkö P, et al. Histological features of malignancy correlate with growth patterns and patient outcome in lung adenocarcinoma. *Histopathology* (2017) 71(3):425–36. doi: 10.1111/his.13236

28. Chen C, Chen Z-J, Li W-J, Pan X-F, Wen Y-Y, Deng T, et al. Impact of minimal solid and micropapillary components on invasive lung adenocarcinoma recurrence. *Ann Diagn Pathol* (2022) 59:151945. doi: 10.1016/j.anndiagpath.2022.151945

29. Huang W, Zhang H, Zhang Z, Zhang B, Sun X, Huo Y, et al. A prognostic nomogram based on a new classification of combined micropapillary and solid components for stage IA invasive lung adenocarcinoma. *J Surg Oncol* (2022) 125 (4):796–808. doi: 10.1002/jso.26760

30. Cao S, Teng J, Xu J, Han B, Zhong H. Value of adjuvant chemotherapy in patients with resected stage IB solid predominant and solid non-predominant lung adenocarcinoma. *Thorac Cancer* (2019) 10(2):249–55. doi: 10.1111/1759-7714.12942

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 17 January 2023
DOI 10.3389/fonc.2022.1098728

Check for updates

OPEN ACCESS

EDITED BY
Hongda Liu,
Nanjing Medical University, China

REVIEWED BY
Terry Moody,
National Cancer Institute (NIH), United States
Lei Chen,
The Second Affiliated Hospital of Harbin Medical University, China
Dongxin Zhang,
Beijing Tongren Hospital, Capital Medical University, China
Zhiqiang Chen,
Qilu Hospital, Shandong University, China

*CORRESPONDENCE
Tao Xin
sbpdz69958264@163.com
drxintao@126.com

SPECIALTY SECTION
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

RECEIVED 15 November 2022
ACCEPTED 22 December 2022
PUBLISHED 17 January 2023

CITATION
Liang N, Sun S, Li Z, Wu T, Zhang C and Xin T (2023) CCKAR is a biomarker for prognosis and asynchronous brain metastasis of non-small cell lung cancer.
Front. Oncol. 12:1098728.
doi: 10.3389/fonc.2022.1098728

COPYRIGHT
© 2023 Liang, Sun, Li, Wu, Zhang and Xin. 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.

CCKAR is a biomarker for prognosis and asynchronous brain metastasis of non-small cell lung cancer

Nan Liang[1], Suohui Sun[1], Zheng Li[1], Tao Wu[1], Chunpu Zhang[1] and Tao Xin[2]*

[1]Department of Neurosurgery, the Second Affiliated Hospital of Shandong First Medical University, Jinan, China, [2]Department of Neurosurgery, the First Affiliated Hospital of Shandong First Medical University and Shandong Provincial Qianfoshan Hospital, Jinan, China

Background: Non-small cell lung cancer (NSCLC) is the most common histological type of lung cancer, and brain metastasis (BM) is the most lethal complication of NSCLC. The predictive biomarkers and risk factors of asynchronous BM are still unknown.

Materials and methods: A total of 203 patients with NSCLC were enrolled into our cohort and followed up. The clinicopathological factors such as tumor size, T stage, lymphatic invasion, metastasis and asynchronous BM were investigated. CCKAR expression in NSCLC and resected BM was assessed by IHC, and CCKAR mRNAs in NSCLC and para-tumor tissues were estimated by qRT-PCR. The correlations between CCKAR expression, BM and other clinicopathological factors were assessed by chi-square test, and prognostic significance of CCKAR was estimated by univariate and multivariate analyses.

Results: CCKAR was highly expressed in NSCLC tissues compared with para-tumor tissues. CCKAR expression in NSCLC was significantly associated with asynchronous BM. The BM percentages for NSCLC patients with low and high CCKAR were surprisingly 5.2% and 66.6%, respectively. CCKAR expression and BM were unfavorable factors predicting unfavorable outcome of NSCLC. Moreover, CCKAR expression in NSCLC was an independent risk factor of asynchronous BM.

Conclusions: CCKAR is a prognostic biomarker of NSCLC. CCKAR expression in NSCLC is positively associated with asynchronous BM, and is a risk factor of asynchronous BM from NSCLC.

KEYWORDS

CCKAR, brain metastasis, non-small cell lung cancer, prognosis, Biomarker

Introduction

Lung cancer is the most common cause of cancer death worldwide, resulting in approximate 1.6 million deaths each year (1). Histologically, lung cancer is mainly divided into small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) (2). NSCLC accounts for about 85% of all the lung cancers (3). Brain metastasis (BM) is the most common and lethal event in patients with lung cancer (4). The development of BM remains one of the main factors associated with poor prognosis and mortality in patients with lung cancer. About 150,000 patients with cancer develop into BM in USA each year (5), and lung is the most common primary site for BM (6, 7). Almost 10% of newly diagnosed patients with NSCLC develop metastasis and 25%–40% acquire BM during the course of disease (8). Although multimodal treatments and advances have been used in BM treatment, the prognosis of BM is extremely poor, ranging from 1.5 to 9.5 months (9, 10). More unfortunately, there is no way which can predict NSCLC patients with high risk to suffer BM during the disease course to date.

Cholecystokinin (CCK) receptor, including CCKAR and CCKBR, is a receptor for cholecystokinin, mainly regulating pancreatic growth and enzyme secretion, smooth muscle contraction of the gallbladder and stomach (11). Previous studies mainly suggest that CCKAR is primarily expressed in the alimentary tract, while CCKBR is mainly found in the brain and the stomach (12). Emerging evidence shows that CCKAR is also expressed in central nervous system and modulates feeding and dopamine-induced behavior (13). The role of CCKAR in cancer is also gradually revealed in these years, including hepatocellular carcinoma, gallbladder carcinoma, gastric cancer, et al (14–16). Moreover, targeting CCK receptor is becoming a promising therapy for cancer treatment. However, the role of CCKAR in lung cancer and its metastasis is still unknown.

Materials and methods

Patient cohort and ethics

Our study enrolled 434 patients who underwent surgical resection of NSCLC, which comprise of the testing cohort, from 2017 to 2020 in the First Affiliated Hospital of Shandong First Medical University and the Second Affiliated Hospital of Shandong First Medical University. From the testing cohort, a total of 203 patients were selected and their outcomes were followed up. The inclusion standards contain that (i) available formalin-fixed tumor tissues and follow-ups, (ii) no pre-operational adjuvant therapy including chemotherapy or radiotherapy. Patients who died during the perioperative period were excluded from the study. All patients received the standard treatment strategies of NSCLC after lung cancer resection including chemotherapy if tumors were invasive, and target therapy if there existed available mutations (17). The study was approved by the Ethics Committee of the First Affiliated Hospital of Shandong First Medical University, and the Second Affiliated Hospital of Shandong First Medical University (Approval No. 2021-091).

Immunohistochemistry and evaluation

CCKAR expression in NSCLC and BM was detected and semi-qualified with IHC by streptavidin peroxidase complex method as previous described (18). Slides were treated with boiled citrate buffer (pH = 6.0) for antigen retrieval. The inactivity of endogenous peroxidase was completed by 3% hydrogen peroxide, and unspecific antigen binding was blocked by 5% bovine serum albumin. The primary antibody of CCKAR (1:100, Santa Cruz Biotechnology, Santa Cruz, CA, USA) or CCK-8 (1:100, Abcam, ab27441, UK) was used to incubate the specimens overnight at 4°, and the corresponding secondary antibody labelled with streptavidin-biotin-peroxidase reagent (Beyotime, Beijing, China) was administrated to incubate the specimens for 30 minutes at room temperature. At last, the antigens were visualized by incubation in the 3,3'-diaminobenzidine solution.

The IHC results were semi-quantified by two aspects, the staining intensity and the positive stained cell percentage, which were the IHC score. The IHC score was assessed by two senior pathologists who were unaware of the clinical data. The IHC score was defined as the score of the staining intensity multiplied by the score of positive stained cell percentage. The scores of staining intensity were defined as: score 0, 1, 2 and 3 for negative staining, weak staining, moderate staining and strong staining, respectively. The scores of positively stained cell percentage were defined as: score 1, 2, 3 and 4 for <25% positive cells, 25%-50% positive cells, 50%-75% positive cells and 75%-100% positive cells respectively. The final IHC score varied from 0 to 12. The cohort was divided into subsets with different CCKAR expression by the cut-off of IHC score, which was identified by receiver operating characteristic (ROC) curve.

RNA extraction and qRT-PCR

qRT-PCR was performed to assess CCKAR expression in fresh NSCLC tissues and para-tumor tissues. The total RNA was extracted by TRIzol agent (Thermo Fisher, Waltham, MA, USA). Reverse transcription of cDNA was accomplished with ReverTra kit

(TOYOBO, Japan), and the quantitative real-time polymerase chain reaction (qRT-PCR) was applied with Thermo Fisher 7500 PCR System and SYBR Premix. The quantification of qRT-PCR results was analyzed by the $2^{-\Delta\Delta Ct}$ method with GAPDH as an internal control. The primers of CCKAR and GAPDH were designed as follows: CCKAR, forward: 5'-ATGGATGTGGTTGACAGCCTT-3', reverse: 5'-AAGCGTCTCATTTTCGAGCCC-3'; GAPDH: forward: 5'-GAGTCAACGGATTTGGTCGT-3', reverse: 5'-GACAAGCTTCCCGTTCTCAG-3'.

Statistical analysis

SPSS 22.0 software (SPSS, Chicago, IL, USA) was used to analyze all the statistical significance. The correlations between CCKAR, BM and other clinicopathological factors were assessed by the chi-square test. The statistical difference between subgroups was assessed by the log-rank test, and overall survival(OS) curves were plotted by Kaplan-Meier method. The Cox-regression hazard model was applied to identify the independent prognostic factors. *P* value less than 0.05 was regarded as statistically significant.

Results

Expression of CCKAR in primary NSCLC

The expressions of CCKAR in NSCLC and corresponding tumor-adjacent tissues were detected with IHC and qRT-PCR. In 203 cases of NSCLC, IHC was performed to show the expression and localization of CCKAR. In our study, CCKAR was mainly expressed in cytoplasm and membrane (Figure 1A). The patients were further divided into subgroups with low or high expression of CCKAR, accounting for 116 and 87 respectively. CCK-8 expression was also detected with IHC in NSCLC. In NSCLC, CCK-8 was hardly detected (Figure 1B).

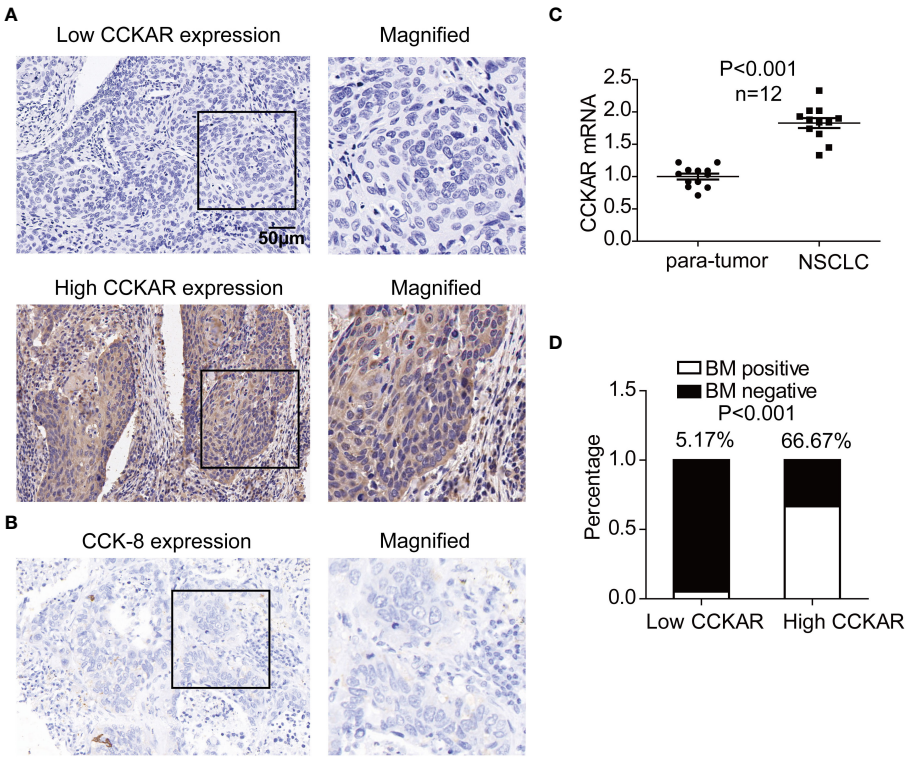


FIGURE 1
Expression of CCKAR in NSCLC. **(A)** The expression of CCKAR in NSCLC was detected by IHC. Patients with NSCLC were divided into subsets with low or high CCKAR expression. **(B)** Expression of CCK-8 in NSCLC was detected with IHC and representative images were shown. **(C)** CCKAR expression in 12 pairs of NSCLC and para-tumor tissues was detected with qRT-PCR. The statistical significance was evaluated by paired t test. **(D)** Asynchronous BM percentages of patients with low and high CCKAR were calculated and compared with chi-square method. Patients with high CCKAR in NSCLC primary tumor were much more susceptible to BM.

In 12 pairs of NSCLC tissues and para-tumor tissues, CCKAR expression was detected with qRT-PCR. As the result, CCKAR in NSCLC was significantly higher than that in para-tumor tissues (Figure 1C), suggesting that CCKAR may play an important role in NSCLC progression.

Correlation between CCKAR, BM and clinicopathological factors

These 203 patients were followed-up and a total of 64 patients suffered from asynchronous BM. The correlations between CCKAR and clinicopathological factors were analyzed with chi-square test (Table 1). Interestingly, high CCKAR in primary NSCLC was significantly associated with positive asynchronous BM (P<0.001). The number of patients with BM was 58 in high-CCKAR patients but only 6 in low-CCKAR patients. The BM percentages for NSCLC patients with low and high CCKAR were surprisingly 5.2% and 66.6%, respectively (Figure 1D). Moreover, CCKAR tended to be significantly associated with advanced TNM stage (P=0.087) and female gender (P=0.067).

The correlation between asynchronous BM and clinicopathological factors was also analyzed (Table 2). Patients with positive lymphatic invasion were more likely to suffer from asynchronous BM (P=0.010). More interestingly, target therapy of NSCLC was associated with low probability of BM (P=0.044).

Prognostic value of CCKAR in NSCLC

The prognostic significance of CCKAR and other clinicopathological parameters were analyzed by univariate and multivariate analyses. The univariate analysis was performed with the log-rank test (Table 3). In the univariate analysis, CCKAR was a significant prognostic factor of NSCLC (Figure 2A). In addition, large tumor size(P =0.006), lymphatic invasion (P<0.001), advanced T stage (P=0.025) and TNM stage (P<0.001) were all substantially associated with unfavorable outcome (Figures 2B–E). Moreover, positive BM was a prognostic indicator of poor prognosis (Figure 2E).

The independent prognostic factors of NSCLC were further validated by multivariate analysis (Table 3). The prognostic factors in univariate analysis were enrolled into the Cox-regression model including tumor size, T stage, lymphatic invasion, BM and CCKAR expression. Positive lymphatic invasion was identified as an independent prognostic parameter of NSCLC. CCKAR and BM were not an independent prognostic biomarker of NSCLC mainly because of they had significant correlations.

Clinical significance of CCKAR expression in BM lesion

A total of 43 patients with BM underwent surgical resection, and CCKAR expression in BM lesion was assessed with IHC and WB (Figure 3A, B). According to the cut-off of CCKAR IHC score in BM, we divided these 43 patients into subgroups with high or low CCKAR expression, accounting for 23.3% (10/43) and 76.7% (33/43) respectively. These results further indicate that CCKAR may be a driving force of BM from NSCLC.

Discussion

Almost 10% of NSCLC are diagnosed with metastasis synchronously, and 25%-40% acquire BM during the course of NSCLC (19). As to patients with mediastinal lymph node metastasis of lung cancer, up to 68% of patients eventually suffer BM (20). BM may cause severe symptoms such as neurological defects, cognitive impairment, and emotional difficulties which require surgical treatment (21). For patients with BM, the prognosis is extremely poor with a median survival time about only 3-6 months (22). However, the life span of BM patients who receive early treatment is much longer than those whose treatment is not prompt (23, 24). So identifying early and effective biomarkers which can predict BM and prognosis is a severe and important task for BM treatment. Here in our study, we identified CCKAR as a prognostic biomarker of NSCLC, and showed that CCKAR in NSCLC was significantly associated with asynchronous BM. CCKAR was significantly associated with asynchronous BM probability of NSCLC. Our results indicated that patients with high CCKAR should receive more severe surveillance for BM. This is very helpful to select the patients with high risk of BM and guide stricter review and early treatment.

Treatment to patients with BM from NSCLC has not got total consensus (25). The treatments depend on the performance status, the effect of other adjuvant therapy and overall health of the patient (18). Many conflicts exist as to the treatment options and the large-scale and multi-centered cohort study should be conducted. Identifying systemic treatment guidelines based on the molecular pattern of BM is an urgent task for the precision treatment of patients with BM. Biomarkers of BM could not only select patients for appropriate cancer management strategies but also help screen more effective drug targets. Our results not only define CCKAR as a BM indicator of NSCLC, but also imply that CCKAR may be a potential drug target of BM from NSCLC. CCKAR has some available inhibitors such as lintitript, rebamipide and loxiglumide (26), which may be promising inhibitor of NSCLC and BM treatment.

10.3389/fonc.2022.1098728

TABLE 1 The correlation between CCKAR expression and clinicopathological factors in NSCLC.

Factors		CCKAR		
	n	Low	High	P*
Sex				
Male	111	57	54	0.067
Female	92	59	33	
Age				
<60	75	46	29	0.356
≥60	128	70	58	
Tumor diameter				
≤3.5cm	95	59	36	0.18
>3.5cm	108	57	51	
Histological grade				
I	121	70	51	0.804
II-III	82	46	36	
T stage				
I+II	130	78	52	0.272
II+IV	73	38	35	
Lymphatic invasion				
Negative	100	63	37	0.097
Positive	103	53	50	
Metastasis				
No	191	109	82	0.932
Yes	12	7	5	
BM				
No	139	110	29	<0.001
Yes	64	6	58	
TNM stage				
I+II	79	51	28	0.087
II+IV	124	65	59	
Chemotherapy				
No	59	34	25	0.939
Yes	144	82	62	
Target therapy				
No	168	94	74	0.453
Yes	35	22	13	

*calculated by chi-square test.

10.3389/fonc.2022.1098728

TABLE 2 The prognostic significance of CCKAR, BM and other clinicopathological factors.

Factors	Univariate analysis		Multiriate analysis		
	5-year OS rate	P*	HR	95% CI	P#
Sex					
Male	21.2	0.258			
Female	24.4				
Age					
<60	27.2	0.233			
≥60	19.6				
Tumor size					
≤3.5cm	33.8	0.006	1		
>3.5cm	12.1		1.36	0.95-1.94	0.095
Histological grade					
I	26	0.173			
II-III	18				
T stage					
I+II	27.6	0.025	1		
II+IV	13.3		0.92	0.64-1.33	0.668
Lymphatic invasion					
Negative	37.1	<0.001	1		
Positive	8.6		2.08	1.46-2.97	<0.001
Metastasis					
No	23.6	0.445			
Yes	8.3				
BM					
No	28.2	<0.001	1		
Yes	9.5		1.41	0.91-2.19	0.123
TNM stage					
I+II	46.7	<0.001			
II+IV	8.8				
CCKAR					
Low	27.8	0.016	1		
High	15.7		1.15	0.75-1.77	0.516
Chemotherapy					
Yes	23.0	0.792			
No	21.2				
Target therapy					
Yes	24.1	0.600			
No	22.4				

*calculated by log-rank test;# calculated by Cox-regression model.

TABLE 3 The correlation between BM and clinicopathological factors in NSCLC.

Factors	BM		
	No	Yes	P*
Sex			
Male	69	35	0.504
Female	70	29	
Age			
<60	57	18	0.077
≥60	82	46	
Tumor diameter			
≤3.5cm	55	24	0.779
>3.5cm	84	40	
Histological grade			
I	88	33	0.113
II-III	51	31	
T stage			
I+II	93	37	0.212
II+IV	46	27	
Lymphatic invasion			
Negative	77	23	0.010
Positive	62	41	
Metastasis			
No	132	59	0.436
Yes	7	5	
Chemotherapy			
No	37	22	0.258
Yes	102	42	
Target therapy			
No	110	58	0.044
Yes	29	6	

*calculated by chi-square test.

CCK is the most abundant peptides in the gastrointestinal tract and also in the central nervous system, functioning as important hormones as well as neurotransmitters (25, 27). CCKAR is a promising drug target for gastrointestinal and metabolic diseases. In cancer study, the oncogenic role and mechanism of CCKAR are occasionally reported. As a hormone modulating gallbladder motility, CCKAR expression or genetic variation is reported to influence hepatocellular carcinoma and biliary tract cancers (16, 28). Compared with CCKBR, the function of CCKAR in tumorigenesis and tumor progression is much less studied (29). Among all G protein-coupled receptors(GPCRs), CCKAR is unique because it can couple with all G-protein subtypes, including Gαs, Gαi and Gαq. Different downstream signaling of CCKAR regulates

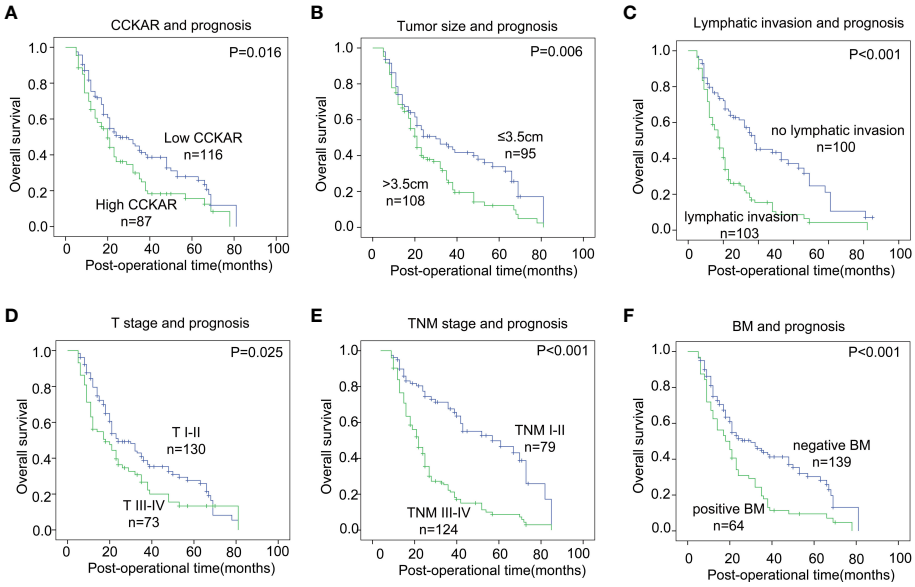


FIGURE 2
The correlation between NSCLC clinicopathological factors and outcome. NSCLC patients were divided into different subgroups according to CCKAR expression **(A)**, tumor size **(B)**, lymphatic invasion **(C)**, T stage **(D)**, TNM stage **(E)** and BM **(F)**. The statistical significance was calculated by the log-rank test.

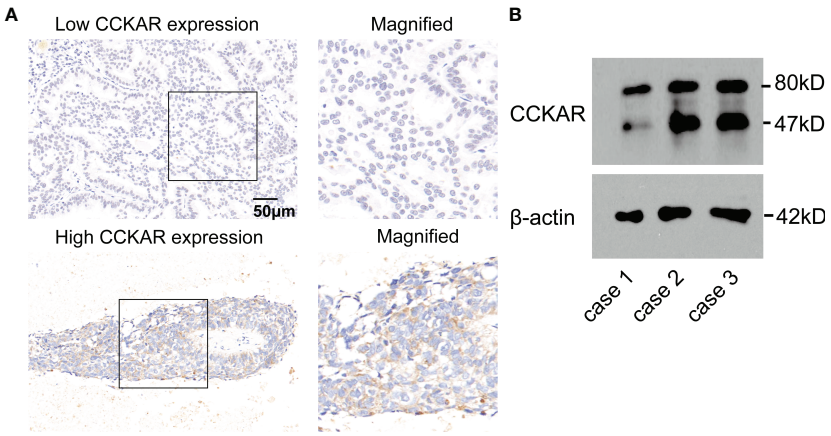


FIGURE 3
CCKAR expression in BM from NSCLC. **(A)** The expression of CCKAR in BM from NSCLC was detected by IHC. Patients with BM from NSCLC were divided into subsets with low or high CCKAR expression. **(B)** The expression of CCKAR in 3 BM tissues was detected with Western Blotting.

distinct functions (30). The underlying mechanism of how CCKAR correlates with poor prognosis of NSCLC is not involved in our study because of the complex signaling network downstream of CCKAR. We wish that more evidence on the important clinical significance of CCKAR could provide more fundamental research of how CCKAR is related with poor prognosis of cancer.

In conclusion, we assessed the expression of CCKAR in NSCLC and BM from NSCLC, and analyzed the correlation between CCKAR expression and clinicopathological factors including BM. We identified CCKAR as a prognostic biomarker of NSCLC. CCKAR expression in NSCLC is positively associated with asynchronous BM, and also is a risk factor of asynchronous BM from NSCLC. Our results indicate that patients with high CCKAR in NSCLC should receive more severe surveillance for BM, which is very helpful to select high-risk patients for individual and early treatment.

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 study was approved by the Ethics Committee of the First and Second Affiliated Hospital of Shandong First Medical University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

NL, SS, ZL, TW, CZ and TX established the cohort and collected the clinical information, NL performed the experiments and analyzed the data. TX wrote the paper. All authors contributed to the article and approved the submitted version.

Funding

The study is supported by Science Innovation Program of Tai'an City (2021NS215).

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 reviewer ZC declared a shared parent affiliation with the author TX to the handling editor at the time of review.

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. 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. Broderick SR. Adjuvant and neoadjuvant immunotherapy in non-small cell lung cancer. *Thorac Surg Clinics* (2020) 30:215–20. doi: 10.1016/j.thorsurg.2020.01.001

3. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. *Nature* (2018) 553:446–54. doi: 10.1038/nature25183

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

5. Bovi JA, White J. Radiation therapy in the prevention of brain metastases. *Curr Oncol Rep* (2012) 14:55–62. doi: 10.1007/s11912-011-0208-6

6. Nathoo N, Chahlavi A, Barnett GH, Toms SA. Pathobiology of brain metastases. *J Clin Pathol* (2005) 58:237–42. doi: 10.1136/jcp.2003.013623

7. Subramanian A, Harris A, Piggott K, Shieff C, Bradford R. Metastasis to and from the central nervous system–the 'relatively protected site'. *Lancet Oncol* (2002) 3:498–507. doi: 10.1016/S1470-2045(02)00819-7

8. Achrol AS, Rennert RC, Anders C, Soffietti R, Ahluwalia MS, Nayak L, et al. Brain metastases. *Nat Rev Dis Primers* (2019) 5:5. doi: 10.1038/s41572-018-0055-y

9. Langer CJ, Mehta MP. Current management of brain metastases, with a focus on systemic options. *J Clin Oncol Off J Am Soc Clin Oncol* (2005) 23:6207–19. doi: 10.1200/JCO.2005.03.145

10. Sperduto PW, Yang TJ, Beal K, Pan H, Brown PD, Bangdiwala A, et al. Estimating survival in patients with lung cancer and brain metastases: An update of the graded prognostic assessment for lung cancer using molecular markers (Lung-molGPA). *JAMA Oncol* (2017) 3:827–31. doi: 10.1001/jamaoncol.2016.3834

11. Zhang X, He C, Wang M, Zhou Q, Yang D, Zhu Y, et al. Structures of the human cholecystokinin receptors bound to agonists and antagonists. *Nat Chem Biol* (2021) 17:1230–7. doi: 10.1038/s41589-021-00866-8

12. Dufresne M, Seva C, Fourmy D. Cholecystokinin and gastrin receptors. *Physiol Rev* (2006) 86:805–47. doi: 10.1152/physrev.00014.2005

13. Bowers ME, Choi DC, Ressler KJ. Neuropeptide regulation of fear and anxiety: Implications of cholecystokinin, endogenous opioids, and neuropeptide y. *Physiol Behav* (2012) 107:699–710. doi: 10.1016/j.physbeh.2012.03.004

14. Baldwin GS, Shulkes A. CCK receptors and cancer. *Curr topics medicinal Chem* (2007) 7:1232–8. doi: 10.2174/156802607780960492

15. Gay MD, Safronenka A, Cao H, Liu FH, Malchiodi ZX, Tucker RD, et al. Targeting the cholecystokinin receptor: A novel approach for treatment and prevention of hepatocellular cancer. *Cancer Prev Res (Phila)* (2021) 14:17–30. doi: 10.1158/1940-6207.CAPR-20-0220

16. Wu W, Ouyang B, Lu Z, Liu H, Tan Y, Cui P. CCK1 receptor is involved in the regulation of protein lysine acetylation in GBC-SD cells and gallbladder carcinoma. *Irish J Med Sci* (2017) 186:883–8. doi: 10.1007/s11845-017-1603-2

17. Postmus PE, Kerr KM, Oudkerk M, Senan S, Waller DA, Vansteenkiste J, et al. Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol Off J Eur Soc Med Oncol* (2017) 28:iv1–iv21. doi: 10.1093/annonc/mdx222

18. Li Z, Liang N, Wang N, Jia Y, Tian C. WDR5 is a prognostic biomarker of brain metastasis from non-small cell lung cancer. *Front Oncol* (2022) 12:1023776. doi: 10.3389/fonc.2022.1023776

19. Lin CY, Chen HJ, Huang CC, Lai LC, Lu TP, Tseng GC, et al. ADAM9 promotes lung cancer metastases to brain by a plasminogen activator-based pathway. *Cancer Res* (2014) 74:5229–43. doi: 10.1158/0008-5472.CAN-13-2995

20. Wang SY, Ye X, Ou W, Lin YB, Zhang BB, Yang H. Risk of cerebral metastases for postoperative locally advanced non-small-cell lung cancer. *Lung Cancer* (2009) 64:238–43. doi: 10.1016/j.lungcan.2008.08.012

21. Laack NN, Brown PD. Cognitive sequelae of brain radiation in adults. *Semin Oncol* (2004) 31:702–13. doi: 10.1053/j.seminoncol.2004.07.013

22. Zhu G, Luo J, Ouyang Z, Cheng Z, Deng Y, Guan Y, et al. The assessment of prostate cancer aggressiveness using a combination of quantitative diffusion-weighted imaging and dynamic contrast-enhanced magnetic resonance imaging. *Cancer Manage Res* (2021) 13:5287–95. doi: 10.2147/CMAR.S319306

23. Goldberg SB, Contessa JN, Omay SB, Chiang V. Lung cancer brain metastases. *Cancer J* (2015) 21:398–403. doi: 10.1097/PPO.0000000000000146

24. Schwer AL, Gaspar LE. Update in the treatment of brain metastases from lung cancer. *Clin Lung Cancer* (2006) 8:180–6. doi: 10.3816/CLC.2006.n.045

25. Rehfeld JF. Cholecystokinin-from local gut hormone to ubiquitous messenger. *Front Endocrinol* (2017) 8:47. doi: 10.3389/fendo.2017.00047

26. Pu Y, Song Y, Zhang M, Long C, Li J, Wang Y, et al. GOLM1 restricts colitis and colon tumorigenesis by ensuring notch signaling equilibrium in intestinal homeostasis. *Signal transduction targeted Ther* (2021) 6:148. doi: 10.1038/s41392-021-00535-1

27. Schubert ML, Rehfeld JF. Gastric peptides-gastrin and somatostatin. *Compr Physiol* (2019) 10:197–228. doi: 10.1002/cphy.c180035

28. Xu HL, Hsing AW, Vogtmann E, Chu LW, Cheng JR, Gao J, et al. Variants in CCK and CCKAR genes to susceptibility to biliary tract cancers and stones: A population-based study in shanghai, China. *J Gastroenterol Hepatol* (2013) 28:1476–81. doi: 10.1111/jgh.12278

29. Smith JP, Fonkoua LK, Moody TW. The role of gastrin and CCK receptors in pancreatic cancer and other malignancies. *Int J Biol Sci* (2016) 12:283–91. doi: 10.7150/ijbs.14952

30. Ning SL, Zheng WS, Su J, Liang N, Li H, Zhang DL, et al. Different downstream signalling of CCK1 receptors regulates distinct functions of CCK in pancreatic beta cells. *Br J Pharmacol* (2015) 172:5050–67. doi: 10.1111/bph.13271

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 20 January 2023
DOI 10.3389/fonc.2023.1083417

Check for updates

OPEN ACCESS

EDITED BY
Hongda Liu,
Nanjing Medical University, China

REVIEWED BY
Yaping Xu,
Tongji University, China
Qi Shen,
The First Affiliated Hospital of University of Science and Technology of China Anhui Provincial Hospital, China

*CORRESPONDENCE
Jinming Yu
sdyujinming@163.com
Jie Liu
linchuangliujie@163.com

SPECIALTY SECTION
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

RECEIVED 29 October 2022
ACCEPTED 04 January 2023
PUBLISHED 20 January 2023

CITATION
Wang B, Zou B, Xu S, Zhao C, Pei J, Wang S, Zhao K, Yu J and Liu J (2023) Postoperative ctDNA detection predicts relapse but has limited effects in guiding adjuvant therapy in resectable stage I NSCLC.
Front. Oncol. 13:1083417.
doi: 10.3389/fonc.2023.1083417

COPYRIGHT
© 2023 Wang, Zou, Xu, Zhao, Pei, Wang, Zhao, Yu and Liu. 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.

Postoperative ctDNA detection predicts relapse but has limited effects in guiding adjuvant therapy in resectable stage I NSCLC

Bolin Wang[1,2,3], Bing Zou[2,3], Shengnan Xu[2,3], Chao Zhao[2,3], Jinli Pei[2,3], Shijie Wang[2,3], Kunlong Zhao[2,3], Jinming Yu[2,3]* and Jie Liu[2,3]*

[1]Lung Cancer Center, West China Hospital, Sichuan University, Chengdu, China, [2]Department of Radiation Oncology and Shandong Provincial Key Laboratory of Radiation Oncology, Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, Jinan, Shandong, China, [3]Research Unit of Radiation Oncology, Chinese Academy of Medical Sciences, Jinan, Shandong, China

Background: To date, identifying resectable stage I non-small cell lung cancer (NSCLC) patients likely to benefit from adjuvant therapy (ADT) remains a major challenge. Previous studies suggest that circulating tumor DNA (ctDNA) is emerging as a promising biomarker for NSCLC. However, the effectiveness of ctDNA detection in guiding ADT for resectable stage I NSCLC patients remains elusive. This study aimed to elucidate the role of ctDNA detection in estimating prognosis and guiding ADT for resectable stage I NSCLC patients.

Methods: Individual patient data and ctDNA results data were collected from 270 patients across four independent cohorts. The detection of ctDNA was conducted at 3 days to 1 month after surgery. The endpoint for this study was relapse-free survival (RFS) and overall survival (OS).

Results: Of the 270 resectable stage I NSCLC patients, 9 patients with ctDNA-positive and 261 patients with ctDNA-negative. We found that the risk of recurrence was significantly lower in the ctDNA-negative group compared to the ctDNA-positive group(HR=0.11, p<0.0001). However, there is no difference in the risk of death between the two groups (p =0.39). In the ctDNA-positive group, there were no significant differences in RFS between patients who received ADT and patients who did not receive ADT (p =0.58). In the ctDNA-negative group, those who received ADT had a worse RFS in comparison with those who did not receive ADT (HR=2.36, p =0.029). No difference in OS was seen between patients who received ADT and patients who did not receive ADT in both the ctDNA-positive group and the ctDNA-negative group (All p values>0.05). Furthermore, there was no difference in RFS and OS between patients who received chemotherapy-based or tyrosine kinase inhibitor-based ADT and patients who did not receive ADT in both the ctDNA-positive group and the ctDNA-negative group (All p values>0.05).

Conclusions: Postoperative ctDNA detection can be a prognostic marker to predict recurrence but has limited effects in guiding ADT for resectable stage I NSCLC. Future prospective investigations are needed to verify these results.

KEYWORDS

circulating tumor DNA, NSCLC, prognosis, recurrence, adjuvant therapy

Introduction

Non-small cell lung cancer (NSCLC) is the most common type of lung cancer, and a significant effort has been devoted in recent years to improving its diagnosis and treatment (1). To our relief, the rate of patients with lung cancer diagnosed at a localized stage increased from 17% in 2013 to 28% in 2018 according to the Cancer Statistics 2022, representing more patients diagnosed at early stages (1). Stage I NSCLC is a common early stages disease with the most promising potentially cured by lobectomy with mediastinal lymphadenectomy (2, 3). However, there are still 20%–40% of stage I NSCLC patients develop postoperative tumor recurrence (4). Substantial evidence has implicated postoperative minimal residual disease (MRD) is a major contributor to postoperative recurrence, and systemic adjuvant therapy (ADT) after surgery for these patients is a rational strategy to eliminate MRD to improve disease outcomes (5–7). Evidence from Lung Adjuvant Cisplatin Evaluation (LACE) meta-analysis suggests that early-stage NSCLC treated with adjuvant chemotherapy versus no treated has improved only modestly with absolute benefits of 5.4% at 5 years, which suggests "treat many to save few" is commonly encountered in clinical practice (8). For resectable stage I NSCLC patients, the US National Comprehensive Cancer Network (NCCN) guidelines recommend the use of adjuvant chemotherapy with stage IB NSCLC with high-risk factors such as poorly differentiated tumors, vascular invasion, and visceral pleural invasion with acknowledging that "these factors independently may not be an indication" (2). This information highlights the urgent need to identify predictive biomarkers which may assist in selecting patients that may benefit from ADT and thus avoid over- or undertreatment.

Recently, circulating tumor DNA (ctDNA) has emerged as a promising noninvasive biomarker that is highly sought after (9, 10). Multiple prospective studies reported that ctDNA detection has the capacity to identify MRD after therapy in various human tumors (11–14). Evidence from TRACERx studies indicates that ctDNA detection could accurately detect MRD in patients who relapsed before their disease was picked up by standard imaging (15). The findings cited above support further exploring whether ctDNA detection can be employed in treatment stratification and prognosis assessment in NSCLC patients. Several studies show that postoperative detectable ctDNA is associated with a very high risk of recurrence and help to select patients who would most benefit from adjuvant therapy (ADT) in early-stage NSCLC patients (16–18). However, the value of ctDNA in resectable stage I NSCLC is still debated.

Recent evidence shows that postoperative ctDNA positivity is significantly correlated with an increased risk of recurrence and death in stage I NSCLC (18, 19). However, several pieces of evidence suggest that some patients with stage I NSCLC either do not release ctDNA or release ctDNA at frequencies below the limit of detection of current technologies and thus ctDNA detection is unlikely to be widely adopted for NSCLC patients with stage I disease (5). Furthermore, most of the previous work focused on the predictive effect of ctDNA detection in guiding ADT in stage II-III disease due to a few stage I NSCLC patients treated with adjuvant therapy (16, 18). Clinical trials testing postoperative ctDNA status as biomarkers to guide prognostic stratification and ADT decisions for stage I NSCLC patients are underway (NCT05079022; NCT04585477). In view of this, understanding the role of postoperative ctDNA status in stage I NSCLC have important implications for clinical care and future clinical trials.

To address this knowledge gap, we conducted a pooled analysis to investigate the role of ctDNA detection in estimating prognosis and guiding ADT for resectable stage I NSCLC patients. We further analyzed whether postoperative ctDNA detection could be applied to guide different ADT strategies in resectable stage I NSCLC patients to guide clinicians in selecting patients who benefit most from particular ADT regimens.

Materials and methods

Patients

Full details of included stage I NSCLC from four cohorts have been published previously (16, 17, 20, 21). The theme of all four included studies was explored in the predictive prognostic value of ctDNA detection in resectable NSCLC patients (16, 17, 20, 21). Data of all patients from four studies were pooled and then ascertained which patients satisfied the necessary inclusion criteria as follows (1): Stage I NSCLC patients, which were defined by pathologic stage or clinical stage, underwent surgical excision of the primary tumor without neoadjuvant therapy; (2) Patients underwent ctDNA detection during collected 3 days to 1 month after surgery with available postoperative ctDNA detection status; (3) Data on ADT and at last follow up and/or date of relapse and death were available. Informed written consent from all study subjects and approval from ethics committees was obtained from each study center.

Study design

The detection time of ctDNA was during 3 days to 1 month after surgery, given perioperative dynamic changes in ctDNA and the time to initiation of ADT. The ctDNA detection of included cohorts was performed based on next-generation high-throughput sequencing

(NGS) and did not consider NGS panel size and content. Detailed methods of ctDNA detection are provided as follows: Peng's cohort, Circulating Single-molecule Amplification and Resequencing technology consisting of 127 genes; Qiu's cohort: Automated Triple Groom Sequencing Technology consisting of 139 genes; Xia's cohort: NGS panel spanning 769 cancer-related genes; Gale's cohort: Personalized RaDaR^{TM} assays. Postoperative ctDNA status was categorized as detectable (ctDNA-positive) or undetectable (ctDNA-negative) based on the result of a single detection. The endpoint for this study was relapse-free survival (RFS) and overall survival (OS). RFS was defined as the time from surgery to disease recurrence due to any cause. OS was defined as the interval between the time of surgery and death.

Statistical analysis

Differences in the distribution of categorical variables among different groups were analyzed using Fisher's exact or chi-squared test. Kaplan–Meier curves of RFS and OS were compared using the standard log-rank test. Univariable and multivariable Cox proportional hazards regression models were used to evaluate proposed prognostic factors. Results are reported as hazard ratios (HR) with 95% confidence intervals (CI). All analyses were performed using SPSS, version 22 (IBM Corp) and R, version 4.1.3 (R Project for Statistical Computing). Two-sided $p \leq 0.05$ was considered statistically significant.

Results

Overview of patient cohort

Figure 1 summarizes the working hypotheses and workflow in this study. A total of 270 stage I NSCLC patients from 4 cohorts who had undergone surgical tumor resection and known postoperative ctDNA status were eventually enrolled (16, 17, 20, 21). The large majority of included patients were nonsmokers, pathologically confirmed adenocarcinoma, had T1 stage disease and did not receive postoperative adjuvant therapy. The patients were assigned to the positive and negative groups stratified by postoperative ctDNA status. Of the 270 patients included in this study, 9 had ctDNA-positive (positive rate, 3.33%). The clinicopathological patient characteristics are shown in Table 1. There were no significant differences in the baseline clinicopathologic characteristics between ctDNA-positive and ctDNA-negative groups.

Postoperative ctDNA detection for predicting recurrence and death

To gain further insight into the potential predictive and prognostic role of postoperative ctDNA detection in resectable stage I NSCLC, we first compared the risk of recurrence between ctDNA-positive and ctDNA-negative groups. Our results indicated that 66.66% (6 of 9) ctDNA-positive patients observed relapse, whereas 12.64% (33 of 261) ctDNA-negative patients experienced relapse (Figure 2A, $p < 0.001$). The Kaplan-Meier (KM) curve showed that relapse risks were higher in the ctDNA-positive group compared with the ctDNA-negative group (HR=0.11, 95% CI:0.05-0.26, $p<0.0001$, Figure 2B). To further explore the relationship between ctDNA status and the risk of recurrence, we conducted univariate and multivariate analyses. The results indicated that ctDNA-positive and T stage was an independent prognostic factor with RFS (Figure 3). We further investigated the differences in the risk of death between ctDNA-positive and ctDNA-negative groups. We observed that there were no deaths among the 4 ctDNA-positive patients, and 7 out of 42 ctDNA-negative patients died (Figure 2C, $p = 1.00$). Furthermore, there were no significant differences in the risk of death between the ctDNA-positive and ctDNA-negative groups (Figure 2D, $p = 0.39$).

Postoperative ctDNA detection for guiding adjuvant therapy

To assess the role of ctDNA detection in directing ADT for resectable stage I NSCLC, we compared prognoses between patients

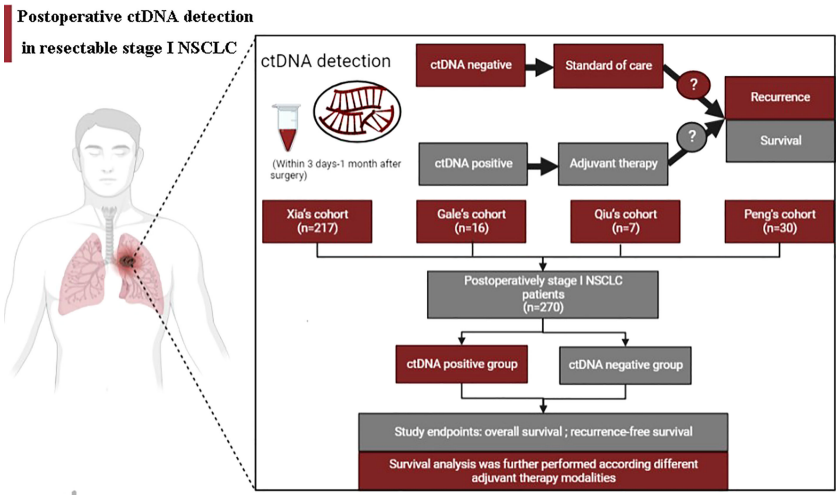


FIGURE 1
Schematic illustration of working hypotheses and workflow.

10.3389/fonc.2023.1083417

TABLE 1 The clinicopathologic characteristics of patients.

Characteristics	ctDNA- (N=261)	ctDNA+ (N=9)	Total (N=270)	p-value
Age (years)				0.32
<60 years	122(46.39%)	6(2.28%)	128(48.67%)	
≥ 60 years	132(50.19%)	3(1.14%)	135(51.33%)	
Sex				0.51
Female	135(51.33%)	6(2.28%)	141(53.61%)	
Male	119(45.25%)	3(1.14%)	122(46.39%)	
Smoker				1
No	180(66.67%)	6(2.22%)	186(68.89%)	
Yes	81(30.00%)	3(1.11%)	84(31.11%)	
Tumor site				1
Left lung	121(47.64%)	4(1.57%)	125(49.21%)	
Right lung	125(49.21%)	4(1.57%)	129(50.79%)	
T stage				0.06
T1	198(77.95%)	4(1.57%)	202(79.53%)	
T2	48(18.90%)	4(1.57%)	52(20.47%)	
Histology subtype				0.42
LUAD	224(82.96%)	7(2.59%)	231(85.56%)	
LUSC	33(12.22%)	2(0.74%)	35(12.96%)	
Other	4(1.48%)	0(0%)	4(1.48%)	
Adjuvant therapy				0.1
No	228(84.44%)	6(2.22%)	234(86.67%)	
Yes	33(12.22%)	3(1.11%)	36(13.33%)	

ctDNA+, ctDNA positive; ctDNA-, ctDNA negative; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma.

treated with ADT and patients who did not receive ADT in ctDNA-positive and ctDNA-negative groups separately. In the ctDNA-positive group,3 of 9 patients received ADT. Bar chart representing clinical features of patients treated with ADT and patients who did not receive ADT, and the groups were not different in their baseline characteristics (Figure 4A). There was no significant difference in relapse risk between patients treated and untreated ADT in the ctDNA-positive group (Figure 4B). In the ctDNA-negative group, most of the patients who did not receive ADT had T1 stage disease, and no differences were found in other clinicopathologic characteristics between patients treated and untreated ADT (Figure 4C). Our results indicated that ctDNA-negative patients who were treated with ADT were found to be at higher risk of relapse than patients who did not (HR=2.36, 95% CI:1.07-5.36, p=0.029, Figure 4D).

We further investigated the risk of mortality according to whether or not received ADT in ctDNA-positive and ctDNA-negative patients. Baseline characteristics of patients who received ADT and those who did not show significant differences in both ctDNA-positive and ctDNA-negative patients Figures 5A, C). There were no statistical differences observed in the risk of death between treated and treated ADT in both ctDNA-positive and ctDNA-negative patients (All p values >0.05, Figures 5B, D).

Postoperative ctDNA detection for guiding different adjuvant therapy strategies

Analyses based on the different ADT types were performed to further understand the role of ctDNA detection in the guided choice of different ADT types. One ctDNA-negative patient who received both CT and TKI ADT was included in CT and TKI ADT subgroup analysis, respectively. Figure 6 displays no significant difference in risk of relapse between treated and untreated CT-based/TKI-based ADT in both ctDNA-positive and ctDNA-negative patients (All p values >0.05, Figure 6). We further analyzed the differences in death risk based on types of adjuvant therapy in ctDNA-positive and ctDNA-negative patients. In the ctDNA-positive group, only one patient received TKI-based ADT. There was no significant risk reduction in patients treated with TKI-based ADT compared with those not treated with ADT (p =1, Figure 5). In the ctDNA-negative group, 4 patients received CT-based ADT. We found no difference in death risk between patients treated with CT-based ADT and those who did not treat with ADT (p =0.42, Figure 5).

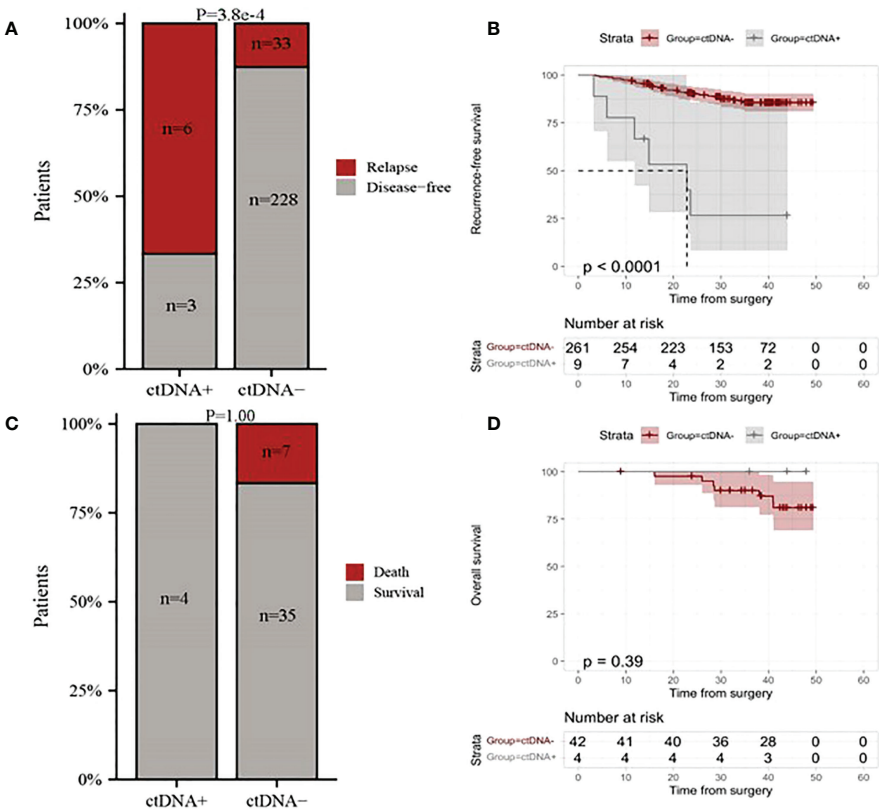
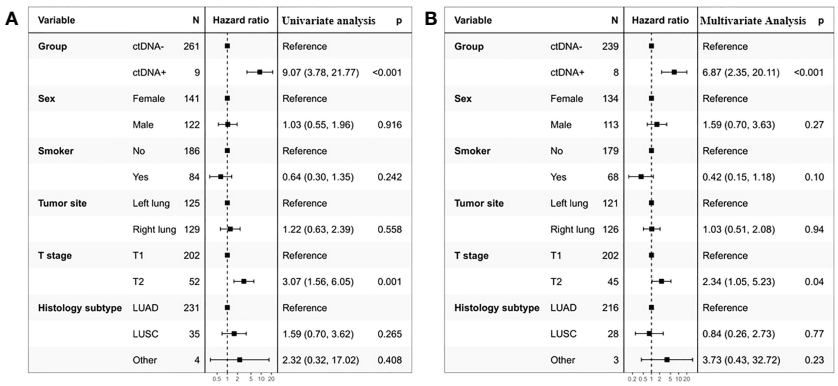


FIGURE 2
The relationship between postoperative ctDNA status and the risk of recurrence and death in resectable stage I NSCLC patients. **(A)** Comparison of overall relapse proportion between ctDNA-positive and ctDNA-negative patients. **(B)** Kaplan-Meier curves for recurrence-free survival (RFS) for patients with NSCLC stratified by the postoperative ctDNA status. **(C)** Comparison of deaths proportion between ctDNA-positive patients and ctDNA-negative patients. **(D)** Kaplan-Meier curves for overall survival for patients with NSCLC stratified by the postoperative ctDNA status. ctDNA+, ctDNA positive; ctDNA-, ctDNA negative.

Discussion

Studies repeatedly demonstrate that ctDNA may serve as a promising biomarker for recurrence (18, 22–25). Notably, several more recent studies have also demonstrated that ctDNA can provide guidance on which patients to treat or not treat with adjuvant chemotherapy injection in resectable NSCLC (17, 18). However, the precise role of postoperative ctDNA detection in stage I NSCLC patients who underwent resection remains unclear. Here, we report a pooled analysis from four cohort studies, focusing on novel endpoints such as the significance of single ctDNA detection, any associations with risk of recurrence and death, and whether they could be used as biomarkers to guide the different types of ADT.

In view of the time to initiate ADT, patient compliance and the association between postsurgical ctDNA status and the prognosis, we



A

Variable		N	Hazard ratio	Univariate analysis	p
Group	ctDNA-	261		Reference	
	ctDNA+	9		9.07 (3.78, 21.77)	<0.001
Sex	Female	141		Reference	
	Male	122		1.03 (0.55, 1.96)	0.916
Smoker	No	186		Reference	
	Yes	84		0.64 (0.30, 1.35)	0.242
Tumor site	Left lung	125		Reference	
	Right lung	129		1.22 (0.63, 2.39)	0.558
T stage	T1	202		Reference	
	T2	52		3.07 (1.56, 6.05)	0.001
Histology subtype	LUAD	231		Reference	
	LUSC	35		1.59 (0.70, 3.62)	0.265
	Other	4		2.32 (0.32, 17.02)	0.408

B

Variable		N	Hazard ratio	Multivariate Analysis	p
Group	ctDNA-	239		Reference	
	ctDNA+	8		6.87 (2.35, 20.11)	<0.001
Sex	Female	134		Reference	
	Male	113		1.59 (0.70, 3.63)	0.27
Smoker	No	179		Reference	
	Yes	68		0.42 (0.15, 1.18)	0.10
Tumor site	Left lung	121		Reference	
	Right lung	126		1.03 (0.51, 2.08)	0.94
T stage	T1	202		Reference	
	T2	45		2.34 (1.05, 5.23)	0.04
Histology subtype	LUAD	216		Reference	
	LUSC	28		0.84 (0.26, 2.73)	0.77
	Other	3		3.73 (0.43, 32.72)	0.23

FIGURE 3
Univariate **(A)** and multivariate **(B)** Cox regression analyses of the association between patient's characteristics and the probability of recurrence-free survival (RFS) in ctDNA-positive group.

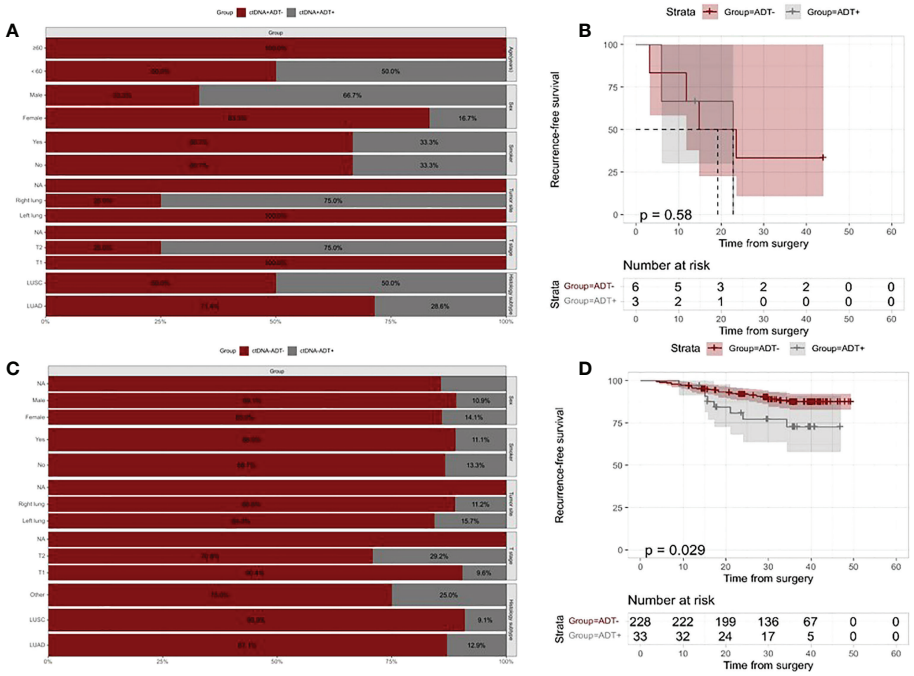


FIGURE 4
Kaplan–Meier analysis of recurrence-free survival according to ADT and postoperative ctDNA status. **(A)** The distribution of clinicopathologic characteristics for patients who received vs patients who did not receive ADT in the ctDNA-positive group. **(B)** Kaplan-Meier curve showing RFS stratified by ADT in the ctDNA-positive group. **(C)** The distribution of clinicopathologic characteristics for patients who received vs patients who did not receive ADT in the ctDNA-negative group. **(D)** Kaplan-Meier curve showing RFS stratified by ADT in the ctDNA-negative group.

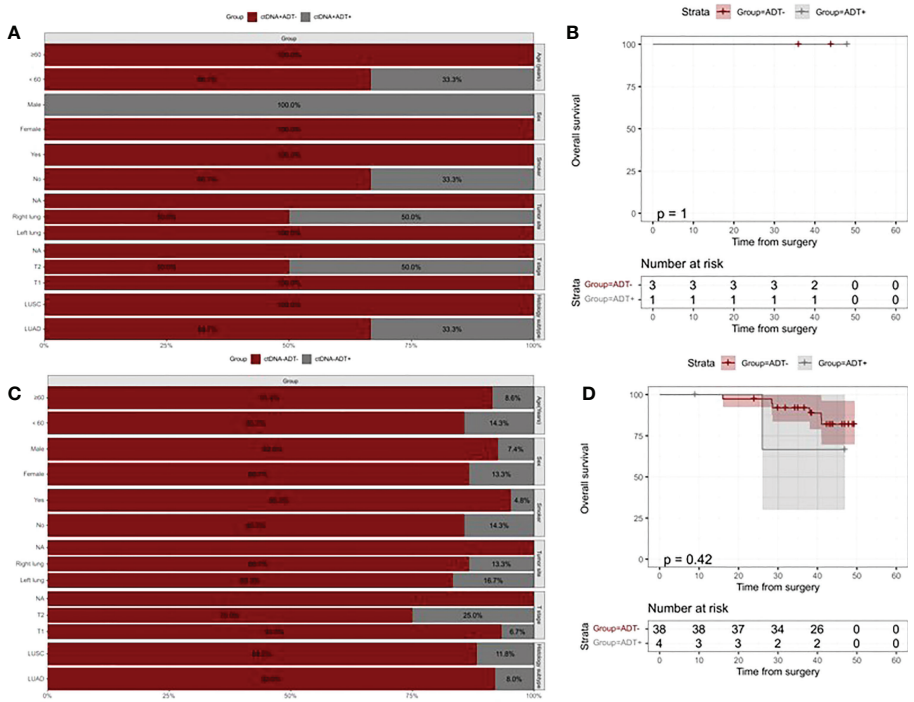


FIGURE 5
Kaplan–Meier analysis of overall survival according to ADT and postoperative ctDNA status. **(A)** The distribution of clinicopathologic characteristics for patients who received vs patients who did not receive ADT in the ctDNA-positive group. **(B)** Kaplan-Meier curve showing RFS stratified by ADT in the ctDNA-positive group. **(C)** The distribution of clinicopathologic characteristics for patients who received vs patients who did not receive ADT in the ctDNA-negative group. **(D)** Kaplan-Meier curve showing RFS stratified by ADT in the ctDNA-negative group.

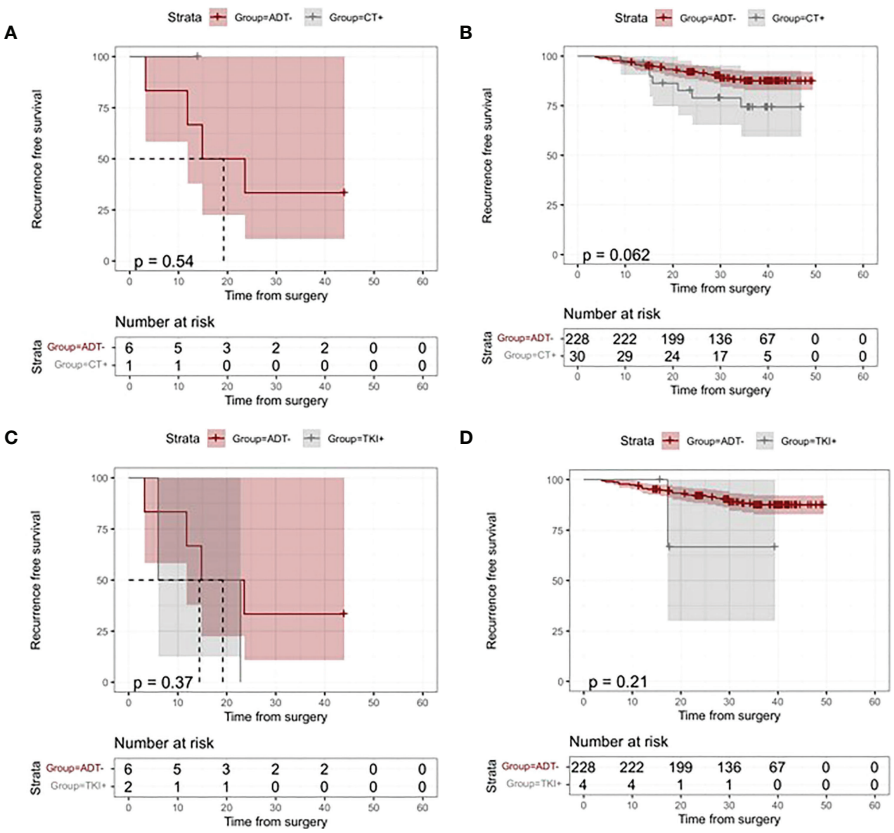


FIGURE 6
Kaplan–Meier analysis of recurrence-free survival (RFS) according to adjuvant treatment modalities and ctDNA status. **(A)** Kaplan-Meier Curves of RFS for patients who received vs patients who did not receive chemotherapy (CT)-based ADT in the ctDNA-positive group. **(B)** Kaplan-Meier Curves of RFS for patients who received vs patients who did not receive CT-based ADT in the ctDNA negative group. **(C)** Kaplan-Meier Curves of RFS for patients who received vs patients who did not receive adjuvant tyrosine kinase inhibitor (TKI) in the ctDNA-positive group. **(D)**Kaplan-Meier Curves of RFS for patients who received vs patients who did not receive TKI-based ADT in the ctDNA-positive group.

focused on single ctDNA detection at postoperative 3 days-1 month in the present study (26, 27). In this study, we first analyzed the postoperative ctDNA-positive detection rate and found only 3.33% of all 270 included patients had ctDNA detected, which was lower than in other studies. The main reasons for the discrepancy can be attributed to variations in the definition of ctDNA-positive. Previous studies examined ctDNA status at multiple time points, and patients with detected ctDNA at one arbitrary time point were defined as ctDNA-positive. However, our findings are based on single ctDNA detection at fixed periods. Furthermore, the association of ctDNA levels with tumor burden has been observed in several studies (15, 26).Thus, stage I NSCLC with low tumor burden had limited ability to detect ctDNA, which is further confirmed by the results shown in our research.

Next, we investigated the prognostic significance of ctDNA status in resectable stage I NSCLC patients. Postoperative ctDNA status was a strong predictive factor for RFS in both univariate and multivariate analysis, further validating previous findings (18, 19). Postoperative ctDNA-positive patients are associated with greater recurrence risk than ctDNA-negative patients. It is worth noting that we observed postoperative ctDNA status was not significantly associated with the risk of death in resectable stage I NSCLC patients. Owing to the low numbers provided by OS information in this study, future studies are required to determine the association between postoperative ctDNA status and death risk.

For additional insights into whether ctDNA status could be used to guide ADT, we compared stage I NSCLC patients' outcomes stratified by whether they received ADT in the ctDNA-positive group. Our results show no significant difference in relapse and death risk between patients treated and untreated with ADT in the ctDNA-positive group. We further performed a statistical analysis stratified on ADT modalities to further verify and demonstrate the above finding. In the ctDNA-positive group, patients treated with CT-based/TKI-based ADT did not alter the risk for relapse and death compared with patients who did not receive ADT. The above results suggested that postoperative ctDNA status based on single detection is not likely to be useful for helping select patients that can benefit from CT-based/TKI-based ADT.

In the ctDNA-negative group, we observed that patients treated with ADT have a higher recurrence risk than patients untreated with ADT. In this study, ADT was administered following standard clinical guidelines based on prognostic stratification by TNM stage and high-

risk factors recommended by the guidelines in the majority of the patient. Thus, most stage IB patients with high-risk factors such as stage T2 disease received ADT, whereas most patients without high-risk factors are free of ADT. Not surprisingly, patients treated with ADT present a higher recurrence risk instead, which is consistent with recent work by Xia et al (16). Furthermore, patients treated with CT-based/TKI-based ADT did not alter the risk for relapse and death compared with patients who did not receive ADT in the ctDNA-negative group.

Ideally, a sensitive enough biomarker could identify high-risk patients who benefit from ADT and low-risk patients who avoid ADT without decreasing the likelihood of prognosis benefit. In light of these findings, postoperative ctDNA status is not a reliable biomarker to guide ADT in resectable stage I NSCLC patients. Evidence from DYNAMIC and TRACER study's showed stage I NSCLC patients have a low tumor burden and limited capacity to detect ctDNA using current ctDNA approaches, which may diminish sensitivity to identify high-risk patients (15, 26). This may be the primary cause limited impact of ctDNA detection in guiding ADT. In view of this, postoperative single ctDNA detection does not improved treatment selection and how to improve the sensitivity of ctDNA detection in stage I NSCLC is a topic worth exploring.

Three major findings are described in this report. Firstly, the postoperative ctDNA-positive detection rate was lower in stage I NSCLC using current technologies and efforts to improve levels of ctDNA detection in this setting are warranted. Secondly, postoperative ctDNA positive was an independent predictor of relapse but not death. Thirdly, the instructive effects of postoperative ctDNA status on the selection of ADT were limited. Given the lower postoperative ctDNA-positive rate and limited effects of ctDNA detection in guiding ADT, we suggest that ctDNA detection used to guide ADT should be considered cautiously in resectable stage I NSCLC. Several ongoing clinical trials explore the instructive effects of postoperative ctDNA status in guiding ADT for resectable NSCLC patients (NCT05167604; NCT05286957; NCT04966663). We believe this is a novel study that provides the preliminary data for avoiding unnecessary workup and increased costs and may be integrated into future trial designs.

Several limitations should also be considered. Firstly, this study is limited by small sample size, short follow-up and inconsistent ctDNA detection time points. Secondly, the size of the panel design and choice of NGS platform for ctDNA detection could have affected the significance of the results. Thirdly, our result mainly targeted CT-based/TKI-based ADT. Fourthly, we pooled data from four cohorts, which may have introduced bias due to differences in study designs.

Conclusion

Our findings suggest that postoperative ctDNA detection can be predictive for relapse but has limited effects in guiding ADT in resectable stage I NSCLC. With these findings, ctDNA detection used to guide ADT should be considered cautiously, and reliable prognostic biomarkers are needed to identify patients at high risk for recurrence to guide ADT in resectable stage I NSCLC.

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

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

BW, JL, and JY were involved in the literature search, figures, study design, data collection, data analysis, writing and article review. BZ, SX, CZ, JP, SW, and KZ were involved in the literature search and data collection. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Academic Promotion Program of Shandong First Medical University (2019ZL002), Research Unit of Radiation Oncology, Chinese Academy of Medical Sciences (2019RU071), the foundation of National Natural Science Foundation of China (81627901, 81972863, 82030082 and 8220103505), the foundation of Natural Science Foundation Of Shandong (ZR201911040452 and ZR2022QH017).

Acknowledgments

We thank all patients that were involved in this study.

Conflict of interest

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

Publisher's note

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

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin* (2022) 72(1):7–33. doi: 10.3322/caac.21708

2. Ettinger DS, Wood DE, Aisner DL, Akerley W, Bauman JR, Bharat A, et al. Non–small cell lung cancer, version 3.2022, NCCN clinical practice guidelines in oncology. *J Natl Compr Cancer Netw* (2022) 20(5):497–530. doi: 10.6004/jnccn.2022.0025

3. Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WE, et al. The IASLC lung cancer staging project: Proposals for revision of the TNM stage groupings in the forthcoming (Eighth) edition of the TNM classification for lung cancer. *J Thorac Oncol* (2016) 11(1):39–51. doi: 10.1016/j.jtho.2015.09.009

4. Vansteenkiste J, Crino L, Dooms C, Douillard JY, Faivre-Finn C, Lim E, et al. 2nd ESMO consensus conference on lung cancer: early-stage non-small-cell lung cancer consensus on diagnosis, treatment and follow-up. *Ann Oncol* (2014) 25(8):1462–74. doi: 10.1093/annonc/mdu089

5. Abbosh C, Birkbak NJ, Swanton C. Early stage NSCLC - challenges to implementing ctDNA-based screening and MRD detection. *Nat Rev Clin Oncol* (2018) 15(9):577–86. doi: 10.1038/s41571-018-0058-3

6. Chae YK, Oh MS. Detection of minimal residual disease using ctDNA in lung cancer: Current evidence and future directions. *J Thorac Oncol* (2019) 14(1):16–24. doi: 10.1016/j.jtho.2018.09.022

7. Wu M, Shen H, Wang Z, Kanu N, Chen K. Research progress on postoperative Minimal/Molecular residual disease detection in lung cancer. *Chronic Dis Transl Med* (2022) 8(2):83–90. doi: 10.1002/cdt3.10

8. Pignon JP, Tribodet H, Scagliotti GV, Douillard JY, Shepherd FA, Stephens RJ, et al. Lung adjuvant cisplatin evaluation: a pooled analysis by the LACE collaborative group. *J Clin Oncol* (2008) 26(21):3552–9. doi: 10.1200/JCO.2007.13.9030

9. Pellini B, Chaudhuri AA. Circulating tumor DNA minimal residual disease detection of non-Small-Cell lung cancer treated with curative intent. *J Clin Oncol* (2022) 40(6):567–75. doi: 10.1200/JCO.21.01929

10. Frankell A, Jamal-Hanjani M. When can we be confident of surgical cure with ctDNA? *Nat Rev Clin Oncol* (2022) 19(9):571–2. doi: 10.1038/s41571-022-00664-8

11. Pantel K, Alix-Panabieres C. Liquid biopsy and minimal residual disease - latest advances and implications for cure. *Nat Rev Clin Oncol* (2019) 16(7):409–24. doi: 10.1038/s41571-019-0187-3

12. Pellini B, Pejovic N, Feng W, Earland N, Harris PK, Usmani A, et al. ctDNA MRD detection and personalized oncogenomic analysis in oligometastatic colorectal cancer from plasma and urine. *JCO Precis Oncol* (2021) 5:378–88. doi: 10.1200/PO.20.00276

13. Parikh AR, Van Seventer EE, Siravegna G, Hartwig AV, Jaimovich A, He Y, et al. Minimal residual disease detection using a plasma-only circulating tumor DNA assay in patients with colorectal cancer. *Clin Cancer Res* (2021) 27(20):5586–94. doi: 10.1158/1078-0432.CCR-21-0410

14. Ueberroth BE, Jones JC, Bekaii-Saab TS. Circulating tumor DNA (ctDNA) to evaluate minimal residual disease (MRD), treatment response, and posttreatment prognosis in pancreatic adenocarcinoma. *Pancreatology* (2022) 22(6):741–8. doi: 10.1016/j.pan.2022.06.009

15. Abbosh C, Birkbak NJ, Wilson GA, Jamal-Hanjani M, Constantin T, Salari R, et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature* (2017) 545(7655):446–51. doi: 10.1038/nature22364

16. 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

17. Qiu B, Guo W, Zhang F, Lv F, Ji Y, Peng Y, et al. Dynamic recurrence risk and adjuvant chemotherapy benefit prediction by ctDNA in resected NSCLC. *Nat Commun* (2021) 12(1):6770. doi: 10.1038/s41467-021-27022-z

18. Zhang JT, Liu SY, Gao W, Liu SM, Yan HH, Ji L, et al. Longitudinal undetectable molecular residual disease defines potentially cured population in localized non-small cell lung cancer. *Cancer Discovery* (2022) 12(7):1690–701. doi: 10.1158/2159-8290.CD-21-1486

19. Yang W, You N, Jia M, Yeung SJ, Ou W, Yu M, et al. Undetectable circulating tumor DNA levels correlate with low risk of recurrence/metastasis in postoperative pathologic stage I lung adenocarcinoma patients. *Lung Cancer* (2020) 146:327–34. doi: 10.1016/j.lungcan.2020.06.009

20. Peng M, Huang Q, Yin W, Tan S, Chen C, Liu W, et al. Circulating tumor DNA as a prognostic biomarker in localized non-small cell lung cancer. *Front Oncol* (2020) 10:561598. doi: 10.3389/fonc.2020.561598

21. 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

22. Yue D, Liu W, Chen C, Zhang T, Ma Y, Cui L, et al. Circulating tumor DNA predicts neoadjuvant immunotherapy efficacy and recurrence-free survival in surgical non-small cell lung cancer patients. *Transl Lung Cancer Res* (2022) 11(2):263–76. doi: 10.21037/tlcr-22-106

23. Li N, Wang BX, Li J, Shao Y, Li MT, Li JJ, et al. Perioperative circulating tumor DNA as a potential prognostic marker for operable stage I to IIIA non-small cell lung cancer. *Cancer* (2022) 128(4):708–18. doi: 10.1002/cncr.33985

24. Kuang PP, Li N, Liu Z, Sun TY, Wang SQ, Hu J, et al. Circulating tumor DNA analyses as a potential marker of recurrence and effectiveness of adjuvant chemotherapy for resected non-Small-Cell lung cancer. *Front Oncol* (2020) 10:595650. doi: 10.3389/fonc.2020.595650

25. Ohara S, Suda K, Sakai K, Nishino M, Chiba M, Shimoji M, et al. Prognostic implications of preoperative versus postoperative circulating tumor DNA in surgically resected lung cancer patients: a pilot study. *Transl Lung Cancer Res* (2020) 9(5):1915–23. doi: 10.21037/tlcr-20-505

26. Chen K, Zhao H, Shi Y, Yang F, Wang LT, Kang G, et al. Perioperative dynamic changes in circulating tumor DNA in patients with lung cancer (DYNAMIC). *Clin Cancer Res* (2019) 25(23):7058–67. doi: 10.1158/1078-0432.CCR-19-1213

27. Wang B, Pei J, Wang S, Cheng K, Yu J, Liu J. Prognostic potential of circulating tumor DNA detection at different time periods in resectable non-small cell lung cancer: Evidence from a meta-analysis. *Crit Rev Oncol Hematol* (2022) 177:103771. doi: 10.1016/j.critrevonc.2022.103771

frontiers | Frontiers in Oncology

TYPE Hypothesis and Theory
PUBLISHED 14 February 2023
DOI 10.3389/fonc.2023.1052147

Check for updates

OPEN ACCESS

EDITED BY
Ping Zhan,
Nanjing University School of Medicine, China

REVIEWED BY
Jung Hun Oh,
Memorial Sloan Kettering Cancer Center, United States
Hongda Liu,
Nanjing Medical University, China

*CORRESPONDENCE
Linlin Wang
wanglinlinatjn@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

RECEIVED 23 September 2022
ACCEPTED 31 January 2023
PUBLISHED 14 February 2023

CITATION
Li B, Yang L, Jiang C, Yao Y, Li H, Cheng S, Zou B, Fan B and Wang L (2023) Integrated multi-dimensional deep neural network model improves prognosis prediction of advanced NSCLC patients receiving bevacizumab.
Front. Oncol. 13:1052147.
doi: 10.3389/fonc.2023.1052147

COPYRIGHT
© 2023 Li, Yang, Jiang, Yao, Li, Cheng, Zou, Fan 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.

Integrated multi-dimensional deep neural network model improves prognosis prediction of advanced NSCLC patients receiving bevacizumab

Butuo Li[1†], Linlin Yang[1†], Chao Jiang[2,3], Yueyuan Yao[1], Haoqian Li[1], Shuping Cheng[1], Bing Zou[1], Bingjie Fan[1] and Linlin Wang[1*]

[1]Department of Radiation Oncology, Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, Jinan, Shandong, China, [2]Department of Otorhinolaryngology Head and Neck Surgery, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, Shandong, China, [3]Shandong Provincial Hospital, Cheeloo College of Medicine, Shandong University, Jinan, Shandong, China

Background: The addition of bevacizumab was found to be associated with prolonged survival whether in combination with chemotherapy, tyrosine kinase inhibitors or immune checkpoint inhibitors in the treatment landscape of advanced non-small cell lung cancer (NSCLC) patients. However, the biomarkers for efficacy of bevacizumab were still largely unknown. This study aimed to develop a deep learning model to provide individual assessment of survival in advanced NSCLC patients receiving bevacizumab.

Methods: All data were retrospectively collected from a cohort of 272 radiological and pathological proven advanced non-squamous NSCLC patients. A novel multi-dimensional deep neural network (DNN) models were trained based on clinicopathological, inflammatory and radiomics features using DeepSurv and N-MTLR algorithm. And concordance index (C-index) and bier score was used to demonstrate the discriminatory and predictive capacity of the model.

Results: The integration of clinicopathologic, inflammatory and radiomics features representation was performed using DeepSurv and N-MTLR with the C-index of 0.712 and 0.701 in testing cohort. And Cox proportional hazard (CPH) and random survival forest (RSF) models were also developed after data pre-processing and feature selection with the C-index of 0.665 and 0.679 respectively. DeepSurv prognostic model, indicated with best performance, was used for individual prognosis prediction. And patients divided in high-risk group were significantly associated with inferior PFS (median PFS: 5.4 vs 13.1 months, P<0.0001) and OS (median OS: 16.4 vs 21.3 months, P<0.0001).

Conclusions: The integration of clinicopathologic, inflammatory and radiomics features representation based on DeepSurv model exhibited superior predictive accuracy as non-invasive method to assist in patients counseling and guidance of optimal treatment strategies.

KEYWORDS

bevacizumab, non-squamous NSCLC, prognosis prediction, deep learning, radiomics

1 Introduction

Bevacizumab, an intravenously administered monoclonal antibody targeting vascular endothelial growth factor (VEGF) pathway, has been proved to have effect on the inhibition of vascular growth, regression of newly formed vessels and normalization of vasculature, thereby facilitate the delivery of cytotoxic chemotherapy (1, 2). The pivotal ECOG4599 demonstrate the efficacy of the addition of bevacizumab to first-line standard chemotherapy in advanced non-squamous NSCLC patients, which can firstly extend the overall survival (OS) to more than 1 year for these patients (3). And the advantages of bevacizumab were also demonstrated in Chinese patient population with the median OS ranging from 17.7 to 24.3 months (4).

With the widespread use of immunotherapy and target therapy, the importance of anti-angiogenesis including bevacizumab seems to be weakened for the first-line treatment of advanced NSCLC patients. It should be noted that the direct action of anti-angiogenesis is vasculature in stroma rather than tumor cells, thus the combination treatment is necessary, and their anti-tumor effect might be magnified in several times (2). The pivotal study IMpower150 also demonstrated the clinical benefit of combination of atezolizumab, bevacizumab and chemotherapy, and the benefit was observed regardless of EGFR and ALK status (5). Thus, bevacizumab likely remains an important role in the treatment landscape for NSCLC patients in the future, especially as a partner in combination treatment strategies. Although the prognostic biomarkers for bevacizumab have been investigated before, such as hypertension, circulating parameters (6), there was still no robust applicable biomarker which is vital for the selection of optimal treatment strategies and individual treatment strategies. Based on the complexity of anti-angiogenesis, multi-dimensional features might exhibit better ability to differentiate survival risk of patients compared with single factor.

The predictive role of clinicopathological and systemic inflammation has been proved in our previous studies (7). Besides, radiomics refers to the highly throughout extraction and analysis of quantitative image features from medical images, and has been used to explore the potential relationship between clinical outcomes and biology of tumor (8). Previous studies have indicated the widespread application of radiomics in lung nodule detection, segmentation, characterization, prognosis prediction and clinical decision making (9).

Deep neural network (DNNs), a subset of artificial intelligence (AI), is an especially promising method that could automatically identify highly intricate and linear/nonlinear associations in data (10), thereby providing evaluations in a quantitative manner. In contrast to machine learning methods, DNNs can realize automated quantification and selection of features and excel at extracting complex features from high-dimensional data and images (11). DNNs have increasingly been deployed in radiomics and development of multi-dimensional models.

The Cox proportional hazard (CPH) regression, which performs the multivariate linear regression between survival time and variables, is the most common survival prediction (12). One limitation of CPH is the linear nature which might result in the neglection of nonlinear relationships between features, while DNN could excel at this task in theory. The potential advantage of DNNs, such as Cox-nnet, DeepSurv (13) and AECOX (14), has been confirmed in predicting prognosis compared to Cox-PH and traditional ML models (15). Yet, the predictive role of DNNs prognostic model based on radiomics is still unclear for advanced NSCLC patients receiving bevacizumab. This study aimed to explore the prognostic effect of radiomics features for advanced NSCLC patients receiving bevacizumab. And we proposed to develop a respective integrated DNN survival model based on three kinds of variables, and perform the comparison of performance between DNNs model and machine learning model to identify the superiority of DNN survival model.

2 Materials and methods

2.1 Patient population

Patients with advanced NSCLC who underwent bevacizumab plus standard chemotherapy in Shandong Cancer Hospital, between July 2014 and October 2019, were enrolled in this study. This study was approved by the Ethics Committee of Shandong Cancer Hospital and was conducted in accordance with the principles of the 1975 Declaration of Helsinki and its later amendments or comparable ethical standards. The inclusion criteria were as follows, (1) radiological and pathological confirmed stage IIIB-IV non-squamous NSCLC; (2) first or second line treated with bevacizumab plus standard chemotherapy for at least two cycles (3 weeks as one cycle); (3) available clinicopathological and hematological data. The exclusion criteria included (1) the synchronization application with target therapy or immunotherapy; (2) combined with other malignancies or hematologic diseases. Moreover, advanced non-squamous NSCLC patients, who met the selection criteria, from

Phase III clinical trials which compared the efficacy between bevacizumab and QL1101-002 or BP102 were retrospectively enrolled into external validation cohort.

2.2 Acquisition of clinical and inflammatory variables

The medical records of each patient were reviewed with respect to consecutive laboratory clinical factors and complete blood count during bevacizumab treatment. All data were acquired retrospectively using uniform database templates to ensure consistent data collection. Specifically, clinical parameters included gender, age, smoking status, EGFR status, anatomical location (central or peripheral), and the presence or absence of liver, brain, or bone metastases.

Inflammatory factors included NLR, PLR, LMR and lactate dehydrogenase (LDH). NLR was defined as the ratio of absolute neutrophil count to absolute lymphocyte count; PLR was the ratio of absolute platelet count to lymphocyte count; LMR was constructed with the ratio of absolute lymphocyte count to absolute monocyte count. The dynamic changes of systemic inflammatory factors were collected during bevacizumab treatment. ROC curves were performed to confirm the cut-off of inflammatory factors.

2.3 Acquisition of CT images

The contrast-enhanced CT images before bevacizumab treatment were extracted from Picture Archiving and Communication Systems using a SOMATOM Definition AS (Siemens Healthineers) for each population in this study. The scanning parameters were as follows: tube voltage, 120 kVp; tube current, 200 mAs; detector, 64 × 0.625 mm; reconstruction thickness, 5 mm; reconstruction interval, 5mm. The CT images were exported and stored in the form of Digital Imaging and Communications for further analysis.

2.4 Patient follow-up and outcomes

The primary endpoint was progression free survival (PFS) and the secondary endpoint was OS. Tumor response was measured according to Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1. PFS and OS were defined as the time from the initiation of bevacizumab to the date of progression and the date of death or last follow-up, respectively. The last follow-up time was until December 2019.

2.5 Training and validation of prognostic model based on the machine learning methods

The region of interest (ROI) in each slice of CT images was defined as the primary lesion of tumor and contoured manually by two radiation oncologists using open-source imaging biomarker explorer (IBEX) software (http://bit.ly/IBEX_MDAnderson) with the window/level of 600/1000 HU. All CT images were countered twice with the interval of about two months in order to reduce the operator bias. The 3D ROI was then preprocessed, and a set of 1041 radiomics features were extracted using IBEX software. All features were divided into nine categories including Intensity, Intensity Histogram, Shape, Gradient Direction Histogram, Gray Level Co-occurrence Matrix 25 (GLCM25), Three-Dimensional Gray Level Co-occurrence Matrix (GLCM3), Two-Dimensional Neighbor Intensity Difference (NID25), Three-Dimensional Neighbor Intensity Difference (NID3), Gray Level Run Length Matrix (GLRLM). Intra-class and Inter-class correlation coefficient (ICC) were calculated for the features extracted from the ROI delineated by two physicians, and features with ICC≥0.75 were screened as stable imaging features. 75% of patients were selected at random to be held out for radiomics feature selection, and validation was done on the remaining 25% of patients.

Univariate and multivariate Cox regression was performed for feature selection among clinicopathological and hematological inflammatory factors. LASSO-Cox analyses were performed to achieve feature dimension reduction for radiomics features. Machine learning prognostic models were performed using CPH and random survival forest (RSF) algorithm.

2.6 Training and validation of multi-dimensional deep neural network survival models

DNN survival models were trained with the network of DeepSurv (16) and Neural Multi-Task Logistic Regression (N-MTLR) (17), and nonlinear variation of parameters was the core of DNN. DeepSurv is a nonlinear extension of CPH model and constructed by using feedforward deep neural network and multilayer perceptron. In addition, the multilayer perceptron is used to estimate the probability of the occurrence of interested events in different time intervals and build the N-MTLR model. Various combinations of hyperparameters were explored in order to optimize the DNN, including batch size, layer of the network and number of neurons of network. The final DNN model was a balance between performance and computing cost, and assigned precise weight to each variable after training and iterations. All patients were divided into high-risk and low-risk group based on the DNN model.

The clinicopathological and inflammatory variables and radiomics features were used as the input of DNN, and PFS time and PFS status were the output. The graphical flow chart of the study was shown in Figure 1.

2.7 Statistical analyses

All patients were randomly divided into training cohort and testing cohort with the split of 70% and 30% once for the train and test of prognostic models. Concordance index (C-index), integrated Brier scores (IBS) and median absolute error (MAE) were used to evaluate and compare the performances of prognostic models of different models. The IBS measures the accuracy of probabilistic predictions of models and the lower IBS indicates the accurate

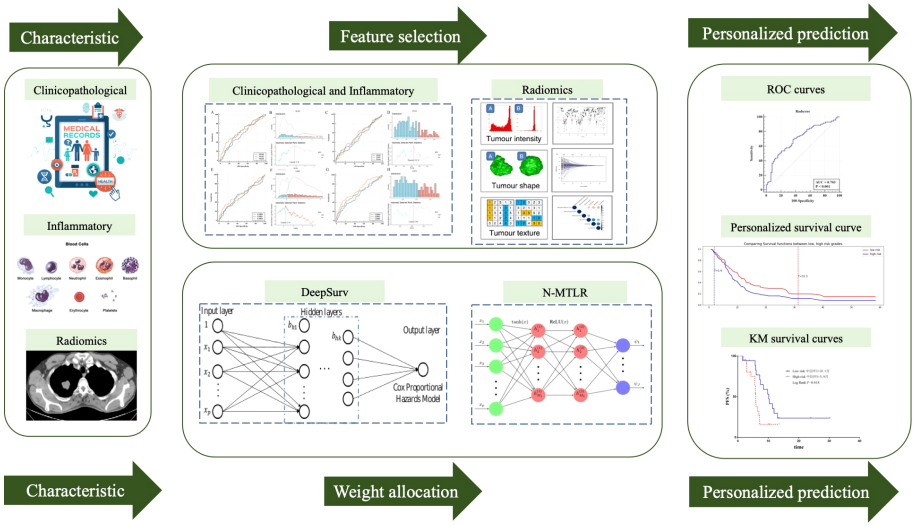


FIGURE 1
Graphic flow chart of development of prognostic model.

predictions. MBE measures the variability between predictions and realities. The evaluation of model was performed three times in testing cohort, and the median of C-index, IBS and MAE were calculated to evaluate the performance.

Kaplan-Meier survival analysis was performed to calculate median survival time and plot the survival curve, and log-rank was used to compare the survival curves. All statistical analyses are two-sided and P value less than 0.5 was considered as significant. Train of DNN is based on the "NumPy" and "SciPy" function, and PyTorch framework, and train of CPH and RSF model is based on "CoxPHModel" and "RandomSurvivalForestModel" function of python 3.6.0.

3 Results

3.1 Patient demographics and characteristics

There were 272 advanced non-squamous NSCLC patients enrolled in this study, and CT images before bevacizumab treatment were available for 195 of them. Entire 195 patients were randomly and automatically split into a training cohort and validation cohort. The baseline detail characteristics of enrolled population were shown in Table 1, and clinicopathological and demographic parameters were balanced between two cohort of patients. 147 patients were included in training cohort, and the median PFS and OS were 8.3 and 26 months respectively.

3.2 Training and validation of prognostic model based on the conventional machine learning methods

3.2.1 Clinicopathological and hematologic inflammatory characteristics preprocessing

There were 220 patients experienced disease progression among 272 patients, and the median PFS was 8.2 months. ROC curves of NLR, PLR, LMR and LDH based on 6-month PFS illustrated that hematologic inflammatory characteristics after 4 cycles treatments were the most predictive, which were included in further feature selection. According to the ROC curves, the value with the maximum Youden index was selected as the cut-off value, and the cut-off values of NLR4, PLR4, LMR4 and LDH4 were set as 2.78, 212.1, 2.18 and 256, respectively (Supplementary Figure 1). All patients were divided into high and low groups.

Univariate and multivariate Cox analysis indicated that smoking history (HR=1.72, P=0.001), anatomical type (HR=1.95, P<0.0001), bone metastasis (HR=1.45, P=0.025), liver metastasis (HR=1.52, P=0.056), NLR4 (HR=1.98, P<0.0001) and LDH4 (HR=1.84, P<0.0001) were the independent prognostic factors for bevacizumab (Table 2).

3.2.2 Radiomic feature pre-processing

Total of 1041 radiomic features were extracted from patients in training cohort. In order to ensure the stability and reproducibility, 740 radiomic features with ICC more than 0.75 were considered as stable features and used in following analyses (Figure 2A). LASSO-Cox analyses were performed to achieve feature dimension reduction (Figure 2B). The model exhibited the optimal performance and the least number of independent variables with the log λ=0.107 (Figure 2C). As the values of λ increased, the LASSO coefficients of these variables were close to zero. As a result, six radiomic features were utilized for the establishment of a prognostic signature. The correlation analysis was further performed to reduce the redundant feature. Finally, Information Measure Corr1, Inverse Variance, Local Std Max, Gauss Area and Spherical Disproportion were selected based on Lasso-Cox analysis (Figure 2D). The Radscore of each patient was calculated according to the weight coefficients of these 5 independent features, with the C-index of 0.65 and AUC of 0.703 after internal validation. (Figure 2E). The result of multivariate cox analysis also showed that the Radscore was the independent prognostic factor for NSCLC patients receiving bevacizumab whether in training cohort and validation cohort (Supplementary Table 1 and Supplementary Table 2).

TABLE 1 Patient Demographics and Characteristics of 195 patients.

Parameters	Training cohort (%)	Validation cohort (%)	P
Age			0.90
≤57	72 (49%)	24 (50%)	
>57	75 (51%)	24 (50%)	
Gender			0.76
Male	82 (55.8%)	28 (58.3%)	
Female	65 (44.2%)	20 (41.7%)	
Smoking History			0.33
No	100 (68%)	29 (60.4%)	
Yes	47 (32%)	19 (39.6%)	
Anatomical type			0.56
Peripheral	107 (72.8%)	37 (77.1%)	
Central	40 (27.2%)	11 (22.9%)	
EGFR			0.31
Wild type	78 (53.1%)	30 (62.5%)	
Sensitive mutation	51 (34.7%)	10 (20.8%)	
Resistance mutation	5 (3.4%)	3 (6.3%)	
NA	13 (8.8%)	5 (10.4%)	
Bone metastasis			0.94
No	94 (63.9%)	31 (64.6%)	
Yes	53 (36.1%)	17 (35.4%)	
Brain metastasis			0.87
No	106 (72.1%)	34 (70.8%)	
Yes	41 (27.9%)	14 (29.2%)	
Liver metastasis			0.85
No	127 (86.4%)	42 (87.5%)	
Yes	20 (13.6%)	6 (12.5%)	

3.2.3 Training and validation of CPH and RSF model

After data pre-processing, smoking history, anatomical type, bone metastasis, liver metastasis, NLR4, LDH4 and Radscore were indicated to be the independent prognostic factors of bevacizumab, and included in the low dimensional feature set. And the CPH model and RSF model were trained in training cohort and estimated in testing cohort. The C-index of CPH prognostic model was 0.665, IBS was 0.13 and MAE was 3.4. The RSF model was trained on the training set with 10,000 trees and the maximum depth of the survival tree is 10. Prediction accuracy was then measured on the test set, with the C-index of 0.679, IBS of 0.14 and MAE of 3.56. The prediction error curve and calibration curve between predicted and actual survival were shown in Figure 3.

3.2.4 Training and validation of prognostic model based on the deep learning methods

Based on the automatic learning characteristics, optimal weight of parameters could be obtained based on deep learning methods, and feature selection was not necessary. DeepSurv and N-MTLR prognostic models were trained based on all parameters included in clinicopathological, inflammatory, and radiomics characteristics. The structure of the final model included one input layer, four hidden layers, and one output layer, and each hidden layer has 100 neurons. Hyperparameters were adjusted in order to achieve the prognostic model with best performance. The He_uniform initialization method applicable to ReLu activation functions was used to initialize weight parameters. Batch normalization was performed to reduce internal variance deviation. Dropout and L2 regularization were applied to reduce overfitting, and the dropout rate was 0.2. Adam optimizer based on gradient was used to obtain stable convergence. After 1000 iterations, the loss value achieved stability gradually, and the final prognostic model was a balance between performance and computing cost. Finally, the C-index of 0.712 and 0.701 was achieved for DeepSurv and N-MTLR prognostic model respectively in testing cohort. The IBS was 0.09 and 0.14, and MAE was 2.6 and 2.4 for DeepSurv and N-MTLR respectively. The prediction error curve and

10.3389/fonc.2023.1052147

TABLE 2 Univariate and multivariate cox analysis of 272 patients.

Parameters	Uni-HR	95% CI	P	Multi-HR	95% CI	P
Age	1.2	0.94-1.6	0.13			
Gender	0.8	0.61-1	0.1			
Smoking History	1.9	1.4-2.5	<0.0001	1.72	1.25-2.37	0.001
Anatomical type	1.3	1.2-1.5	<0.0001	1.95	1.42-2.68	<0.001
EGFR			0.15			
Sensitive mutation	0.74	0.55-1.1	0.057			
Resistance mutation	1.02	0.53-1.9	0.96			
Bone metastasis	1.3	0.97-1.7	0.087	1.42	1.05-1.92	0.025
Brain metastasis	0.94	0.7-1.3	0.67			
Liver metastasis	2.1	1.4-3	0.00012	1.52	0.99-2.34	0.056
NLR4	2.3	1.6-3.1	<0.0001	1.98	1.42-2.74	<0.001
PLR4	1.4	1-1.9	0.031			
LMR4	0.59	0.44-0.8	0.001			
LDH4	1.9	1.4-2.6	<0.0001	1.84	1.32-2.57	<0.001

calibration curve between predicted and actual survival of deep learning prognostic models were shown in Figure 4.

The comparison of performance between machine learning and deep learning prognostic model was illustrated in Figure 5A. All patients were divided into high-risk and low-risk group based on DeepSurv prognostic model (Figure 5B), and the performance was further validated with Kaplan-Meier curves. Patients in high-risk group were significantly associated with inferior PFS (median PFS: 5.4 vs 13.1 months, P<0.0001) and OS (median OS: 16.4 vs 21.3 months, P<0.0001) compared to patients in low-risk group (Figure 5C, D).

3.2.5 Validation of DeepSurv model in external validation cohort

There were 39 advanced non-squamous NSCLC patients from Phase III clinical trials receiving bevacizumab or QL1101-002 or BP102 were enrolled into external validation cohort, and median PFS was 7.1 months (95%CI 5.7-8.5 months). Baseline characteristics were balanced between patients in retrospective and external validation cohort (Table 3). And the Deepsurv model also performed well in external validation cohort, with the C-index of 0.73 and the IBS of 0.15. Patients were also divided into high-risk and low-risk group.

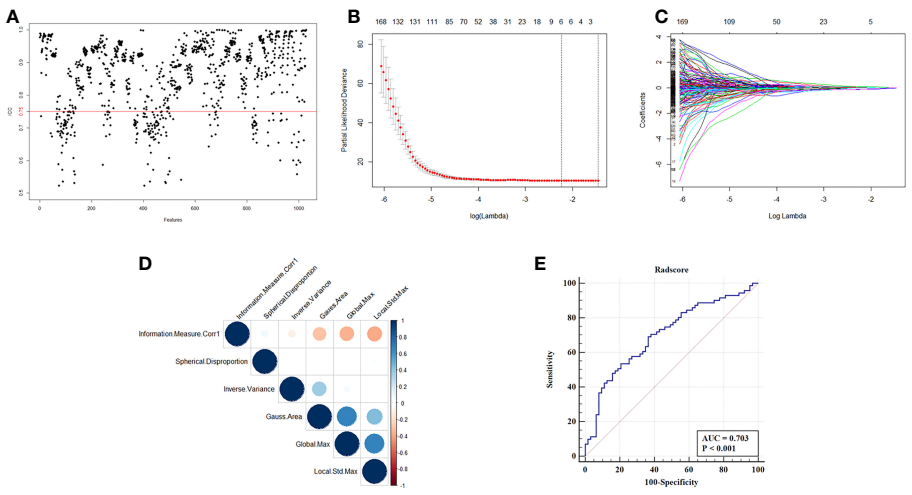


FIGURE 2
Selection of radiomics features and construction of Radscore. **(A)** ICC map of radiomics feature from two independent radiologist. **(B)** Partial likelihood deviance of radiomics features revealed by the LASSO-Cox regression model. The red dots represented the partial likelihood of deviance values, the gray lines represented the standard error (SE), the two vertical dotted lines on the left and right represented optimal values by minimum criteria and 1-SE criteria, respectively. Minimum criteria were used to select host radiomics features in model. **(C)** LASSO coefficient profiles of the survival-related radiomics features. **(D)** Correlation analysis of radiomics features selected by LASSO-Cox regression model. **(E)** ROC curve of Radscore with the AUC of 0.703.

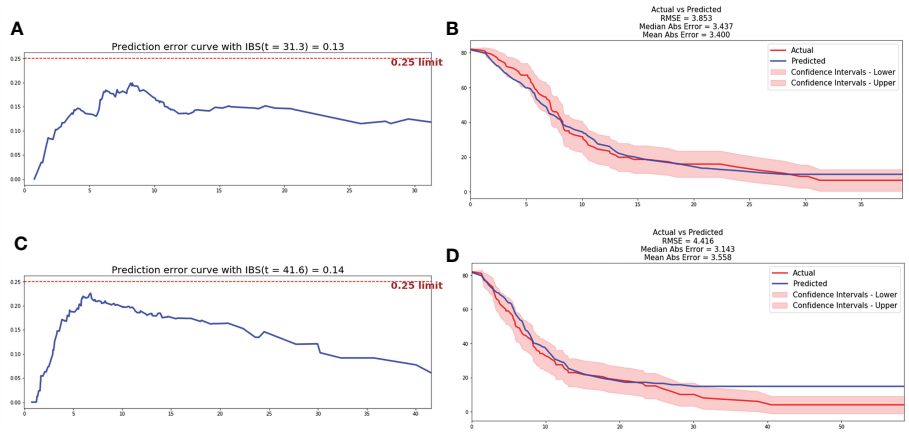


FIGURE 3
Prediction error curve and correction curve of CPH and RSF model. **(A)** Prediction error curve of CPH model. **(B)** Prediction error curve of RSF model. **(C)** Correction curve of CPH model. **(D)** Correction curve of RSF model.

The median PFS was 10.1 months of high-risk patients and was significantly superior to patients in low-risk group, which illustrated the external applicability and generalization ability of DeepSurv prognostic model (Figure 6).

4 Discussion

In this study, we initially demonstrated the predictive role of radiomics for the prognosis of advanced NSCLC patients receiving bevacizumab. More importantly, a robust DNN prognostic model was developed using DeepSurv and N-MTLR algorithm respectively, which could be conveniently used by physicians for accurate prognosis prediction and development of individual treatment strategies for advanced NSCLC patients.

Radiomics is an emerging field in quantitative imaging, and has been widespread adopted in diagnosis, staging and evaluation of clinical outcomes for cancer patients, which might have fantastic application prospects in personalized medicine (8). Radiographic phenotypes were found to be capable of representing the underlying pathophysiology and microenvironment of tumor (18), and thus were more suitable for predicting prognosis and therapeutic response of bevacizumab which acts on the vasculature of tumor. This study firstly indicated the prognostic role of radiomics for bevacizumab treatment in NSCLC patients by development of Radscore using Lasso-Cox. And the Radscore was also found to be the independent factor of bevacizumab.

However, the predictive ability of single type of variables were limited, and clinicopathological and systemic inflammation were also included to develop a prognostic model in this study. CPH and RSF is by far the most commonly used method for survival analysis. While, the setting outcomes of these methods were the linear fitting of covariates, whose intrinsic complex nonlinearities were largely ignored (19). With the extensive application of artificial intelligence in cancer research process, DNN provide a new prospective on non-linear prognostic model by construction of complex correlations through multiple hidden layers. The DNN has been widely used in clinical and translational cancer research including diagnosis, staging,

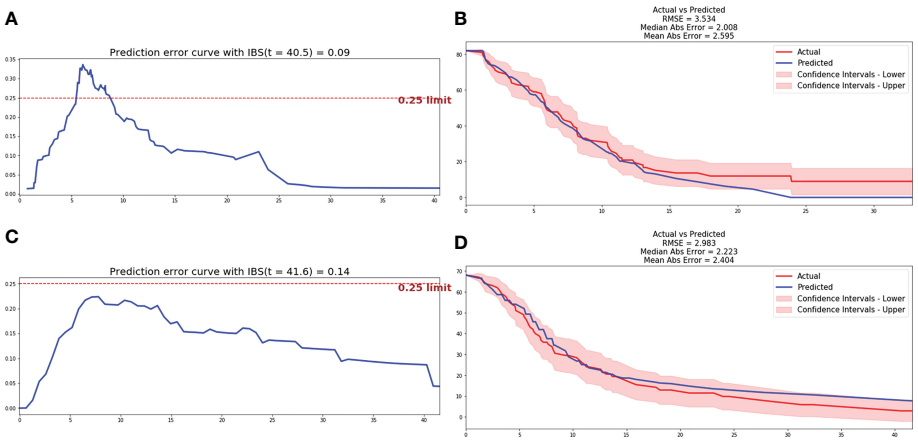


FIGURE 4
Prediction error curve and correction curve of DeepSurv and N-MTLR model. **(A)** Prediction error curve of DeepSurv model. **(B)** Prediction error curve of DeepSurv model. **(C)** Correction curve of N-MTLR model. **(D)** Correction curve of N-MTLR model.

10.3389/fonc.2023.1052147

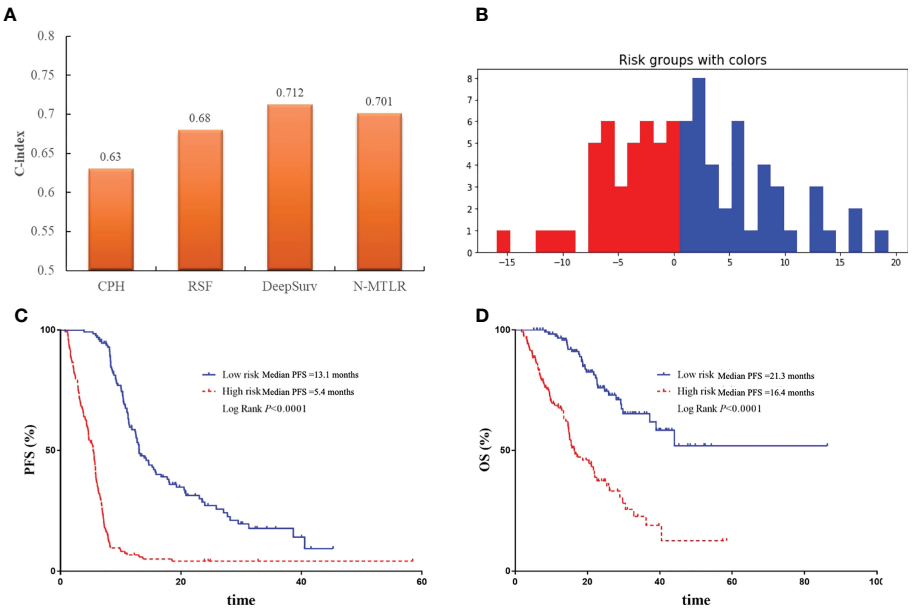


FIGURE 5
The comparison and validation of the performance of DeepSurv models. **(A)** The comparison of performance of machine learning and deep learning models. **(B)** All patients were divided into high-risk and low-risk group according to DeepSurv model. **(C)** Comparison of Kaplan-Meier survival curves of PFS between high-risk and low-risk patients. **(D)** Comparison of Kaplan-Meier survival curves of OS between high-risk and low-risk patients.

evaluation of efficacy and adverse effect (20, 21). Hosny A et al. have trained a CNN for automatic quantification and feature selection, thereby to prediction the 2-year OS of NSCLC patients. And the results indicated the reliable prediction performance of CNN with an AUC of 0.70, better than that of RSF, and were also robust in independent external datasets (22). However, the survival outcomes of this study were dichotomous variable rather than the survival or censor time, which cause the loss of outcome information and might contribute to the reduction of the prediction accuracy.

This study firstly trained the DNN prognostic model applying DeepSurv and N-MTLR for prognosis prediction of bevacizumab, whose output variables were set as PFS time and status. And the input layer of DNN consisted of clinicopathological, systemic inflammation, which has been demonstrated to be related to the

TABLE 3 Comparison of baseline characteristics of patients in retrospective and external validation cohort.

Parameters	Retrospective cohort (%)	External Validation cohort (%)	P
Age			0.54
≤57	139 (51.1%)	22 (56.4%)	
>57	133 (48.9%)	17 (43.6%)	
Gender			0.50
Male	155 (57%)	20 (51.3%)	
Female	117 (43%)	19 (48.7%)	
Smoking History			0.12
No	175 (64.3%)	20 (51.3%)	
Yes	97 (35.7%)	19 (48.7%)	
Anatomical type			0.16
Peripheral	194 (71.3%)	32 (82.1%)	
Central	78 (28.7%)	7 (17.9%)	
EGFR			0.26
Wild type	155 (57%)	16 (41%)	

(Continued)

TABLE 3 Continued

Parameters	Retrospective cohort (%)	External Validation cohort (%)	P
Sensitive mutation	81 (29.8%)	16 (41%)	
Resistance mutation	11 (4%)	3 (7.7%)	
NA	25 (9.2%)	4 (10.3%)	
Bone metastasis			0.91
No	170 (62.5%)	24 (61.5%)	
Yes	102 (37.5%)	15 (38.5%)	
Brain metastasis			0.15
No	193 (71%)	32 (82.1%)	
Yes	79 (29%)	7 (17.9%)	
Liver metastasis			0.15
No	236 (86.8%)	37 (94.8%)	
Yes	36 (13.2%)	2 (5.1%)	

efficacy of bevacizumab before (7), and radiomics variables. Besides, feature selection is not necessary as weight allocation has been performed for all parameters. DeepSurv takes the top layer of hidden layer as the input variable of proportional hazards model (13), and N-MTLR builds multi-layer neural network at different time intervals to evaluate the probability of event occurrence (17). The result indicated the best performance of DeepSurv model with the C-index of 0.712 compared to CPH and RSF models. The robust performance of DeepSurv model was also validated in external dataset, which indicated the extensive generalization and applicability of DNN model for prognostic prediction of bevacizumab. Our results indicated the superiority of DNN in prognosis prediction, especially in the analysis of high-dimensional features. It can be seen that DNN prognosis prediction is not only theoretically feasible but also can be extended to clinical practice to assist decision-making.

The prognostic biomarkers of bevacizumab were still inconclusive for advanced NSCLC patients. Our previous studies have found the predictive value of clinicopathological and systemic inflammatory factors for bevacizumab in NSCLC patients (7, 23). This study also indicated the prognostic value of radiomics features. Compared to single factor, the integrative model contains more information and can achieve better prediction performance. Thus, we developed DeepSurv prognostic model, which can be conveniently used by clinicians before and during treatment. For patients in high-risk group, another treatment strategy such as immunotherapy or combination treatment might be selected.

Although our study had many strengths, several limitations should be addressed here. Firstly, the sample size was still small, which might inevitably limit the performance and stability of DNN. Secondly, although thousands of variables were included in this study, they were still uncomprehensive and genomics, histological feature and others were not included. Thus, large cohort with more comprehensive variables are needed to optimize the DNN. Besides, DNNs are still more or less a kind of "black box" which could automatically modulate the weights of every variable upon the outcome, the potential biological mechanism needed to be further investigated in future studies.

5 Conclusions

Integrative DNN prognostic model was initially developed by combining radiomics signature with clinicopathological and

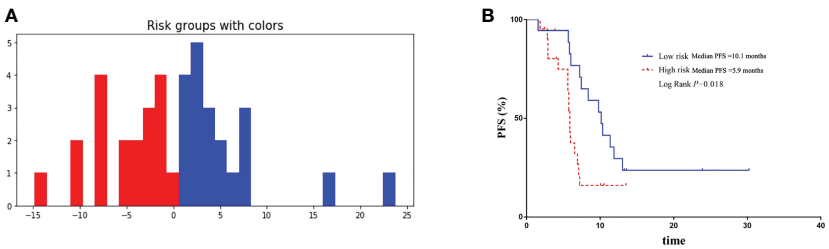


FIGURE 6
Validation of the performance of DeepSurv models in external validation cohort. **(A)** Patients in external validation cohort were divided into high-risk and low-risk group. **(B)** Comparison of Kaplan-Meier survival curves of PFS between high-risk and low-risk patients.

inflammatory feature using DeepSurv. The superior performance and robustness of DeepSurv model were observed, which open up prospects for the cross disciplines between AI and survival analysis of cancer patients. This easy-to-operated DNN model could not only assist in personalized treatment and surveillance strategies, but also provide patient consultation services, and was strongly suggested to be widely applied in clinical practice.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee of Shandong Cancer Hospital. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

(I) Conception and design: LW, BL and CJ. (II) Administrative support: YY, BZ and BF. (III) Provision of study materials or patients: LY, HL and SC. (IV) Collection and assembly of data: BL and CJ. (V) Data analysis and interpretation: LW and BL. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by National Natural Science Foundation of China, grant number 82172865 and 82203198, Natural Science Foundation of Shandong Province, grant number ZR2022QH023, Jinan Clinical Medical Science and Technology Innovation Plan, grant number 202019043.

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.2023.1052147/full#supplementary-material

References

1. Ferrara N, Hillan KJ, Gerber HP, Novotny W. Discovery and development of bevacizumab, an anti-VEGF antibody for treating cancer. *Nat Rev Drug discovery.* (2004) 3(5):391–400.

2. Garcia J, Hurwitz HI, Sandler AB, Miles D, Coleman RL, Deurloo R, et al. Bevacizumab (Avastin(R)) in cancer treatment: A review of 15 years of clinical experience and future outlook. *Cancer Treat Rev* (2020) 86:102017.

3. Dahlberg SE, Sandler AB, Brahmer JR, Schiller JH, Johnson DH. Clinical course of advanced non-small-cell lung cancer patients experiencing hypertension during treatment with bevacizumab in combination with carboplatin and paclitaxel on ECOG 4599. *J Clin Oncol Off J Am Soc Clin Oncol* (2010) 28(6):949–54.

4. Zhou C, Wu YL, Chen G, Liu X, Zhu Y, Lu S, et al. BEYOND: A randomized, double-blind, placebo-controlled, multicenter, phase III study of first-line Carboplatin/Paclitaxel plus bevacizumab or placebo in Chinese patients with advanced or recurrent nonsquamous non-Small-Cell lung cancer. *J Clin Oncol* (2015) 33(19):2197–204.

5. Reck M, Karagiannis T, Wehler T, Shtivelband M, Gonzalez-Larriba JL, Rothenstein J, et al. Patient-reported outcomes (PROs) in the randomized, phase III IMpower150 study of atezolizumab (atezo) + chemotherapy (chemo) ± bevacizumab (bev) vs chemo + bev in 1L nonsquamous metastatic NSCLC (mNSCLC). conference abstract. *J Clin Oncol* (2018) 36(15):9047.

6. Greillier L, Tomasini P, Barlesi F. Bevacizumab in the treatment of nonsquamous non-small cell lung cancer: Clinical trial evidence and experience. *Ther Adv Respir Dis* (2016) 10(5):485–91.

7. Li B, Wang S, Li C, Guo M, Xu Y, Sun X, et al. The kinetic changes of systemic inflammatory factors during bevacizumab treatment and its prognostic role in advanced non-small cell lung cancer patients. *J Cancer.* (2019) 10(21):5082–9.

8. 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(12):749–62.

9. Wu G, Jochems A, Refaee T, Ibrahim A, Yan C, Sanduleanu S, et al. Structural and functional radiomics for lung cancer. *Eur J Nucl Med Mol imaging.* (2021) 48 (12):3961–74.

10. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* (2015) 521(7553):436–44.

11. Aggarwal R, Sounderajah V, Martin G, Ting DSW, Karthikesalingam A, King D, et al. Diagnostic accuracy of deep learning in medical imaging: A systematic review and meta-analysis. *NPJ Digit Med* (2021) 4(1):65.

12. Solvang HK, Lingjaerde OC, Frigessi A, Borresen-Dale AL, Kristensen VN. Linear and non-linear dependencies between copy number aberrations and mRNA expression reveal distinct molecular pathways in breast cancer. *BMC Bioinf* (2011) 12:197.

13. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: Personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC Med Res Methodol* (2018) 18(1):24.

14. Huang Z, Johnson TS, Han Z, Helm B, Cao S, Zhang C, et al. Deep learning-based cancer survival prognosis from RNA-seq data: Approaches and evaluations. *BMC Med Genomics* (2020) 13(Suppl 5):41.

15. Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell N. Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med* (2021) 13 (1):152.

16. Katzman J, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. Deep survival: A deep cox proportional hazards network. *ArXiv abs/1606.00931* (2016).

17. Fotso S. Deep neural networks for survival analysis based on a multi-task framework. *ArXiv abs/1606.00931* (2018).

18. Zhao B, Tan Y, Tsai WY, Qi J, Xie C, Lu L, et al. Reproducibility of radiomics for deciphering tumor phenotype with imaging. *Sci Rep* (2016) 6:23428.

19. Cox DR. Regression models and life-tables. *J R Stat Society: Ser B (Methodological)* (1972) 34(2):187–202.

20. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyo D, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med* (2018) 24(10):1559–67.

21. Xu Y, Hosny A, Zeleznik R, Parmar C, Coroller T, Franco I, et al. Deep learning predicts lung cancer treatment response from serial medical imaging. *Clin Cancer Res* (2019) 25(11):3266–75.

22. Hosny A, Parmar C, Coroller TP, Grossmann P, Zeleznik R, Kumar A, et al. Deep learning for lung cancer prognostication: A retrospective multi-cohort radiomics study. *PloS Med* (2018) 15(11):e1002711.

23. Li B, Li C, Guo M, Shang S, Li X, Xie P, et al. Predictive value of LDH kinetics in bevacizumab treatment and survival of patients with advanced NSCLC. *Onco Targets Ther* (2018) 11:6287–94.

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 03 March 2023
DOI 10.3389/fonc.2023.1146362

Check for updates

OPEN ACCESS

EDITED BY
Hongda Liu,
Nanjing Medical University, China

REVIEWED BY
Shawn D. Spencer,
Philadelphia College of Osteopathic Medicine (PCOM), United States
Carmelo Tibaldi,
S.Luca Hospital, Italy
Vianney Poinsignon,
Bicêtre Hospital, France

*CORRESPONDENCE
Jiang Lou
Jianglou5855@163.com

SPECIALTY SECTION
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

RECEIVED 17 January 2023
ACCEPTED 20 February 2023
PUBLISHED 03 March 2023

CITATION
Chen L, Jiang H, Rao J-j, Wang L-s, Yan W, Ye J and Lou J (2023) Association between anlotinib trough plasma concentration and treatment outcomes in advanced non-small-cell lung cancer.
Front. Oncol. 13:1146362.
doi: 10.3389/fonc.2023.1146362

COPYRIGHT
© 2023 Chen, Jiang, Rao, Wang, Yan, Ye and Lou. 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 anlotinib trough plasma concentration and treatment outcomes in advanced non-small-cell lung cancer

Ling Chen[1,2], Hong Jiang[3], Jun-jie Rao[4], Liu-sheng Wang[5], Wei Yan[1,2], Jian Ye[6] and Jiang Lou[1,2]*

[1]Department of Pharmacy, Affiliated Hangzhou First People's Hospital, Zhejiang University School of Medicine, Hangzhou, China, [2]Department of Clinical Pharmacy, Key Laboratory of Clinical Cancer Pharmacology and Toxicology Research of Zhejiang Province, Affiliated Hangzhou First People's Hospital, Zhejiang University School of Medicine, Hangzhou, China, [3]Department of Cardiothoracic Surgery, Affiliated Hangzhou First People's Hospital, Zhejiang University School of Medicine, Hangzhou, China, [4]Department of Clinical Laboratory, Affiliated Hangzhou First People's Hospital, Zhejiang University School of Medicine, Hangzhou, China, [5]Department of Respiratory, Affiliated Hangzhou First People's Hospital, Zhejiang University School of Medicine, Hangzhou, China, [6]Department of Respiratory, Zhejiang Hospital, Hangzhou, China

Background: Efficacy and toxicities of anlotinib (ANL) show large inter-patient variation, which may partly be explained by differences in ANL exposure. Exposure-response/toxicities relationship have not been investigated for ANL. Therefore, the aim of the present study was to explore the association between the trough plasma concentration (C_{trough}) of ANL and treatment outcomes in Chinese patients with advanced non-small cell lung cancer (NSCLC).

Methods: Patients with advanced NSCLC who started third-line or further ANL alone therapy between January 2021 and October 2022. This study examined the ANL C_{trough} and clinical response evaluation at day 43 after initiation of ANL treatment. We evaluated the association between the ANL C_{trough} and clinical efficacy and toxicities. Additionally, this study defined patients with complete response (CR), partial response (PR) and stable disease (SD) as responder. The receiver-operating characteristic (ROC) curve combined with Youden index was identify the potential threshold value of ANL C_{trough} for the responder.

Results: 52 patients were evaluated for analyses. The median ANL C_{trough} was 11.45ng/ml (range, 3.69-26.36 ng/ml). The ANL C_{trough} values in the PR group (n=6, 15.51 ng/ml (range, 8.19-17.37 ng/ml)) was significantly higher than in the PD group (n=8, 7.44 ng/ml (range, 5.41-14.69 ng/ml), p=0.001). The area under the ROC curve (AUC_{ROC}) was 0.76 (95% confidence interval (CI), 0.58-0.93; p=0.022) and threshold value of ANL C_{trough} predicting responder was 10.29 ng/ml (sensitivity 65.9% and specificity 87.5%, the best Youden index was 0.53). The disease control rate (DCR) was 84.6%, and DCR was significantly higher in the high-exposure group (≥10.29ng/ml) than low-exposure group (<10.29ng/ml) (96.67% vs 68.18%, p=0.005). Although there was no significant difference in ANL C_{trough} between grade ≥ 3 and grade ≤2 toxicities, the incidence of any grade hand-foot syndrome (70.0% vs 36.36%, p=0.016) and thyroid-stimulating

hormone elevation (53.33% vs 22.73%, p =0.026) was significantly higher in the high-exposure group compared with the low-exposure group.

Conclusions: Considering these results, we propose that maintaining ANL C_{trough} ≥ 10.29ng/ml was important for achieving the response in advanced NSCLC patients treated with ANL.

KEYWORDS

anlotinib, trough plasma concentration, clinical efficacy, toxicities, non-small cell lung cancer

Introduction

Anlotinib (ANL) is a multi-targeted receptor tyrosine kinase inhibitor (TKI), which was developed by Jiangsu Chia-Tai Tianqing Pharmaceutical Group Co., Ltd. The major targets of ANL including vascular endothelial growth factor receptors (VEGF) 2 and 3, fibroblast growth factor receptors 1 to 4, platelet-derived growth factor receptors, c-Kit and Ret, resulting in treatment of various tumors (1, 2). ANL was approved as a third-line or further treatment for refractory advanced non-small cell lung cancer (NSCLC) by the National Medical Products Administration (NMPA) of China in 2018 (3). In phase 3 clinical study (ALTER 0303), ANL can significantly prolong overall survival (OS) and progression-free survival (PFS) and improve the quality of life in advanced NSCLC (4). The common toxicities during the treatment of NSCLC were hypertension, gastrointestinal reaction, hand-foot syndrome, and so on (4). Toxicities of grade 3 or higher were reported 21.67-61.9% and some patients led to dose reduction or treatment interruption (5; Han and Li and 6). Therefore, the optimal dose adjustment of ANL is crucial for avoiding serious toxicities and maximize its clinical efficacy.

As with other TKIs, the dose regimen of ANL are fixed and was adjusted mainly according to the degree of toxicities, the treatment regimen is 8mg, 10mg, or 12mg daily dose for 2 weeks on treatment followed by 1 week off treatment. In order to improve clinical efficacy and reduce toxicities, therapeutic drug monitoring (TDM) of other TKIs has been attempted for optimal dosage. For example, the sum of the total trough plasma concentration of both sunitinib and its active metabolite SU12662 should be set above 50ng/ml to obtain clinical efficacy and below 100ng/ml to avoid serious toxicities in sunitinib treated with metastatic renal cell carcinoma (7). In multiple studies, TDM may be useful for reducing toxicities and increasing clinical efficacy of imatinib, sunitinib and pazopanib (7). According to the phase I clinical trial, pharmacokinetic (PK) show that ANL is rapidly absorbed, high protein binding rates (93%) and eliminated slowly with a half-life ($t_{1/2}$) of 96h, the trough plasma concentration (C_{trough}) of ANL was reached at day 22 before the next treatment cycle was 5.05-28.5ng/ml (8). *Zhang et al.* have been reported that the median ANL C_{trough} was 25.38ng/ml (range 11.61-62.9ng/ml), which was characterized by high inter-patient variability in PK with 47.5% coefficients of variation (CV) (9). Several reports have also shown that ANL PK are partially dose-dependent and large inter-patient variability between 39.83% and 79.36%, which was similar to the other TKIs, such as imatinib, pazopanib and sunitinib (10, 11). ANL is mainly metabolized by CYP1A2 and CYP3A4/5 and secondary metabolic enzymes including CYP2C9, CYP2C19 and CYP2D6 (12, 13). Dexamethasone is a moderate inducer of CYP3A, which could accelerate the metabolism of ANL and reduce exposure (14). CYP2C19 rs3814637 and rs11568732 gene polymorphisms may also influence the ANL exposure (15). In addition, *Li et al.* and *Yu et al.* has been reported that ANL plasma concentration is higher in female patients than male patients under the same dose (14, 16). Beside these known contributions to variability in ANL exposure, patient characteristics such as age, body mass index, renal function and liver function may also contribute to variability in ANL exposure. Although those factors did not affect exposure separately, a relevant cumulative effect on plasma exposure cannot be excluded. Hence, TDM could be an effective tool for dose optimization of ANL according to plasma concentration. However, the information on the association between ANL exposure and treatment outcomes is limited.

In the present study, we aimed to evaluate the association of ANL C_{trough} with clinical efficacy and toxicities in Chinese patients with advanced NSCLC.

Material and methods

Patients

This prospectively study was performed in advanced NSCLC patient treatment with ANL alone at Hangzhou First People's Hospital between January 2021 and October 2022. All patients were disease progression after at least 2 lines of chemotherapy combined with or without immune check point inhibitors for patients without driver alterations as well as disease progression after at least 1 line of chemotherapy combined with or without immune check point inhibitors and TKI therapy for patients with driver alterations. The dosage regimen of ANL is 8mg, 10mg, or

12mg daily dose for 2 weeks on treatment followed by 1 week off treatment of a 21-day cycle. Daily dosing was determined individually by each attending doctor based on the patient's age, bodyweight, Eastern Cooperative Oncology Group (ECOG) Performance Status (PS) and so on. Patients were evaluated for clinical efficacy and plasma concentration on day 43 after initiation of ANL treatment. This study was approved by the Ethics Committee of the Hangzhou First People's Hospital (Ethical approval documents: No. 2020-098-01) and the registration number ChiCTR2100043709 was granted by the Chinese Clinical Trial Registry, and written informed consent was obtain from all patients. We extracted clinical data including age, sex, bodyweight, dosage, metastasis, etiology, TNM stage, PS score from the electronic medical records.

Measurement of ANL C_{trough}

Previous studies have reported that ANL plasma concentration were measured using ultra-high-performance liquid chromatography coupled to tandem mass spectrometry (UPLC-MS/MS) (10, 17, 18). However, we were unable to obtain the internal standard in the literature. Trough methodological research, this study found that zanubrutinib is a suitable new internal standard. A new specific and sensitive UPLC-MS/MS has been developed and application for determination of ANL C_{trough} in this study. Briefly, plasma sample were treated by protein precipitation and addition of zanubrutinib as internal standard, then 5.0μL of the supernatant was analyzed and separated on an Ultimate XB-C18 column (100mm × 2.1mm, 3.0μm particle size) with optimized gradient of mobile phase A (consisted of 10mM ammonium acetate-0.1% formic acid) and mobile phase B (acetonitrile) was used to elute ANL, the gradient elution procedure was set as follows: 0-2.5min, 10%B→50%B; 2.5-2.6min, 50%B→100%B; 2.6-3.5min, 100%B; 3.5-5.0min, 100%B→10%B. The Mass spectrometric detection was performed on a Xevo TQD Triple Quadrupole Mass Spectrometry (Waters, USA) with an electrospray ionization source coupled with the above described UPLC system. Quantitative analysis was conducted using MRM transitions of m/z 408.1→339.1 for ANL and 472.2→290.1 for zanubrutinib, respectively. The calibration curve of ANL in plasma was linear over the concentration range of 1.0-100.0ng/mL with a correlation coefficient r^2>0.990. The precision and accuracy was less than 8.64% and within ±104.61%, respectively. The extraction recovery and matrix effect was 104.81-107.32% and 102.54%-104.26%, respectively. The method was validated and accorded with the criteria of industrial guidance for bioanalytical method validation of the Food and Drug Administration. (FDA. 2018. Guidance for industry, bioanalytical method validation. FDA, Washington, DC.) Blood samples (4ml aliquots) were collected from 52 patients on day 43 after the initiation of ANL treatment to measure ANL C_{trough}. Patients could not be evaluated for ANL C_{trough} were excluded.

Assessment of efficacy and safety

Efficacy evaluation was assessed in accordance with the Response Evaluation Criteria in Solid Tumors (RECIST) guidelines, version 1.1, which were included complete response (CR), partial response (PR), stable disease (SD), or progressive disease (PD) based on the computed tomography or magnetic resonance imaging. The assessment of clinical efficacy was performed at day 43 after the initiation of ANL treatment. According to the *Noda S et al.* study, we defined patients with CR, PR and SD as responder, and defined patients with PD as non-responder in this study (19). Disease control rate (DCR) was defined as the percentage of patients with CR, PR and SD. Patients could not be evaluated for efficacy were excluded. Toxicities were registered from the start of treatment until at day 43 according to the Common Terminology Criteria for Adverse Events (CTCAE) version 5.0. Toxicities of worst grade were recorded within first two cycles. Additionally, we examined association between the median ANL C_{trough} and clinical efficacy and toxicities.

Statistics analysis

Categorical variables statistically analyzed with the Chi-square test or Fisher's exact test, and continuous variables were statistically analyzed using the Mann-Whitney U-test, significance was set at p<0.05. Kruskal-Wallis H with Bonferroni-corrected test was performed for the three-group comparison. The adjusted p values by Bonferroni-corrected test multiplied the p values of each paired comparison group by 3, and adjusted p values of < 0.017 were considered statistically significant for the comparison between the two groups. Patients were divided according to the PR, SD and PD, different dosing doses of ANL and toxicities (grade ≥3 or graded ≤2 toxicities). Comparing the ANL C_{trough} between PR, SD and PD, and different dosing doses of ANL using Kruskal-Wallis H with Bonferroni-corrected test. Comparing the toxicities grade ≥3 and graded ≤2 toxicities with Mann-Whitney U-test. The sensitivity and specificity of the analyzed data were calculated by receiver operating characteristic (ROC) curves, and predictive capacity is expressed by the area under the curve (AUC) value. The optimal ANL C_{trough} for prediction of responder was determined by Youden index (sensitivity +specificity-1). Patients also were divided into high-exposure group (above prediction concentration) and low-exposure group (below prediction concentration) according to the prediction concentration. Patient characteristics and the incidence of any grade toxicities and grade ≥ 3 toxicities were compared between low-exposure group and high-exposure group. Univariate and multivariate logistic regression analyses were evaluated to potential factors associated with DCR. Variables with borderline significance (p<0.2) on the univariate analysis were subjected to multivariate logistic regression analyses, p<0.05 were considered statistically significance in the multivariate analysis (20). All statistical analyses were performed using SPSS 22.0.

10.3389/fonc.2023.1146362

TABLE 1 Baseline characteristics of patients.

Characteristic	Value (median, range) or (%)
Age(years)	66 (44-84)
Body weight (kg, range)	60 (41-89)
Body surface area (m^2, range)	1.61 (1.38-2.01)
Dosage	
12mg	15 (28.85%)
10mg	24 (46.15%)
8mg	13 (25.00%)
Gender, n (%)	
Male	31 (59.62%)
Female	21 (40.38%)
Smoking history, n (%)	
Never	29 (55.76%)
Current/former	23 (44.23%)
ECOG PS, n (%)	
0	12 (23.08%)
1	31 (59.62%)
2	9 (17.30%)
Histology, n (%)	
Adenocarcinoma	31 (59.62%)
Squamous	21 (40.38%)
TNM stage, n (%)	
III stage	18 (34.62%)
IV stage	34 (65.38%)
Driver alterations	
No	37 (71.15%)
Yes	15 (28.85%)
EGFR mutation	10 (19.23%)
ALK rearrangement	2 (3.85%)
Other mutation	3 (5.77%)
Number of metastases	
≤3	21 (40.38%)
>3	31 (59.62%)
Efficacy of previous therapy	
Yes	48 (92.31%)
No	4 (7.69%)
Drugs of previous therapy	
Chemotherapy	8 (15.38%)
Chemotherapy+Immunotherapy	29 (55.77%)
Chemotherapy+Immunotherapy +Target	15 (28.85%)

ECOG, Eastern Cooperative Oncology Group; PS, performance status; TNM, tumor node metastasis.

Results

Patient characteristics

58 patients treated with ANL were enrolled in this study. We excluded one patient for whom treatment was discontinued within 43 days owing to grade 3 hand-foot syndrome toxicity. Two patients who were rapid clinical deterioration and could not be evaluated for ANL C_{trough} at day 43 after the initiation of ANL treatment. Three patients could not be evaluated for efficacy who were changed hospitals. The above-mentioned six patients were excluded from analysis. 52 patients were included and baseline characteristics are shown in Table 1. The median age was 66 years (range, 44-84 years) and median body weight was 60 kg (range, 41-89 kg). The dosage regimen of ANL was 12mg in 15 patients, 10mg in 24 patients, and 8mg in 13 patients. ANL C_{trough} were higher than LLOQ (1.0 ng/ml) from 52 patients. The median (minimum-maximum) and mean ± SD plasma concentrations of ANL C_{trough} was 11.45ng/ml (range, 3.69-26.36 ng/ml) and 11.38 ± 4.29ng/ml (n=52, CV=37.66%), which shown large inter-patient variability in ANL $C_{trough.}$ The distribution of ANL C_{trough} were presented in Figure 1.

Association of efficacy with ANL C_{trough}

CR, PR, SD and PD were observed in 0, 6, 38 and 8 patients, respectively. Based on the results of plasma concentration, the median (minimum-maximum) and mean C_{trough} of ANL treatment with 8mg, 10mg and 12mg were 10.10ng/ml (range, 3.69-16.49) and 10.06ng/ml (CV=38.46%), 11.51ng/ml (range, 3.75-26.36ng/ml) and 11.42ng/ml (CV=41.52%), and 13.19ng/ml (range, 5.41-17.37 ng/ml) and 12.47ng/ml (CV=30.31%), respectively. Although there was high inter-individual variability among patients treated with the same dose of ANL, there was no significant correlation between the median ANL C_{trough} versus daily dose, which was shown in Figure 1S. The median C_{trough} of the PR (n=6), SD (n=38) and PD (n=8) groups were 15.51ng/ml (range, 8.19-17.37ng/ml), 11.56ng/ml (range, 3.69-26.36ng/ml) and 7.44 ng/ml (range, 5.41-14.69ng/ml), respectively, the result was presented in Figure 2. The median C_{trough} in the PR

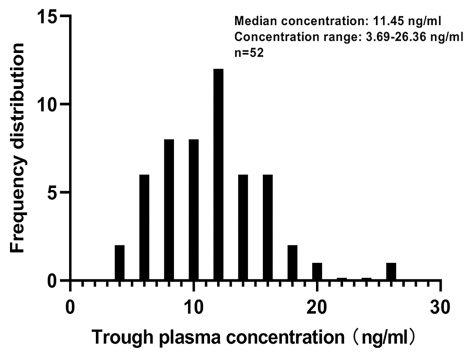


FIGURE 1
Distribution of anlotinib trough plasma concentration.

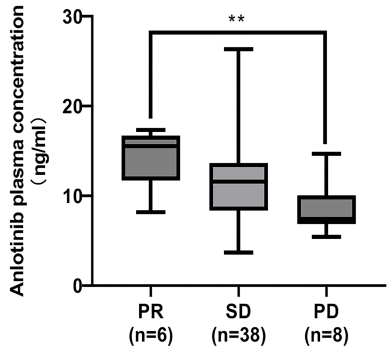


FIGURE 2
Comparison of the median ANL C_{trough} in patients with partial response (PR), stable disease (SD) and progressive disease (PD). The adjusted p values by Bonferroni-corrected test multiplied the p values of each paired comparison group by 3. **represents p<0.01.

group was significantly higher than PD group (p=0.001), but median C_{trough} in the SD group was not significantly difference in the PD group (p=0.148), and median C_{trough} in the PR group was also not significantly difference in the SD group (p=0.181). The area under the ROC curve (AUC_{ROC}) was 0.76 (95% confidence interval (CI), 0.58-0.93; p=0.022) and threshold value of ANL C_{trough} predicting responder was 10.29 ng/ml (sensitivity 65.9% and specificity 87.5%, the best Youden index was 0.53), which were depicted in Figure 3. Additionally, the median ANL C_{trough} was 12.39 ng/ml (range, 3.69-26.36ng/ml) in the responder group (n=44) was significantly higher than the non-responder group (n=8, 7.44ng/ml (range, 3.69-26.36ng/ml), p=0.022). We divided the patients into the high-exposure group (≥10.29ng/ml) and low-exposure group (<10.29ng/ml) according to the threshold value. DCR was significantly higher in the high-exposure group than in the low-exposure group (96.67% vs 68.18%, p=0.005).

The results of the comparison of patient characteristics in the high-exposure group and low-exposure group are shown in Table 2. ANL C_{trough} was significantly higher in the high-exposure group compared with the low-exposure group (11.90ng/ml vs 8.55ng/ml, p=0.021). There was significant difference in the dosage of administration between high-exposure group and low-exposure group (p=0.023). No significant difference was observed in the sex (p=0.613), age (p=0.290), body weight (p=0.242), body surface area (p=0.824), smoking history (p=0.473), ECOG PS (p=0.104), histology (p=0.613), driver alterations (p=0.404), number of metastases (p=0.281), TNM stage (p=0.414) and efficacy of previous therapy (p=0.466) among high-exposure group and low-exposure group.

Factors associated with the DCR

Table 3 illustrated the results of univariate and multivariate logistic regression analyses of the DCR in 52 patients. In the univariate analysis, ANL C_{trough} ≥10.29 ng/ml (p=0.020), age ≥65 years (p=0.133) and dosage ≥10mg (p=0.089) were below borderline significance (p<0.2). In the multivariate analysis, ANL C_{trough}

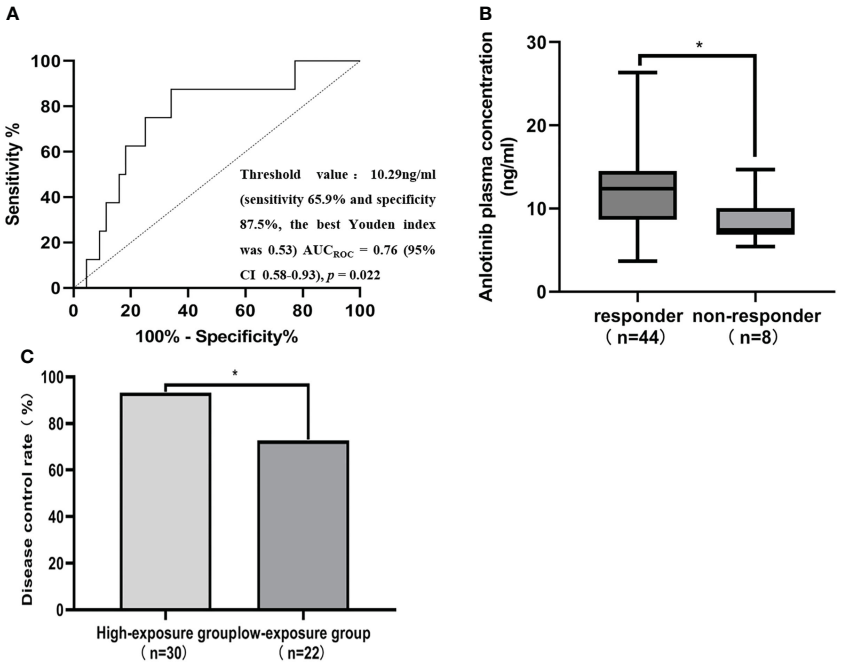


FIGURE 3
ANL C_{trough} threshold for responder and disease control rate (DCR). **(A)** Receiver operating characteristics (ROC) curves for predicting responder status. AUC_{ROC} area under the receiver operating characteristic curve, CI Confidence interval. **(B)** Comparison of ANL C_{trough} between patients with responder and non-responder. Responders were defined as the patients with CR and SD at best response, while non-responder were defined the patients with PD at best response. **(C)** Comparison of the DCR between patients with a median ANL C_{trough} ≥ 10.29 ng/ml (high-exposure group) and <10.29 ng/ml (low-exposure group), * represents p<0.05.

TABLE 2 Comparison between patients with ANL C_{trough} ≥10.29ng/ml (high-exposure group) and <10.29ng/ml (low-exposure group).

Characteristic	high-exposure group (n=30)	low-exposure group (n=22)	*p* values
Age (years, range)	69 (48-81)	65 (44-84)	0.29
Body weight (kg, range)	58 (43-89)	60 (41-80)	0.242
Body surface area (m^2, range)	1.61 (1.38-2.01)	1.66 (1.40-1.94)	0.824
Sex			
Male	17	13	0.613
Female	13	8	
Dosage			
≥10mg	26	13	**0.023**
8mg	4	9	
Smoking history			
Never	12	11	0.473
Current/former	18	11	
ECOG PS			
≤1	27	16	0.104
≥2	3	6	
Histology			
Adenocarcinoma	17	14	0.613
Squamous	13	8	
TNM stage			
III stage	9	9	0.414
IV stage	21	13	
Driver alterations			
No	20	17	0.404
Yes	10	5	
Number of metastases			
>3	14	7	0.281
≤3	16	15	
Efficacy of previous therapy			
Yes	27	21	0.466
No	3	1	
Drug Interactions			
Dexamethasone	2	0	0.217
Amlodipine	4	2	0.636
Nifedipine	3	2	0.913
losartan	1	3	0.168
Metformin	4	3	0.975
Insulin	3	1	0.466
Levothyroxine	2	4	0.199
Atorvastatin	6	3	0.549
Rosuvastatin	3	5	0.209

ECOG, Eastern Cooperative Oncology Group; PS, performance status; TNM, tumor node metastasis.
Bold values indicate that the difference were statistically significant.

10.3389/fonc.2023.1146362

TABLE 3 Univariate and multivariate analyses of potential factors associated with disease control rate.

	No. of case	Univariate analysis		Multivariate analysis	
		ORs (95% CI)	*p* values	ORs (95% CI)	*p* values
Sex					
Male	31	0.44	0.344		
Female	21	(0.08-2.42)			
Age				0.17	
≥65 years	32	0.19	0.133	(0.02-1.79)	0.139
<65 years	20	(0.02-1.66)			
Body weight					
≥65 kg	18	1.71	0.538		
<65 kg	34	(0.31-9.52)			
Smoking history					
Never	23	1.39	0.678		
Current/former	29	(0.30-6.54)			
ECOG PS					
≤1	43	1.76	0.536		
≥2	9	(0.29-10.58)			
Dosage					
≥10mg	39	0.26	0.089	0.3	0.194
8mg	13	(0.05-1.23)		(0.05-1.85)	
ANL C_{trough}					
≥10.29ng/ml	32	0.07	0.02	0.09	0.034
<10.29ng/ml	20	(0.01-0.69)		(0.01-0.83)	
Histology					
Adenocarcinoma	31	0.87	0.857		
Squamous	21	(0.18-4.09)			
TNM stage					
III stage	18	0.86	0.862		
IV stage	34	(0.18-4.11)			
Driver alterations					
No	37	1.6	0.559		
Yes	15	(0.33-7.55)			
AEs: Hypertension					
Yes	27	1.32	0.721		
No	25	(0.29-5.95)			
AEs: Hand-foot syndrome	29	2.41	0.267		
Yes	23	(0.51-11.37)			
No					
AEs: Thyroid stimulating hormone elevation					
Yes	21	1.39	0.678		

(Continued)

TABLE 3 Continued

	No. of case	Univariate analysis		Multivariate analysis	
		ORs (95% CI)	*p* values	ORs (95% CI)	*p* values
No	31	(0.30-6.54)			
AEs: Hypertriglyceridemia					
Yes	22	1.27	0.765		
No	30	(0.27-5.97)			

ANL, anlotinib; AEs, adverse events; ORs, odds ratio. ECOG, Eastern Cooperative Oncology Group; PS, performance status; TNM, tumor node metastasis.

≥10.29 ng/ml was independently associated with DCR (odds ratio, 0.09; 95% CI 0.01-0.83; p=0.034).

Association of toxicity with ANL C_{trough}

In the present study, the most common toxicities were hypertension, hand-foot syndrome, hypertriglyceridemia and thyroid-stimulating hormone elevation. Of these, the most common grade ≥3 toxicities were hypertension (21.15%), hand-foot syndrome (3.85%) and thyroid-stimulating hormone elevation (3.85%). No significant difference in ANL C_{trough} was found when comparing grade ≥ 3 toxicities with grade ≤2 toxicities (p=0.271), which were depicted in Figure 4. All patients in the high-exposure group and low-exposure group experienced toxicities induced by ANL in Table 4. Although there was no significant difference in the incidence of most toxicities among high-exposure group and low-exposure group, the incidence of any grade hand-foot syndrome (70.00% vs 36.36%, p=0.016) and any grade thyroid-stimulating hormone elevation (53.33% vs 22.73%, p=0.026) was significantly higher in the high-exposure group than the low-exposure group.

Discussion

In the present study, the median ANL C_{trough} was 11.45ng/ml (range, 3.69-26.36 ng/ml) and shows large inter-patient variability with 37.66% CV, which was similar to the previous studies, and ANL occasionally cause no response in patients with NSCLC (9–11). The influence of plasma concentration on clinical response should be considered during ANL therapy. A preclinical study showed that ANL have high selectivity for vascular endothelial growth factor (VEGF) family members, especially VEGFR2, with half maximal inhibitory concentration (IC50) value was 0.2 ng/ml in endothelial cell (2b). However, information on the association between ANL exposure and treatment outcomes is limited. To our knowledge, this study was the first reported that ANL C_{trough} was higher in the PR group than in the PD groups, and the threshold value of C_{trough} associated with the responder for advanced NSCLC was 10.29 ng/ml using ROC analysis combined with Youden index. In the non-responder group (n=8), ANL C_{trough} of 87.5% (7/8) non-responder patients were below the 10.29ng/ml, which means that threshold value (10.29ng/ml) could predict non-responder in advanced NSCLC treated with ANL. Furthermore, ANL C_{trough} show partially dose-dependent and there was significant difference between the high-exposure group and low-exposure group at a dosage of 8mg versus ≥10mg of ANL per day, which is consistent with the *Tan T et al.* study (15). ANL C_{trough} ≥ 10.29ng/ml was achieved by changing the dose administered, which might obtain sufficient efficacy. *Yu M et al.* and *Li Z et al.* have been reported that the plasma concentration of ANL in female was higher than male patients, but this study found that gender could not affect the ANL C_{trough} (14, 16). Dexamethasone and CYP2C19 gene polymorphisms (rs3814637 and rs11568732) may also influence the ANL exposure (14, 15). However, we did not found significance interaction of ANL with other drugs, which were shown in Table 2. Subsequent studies could focus on the influence of gene polymorphisms, drug interactions and other factors on ANL C_{trough}.

In this study, the DCR was 84.6% (44/52), which was similar to the ALTER 0302 (DCR 83.3%) and ALTER 0303 study (DCR 81.0%) (Han and Li and 6, 21). However, this study divided the patients into the high-exposure group (≥10.29ng/ml) and low-exposure group (<10.29ng/ml) according to the threshold value. The DCR can reach 96.67% (29/30) in the high-exposure group, which was higher than 83.3% and 81.0%. *Wang J et al.* reported that ANL-induced hypertension, hand-foot syndrome, hypertriglyceridemia and elevation thyroid stimulating hormone were independent predictor factors of PFS in refractory NSCLC treated with ANL(6). However, our present study has not investigated the association between the ANL-induced hypertension, hand-foot syndrome,

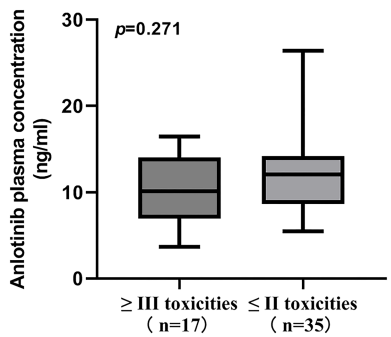


FIGURE 4
Comparison of ANL C_{trough} between patients with grade ≥ III toxicities and grade ≤ II toxicities.

TABLE 4 Comparison of adverse events in patients with ANL C_{trough} ≥10.29ng/ml (high-exposure group) and <10.29ng/ml (low-exposure group).

Type	Level	high-exposure group (n=30) (%)	low-exposure group (n=22) (%)	*p* values
Hypertension	Any grade	16 (53.33)	11 (50.0)	0.812
	≥ Grade III	5 (16.67%)	6 (27.28)	0.355
Hypertriglyceridemia	Any grade	12 (40.0)	10 (45.45)	0.694
	≥ Grade III	1 (3.33)	1 (4.55)	0.822
Hand-foot syndrome	Any grade	21 (70.0)	8 (36.36)	0.016
	≥ Grade III	2 (6.67)	0 (0)	0.217
Diarrhea	Any grade	6 (20.0)	6 (27.27)	0.539
	≥ Grade III	0 (0)	1 (4.55)	0.238
Thrombocytopenia	Any grade	3 (10.0)	2 (9.09)	0.972
	≥ Grade III	1 (3.33)	0 (0)	0.387
Fatigue	Any grade	8 (26.67)	6 (27.27)	0.961
	≥ Grade III	0 (0)	1 (4.55)	0.238
Proteinuria	Any grade	5 (16.67)	6 (27.27)	0.355
	≥ Grade III	1 (3.33)	0 (0)	0.387
Hemoptysis	Any grade	6 (20.0)	3 (13.64)	0.549
	≥ Grade III	1 (3.33)	0 (0)	0.387
Thyroid-stimulating hormone elevation	Any grade	16 (53.33)	5 (22.73)	**0.026**
	≥ Grade III	2 (6.67)	0 (0)	0.217

Bold values indicate that the difference were statistically significant.

hypertriglyceridemia and elevation thyroid stimulating hormone and DCR, which were shown in Table 3. Multivariate logistic regression indicated that ANL C_{trough} ≥10.29ng/ml was an independent prognostic factor with DCR in this study. It is possible that PFS and DCR are different efficacy evaluation indicators. According to the results of this study, we proposed that maintaining ANL C_{trough} ≥10.29ng/ml is crucial for achieving the DCR.

In clinical practice, toxicities often lead to treatment discontinuation or dose reductions. The most common toxicities associated with ANL were hypertension, hand-foot syndrome, hypertriglyceridemia and thyroid-stimulating hormone elevation in this study. Notably, hypertension was the most common adverse events during ANL treatment, which is consistent with the ALTER 0302 and ALTER 0303 (21, 21). According to the ALTER 0303, most of toxicities occurred within the first two cycles of ANL treatment, the follow-up period of toxicities was also set the first two cycles in this study. Although there was no significantly difference with grade ≥ 3 toxicities and grade ≤ 2 toxicities between the high-exposure group and low-exposure group, the incidence of any grade hand-foot syndrome and thyroid-stimulating hormone elevation was significantly higher in the high-exposure group in this study. Based on the exploratory data from a phase I trial, ANL exhibited a quite long $t_{1/2}$ (96 ± 17h) and accumulation of plasma over time with accumulation ratio of 12, suggesting that drug accumulation over time (8). *Zhang J et al.* has been reported that the initial dosage was 12mg ANL in a patient, C_{trough} level in cycle 1 and 2 of treatment was 12.75 and 15.83ng/ml, respectively. However, C_{trough} level increased to 44.01 ng/ml in cycle 4 with intolerance to toxicity, then the dose was reduced to 10mg from cycle 5. Thereafter, C_{trough} level decreased to 20.84 ng/ml (22). The association between ANL C_{trough} and toxicities was not examined at long-term in the present study. Considering the accumulation effect of ANL, it is necessary to explore the association between the ANL C_{trough} and toxicities by long-term follow-up period.

There were some limitations to the present study. First, we were unable to estimate the area under the plasma concentration-time curve of ANL, which is an indicator of the drug exposure *in vivo*. However, for molecularly targeted drugs, such as sunitinib, imatinib and pazopanib, TDM based on C_{trough} level is recommended (11, 23). Second, this study was a single-center study. The number of patients involved was small and the follow-up period was short. For confirming these findings, multiple-center study and large PK data should be performed by long-term follow-up. Third, this study could not evaluate the association between the ANL C_{trough} and PFS or OS. The contribution of ANL C_{trough} to PFS and OS to be further defined.

In conclusion, this is the first study to elucidate the relationship between ANL C_{trough} and clinical efficacy and toxicities in Chinese patients with advanced NSCLC. The present study demonstrated

that there was an extensive inter-individual variability in ANL C_{trough}, and maintaining C_{trough} ≥10.29ng/ml is crucial for achieving the responder in advanced Chinese NSCLC patients treated with ANL. The ANL C_{trough} could be used as a guide to improving efficiency.

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 human participants were reviewed and approved by the Ethics Committee of the Hangzhou First People's Hospital (Ethical approval documents: No.2020-098-01). The patients/participants provided their written informed consent to participate in this study.

Author contributions

LJ and LC conceptualized and supervised the project. L-SW and JY carried out the study and interpreted the results. HJ drafted the initial version of the manuscript. WY and J-JR searched the literature and collected data. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by Zhejiang Provincial Natural Science Foundation of China under Grant No. LGF21H31003 and Zhejiang Provincial Medical and Health Technology Project (grant. 2021433647), Zhejiang Pharmaceutical Association Hospital Pharmacy special research support project (NO.2022ZYY09), the Key Medical Discipline of Hangzhou City (2021, No.21) and Key Laboratory of Clinical Cancer Pharmacology and Toxicology Research of Zhejiang Province (No. 2020E10021).

Acknowledgments

We would like to take the opportunity to thank the patients and their families, and all of the research members in this study.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.1146362/full#supplementary-material

References

1. Taurin S, Yang CH, Reyes M, Cho S, Coombs DM, Jarboe EA, et al. Endometrial cancers harboring mutated fibroblast growth factor receptor 2 protein are successfully treated with a new small tyrosine kinase inhibitor in an orthotopic mouse model. *Int J Gynecol Cancer* (2018) 28:152–60. doi: 10.1097/IGC.0000000000001129

2. Xie C, Wan X, Quan H, Zheng M, Fu L, Li Y, et al. Preclinical characterization of anlotinib, a highly potent and selective vascular endothelial growth factor receptor-2 inhibitor. *Cancer Sci* (2018) 109:1207–19. doi: 10.1111/cas.13536

3. Syed YY. Anlotinib: first global approval. *Drugs* (2018) 78:1057–62. doi: 10.1007/s40265-018-0939-x

4. Han B, Li K, Wang Q, Zhang L, Shi J, Wang Z, et al. Effect of anlotinib as a third-line or further treatment on overall survival of patients with advanced non-small cell lung cancer: the alter 0303 phase 3 randomized clinical trial. *JAMA Oncol* (2018) 4:1569–75. doi: 10.1001/jamaoncol.2018.3039

5. Si X, Zhang L, Wang H, Zhang X, Wang M, Han B, et al. Management of anlotinib-related adverse events in patients with advanced non-small cell lung cancer: experiences in alter-0303. *Thorac Cancer* (2019) 10:551–6. doi: 10.1111/1759-7714.12977

6. Wang J, Zhao Y, Wang Q, Zhang L, Shi J, Wang Z, et al. Prognostic factors of refractory nsclc patients receiving anlotinib hydrochloride as the third- or further-line treatment. *Cancer Biol Med* (2018) 15:443–51. doi: 10.20892/j.issn.2095-3941.2018.0158

7. Verheijen RB, Yu H, Schellens J, Beijnen JH, Steeghs N, Huitema A. Practical recommendations for therapeutic drug monitoring of kinase inhibitors in oncology. *Clin Pharmacol Ther* (2017) 102:765–76. doi: 10.1002/cpt.787

8. Sun Y, Niu W, Du F, Du C, Li S, Wang J, et al. Safety, pharmacokinetics, and antitumor properties of anlotinib, an oral multi-target tyrosine kinase inhibitor, in patients with advanced refractory solid tumors. *J Hematol Oncol* (2016) 9:105. doi: 10.1186/s13045-016-0332-8

9. Zhou L, Wang S, Chen M, Huang S, Zhang M, Bao W, et al. Simultaneous and rapid determination of 12 tyrosine kinase inhibitors by lc-ms/ms in human plasma: application to therapeutic drug monitoring in patients with non-small cell lung cancer. *J Chromatogr B Analyt Technol BioMed Life Sci* (2021) 1175:122752. doi: 10.1016/j.jchromb.2021.122752

10. Du P, Guan Y, An Z, Li P, Liu L. A selective and robust uplc-ms/ms method for the simultaneous quantitative determination of anlotinib, ceritinib and ibrutinib in rat plasma and its application to a pharmacokinetic study. *Analyst* (2019) 144:5462–71. doi: 10.1039/c9an00861f

11. Westerdijk K, Desar I, Steeghs N, van der Graaf W, van Erp NP. Imatinib, sunitinib and pazopanib: from flat-fixed dosing towards a pharmacokinetically guided personalized dose. *Br J Clin Pharmacol* (2020) 86:258–73. doi: 10.1111/bcp.14185

12. Sun W, Wang Z, Chen R, Huang C, Sun R, Hu X, et al. Influences of anlotinib on cytochrome p450 enzymes in rats using a cocktail method. *BioMed Res Int* (2017) 2017:3619723. doi: 10.1155/2017/3619723

13. Jin Z, He Q, Zhu X, Zhu M, Wang Y, Wu XA, et al. Application of physiologically based pharmacokinetic modelling for the prediction of drug-drug interactions involving anlotinib as a perpetrator of cytochrome p450 enzymes. *Basic Clin Pharmacol Toxicol* (2022) 130:592–605. doi: 10.1111/bcpt.13721

14. Li Z, Chen H, Shi J, Wang F, Liu Y. A novel method for the quantification of anlotinib in human plasma using two-dimensional liquid chromatography. *BioMed Chromatogr* (2021) 35:e5218. doi: 10.1002/bmc.5218

15. Tan T, Han G, Cheng Z, Jiang J, Zhang L, Xia Z, et al. Genetic polymorphisms in cyp2c19 cause changes in plasma levels and adverse reactions to anlotinib in chinese patients with lung cancer. *Front Pharmacol* (2022) 13:918219. doi: 10.3389/fphar.2022.918219

16. Yu M, Liang Z, Xiong L, Liu F, Cheng L, Jia C, et al. Establishment and validation of a lc-ms/ms method for the determination of anlotinib in human plasma: application to therapeutic drug monitoring. *BioMed Chromatogr* (2022) 36:e5501. doi: 10.1002/bmc.5501

17. Zhong CC, Chen F, Yang JL, Jia WW, Li L, Cheng C, et al. Pharmacokinetics and disposition of anlotinib, an oral tyrosine kinase inhibitor, in experimental animal species. *Acta Pharmacol Sin* (2018) 39:1048–63. doi: 10.1038/aps.2017.199

18. Ma Z, Lu S, Zhou H, Zhang S, Wang Y, Lin N. Determination of intracellular anlotinib, osimertinib, afatinib and gefitinib accumulations in human brain microvascular endothelial cells by liquid chromatography/tandem mass spectrometry. *Rapid Commun Mass Spectrom* (2021) 35:e8955. doi: 10.1002/rcm.8955

19. Noda S, Iida H, Fujimoto T, Wakasugi Y, Yabuta N, Sudou M, et al. Exploratory analysis of target concentration of lenvatinib in the treatment of hepatocellular carcinoma. *Cancer Chemother Pharmacol* (2021) 88:281–8. doi: 10.1007/s00280-021-04286-2

20. Hata K, Suetsugu K, Egashira N, Makihara Y, Itoh S, Yoshizumi T, et al. Association of lenvatinib plasma concentration with clinical efficacy and adverse events in patients with hepatocellular carcinoma. *Cancer Chemother Pharmacol* (2020) 86 (6):803–13. doi: 10.1007/s00280-020-04178-x

21. Han B, Li K, Zhao Y, Li B, Cheng Y, Zhou J, et al. Anlotinib as a third-line therapy in patients with refractory advanced non-small-cell lung cancer: a multicentre, randomised phase ii trial (alter0302). *Br J Cancer* (2018) 118:654–61. doi: 10.1038/bjc.2017.478

22. Zhang J, Wang W, Du J, Li C, Wang S, Han Y, et al. A rapid and sensitive ultra-performance liquid chromatography tandem mass spectrometry method for determination of anlotinib in plasma and dried blood spots: method development, validation, and clinical application. *Rapid Commun Mass Spectrom* (2022) 36:e9372. doi: 10.1002/rcm.9372

23. Lankheet N, Desar I, Mulder SF, Burger DM, Kweekel DM, van Herpen C, et al. Optimizing the dose in cancer patients treated with imatinib, sunitinib and pazopanib. *Br J Clin Pharmacol* (2017) 83:2195–204. doi: 10.1111/bcp.13327

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 15 March 2023
DOI 10.3389/fonc.2023.1041140

Check for updates

OPEN ACCESS

EDITED BY
Ping Zhan,
Nanjing University School of Medicine, China

REVIEWED BY
Xiaoding Hu,
University of Texas MD Anderson Cancer Center, United States
Ahmed Mahmoud Hamdy,
Assiut University, Egypt

*CORRESPONDENCE
Meng Wang
wangmeng@hrbmu.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

RECEIVED 10 September 2022
ACCEPTED 28 February 2023
PUBLISHED 15 March 2023

CITATION
Wu Y, Yang J, Qiao X, Li Y, Zhao R, Lin T, Li X and Wang M (2023) Use of the prognostic nutrition index as a predictive biomarker in small-cell lung cancer patients undergoing immune checkpoint inhibitor treatment in the Chinese alpine region.
Front. Oncol. 13:1041140.
doi: 10.3389/fonc.2023.1041140

COPYRIGHT
© 2023 Wu, Yang, Qiao, Li, Zhao, Lin, Li 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.

Use of the prognostic nutrition index as a predictive biomarker in small-cell lung cancer patients undergoing immune checkpoint inhibitor treatment in the Chinese alpine region

Yunjiao Wu[1], Jing Yang[2], Xinyi Qiao[1], Yingjie Li[1], Rui Zhao[1], Tie Lin[3†], Xiaoli Li[1†] and Meng Wang[1*]

[1]Department of Respiratory Medical Oncology, Harbin Medical University Cancer Hospital, Heilongjiang, Harbin, China, [2]Chongqing Engineering Research Center for Processing and Storage of Distinct Agricultural Products, Chongqing Technology and Business University, Chongqing, China, [3]Department of Surgery, The First Affiliated Hospital of Harbin Medical University, Heilongjiang, Harbin, China

Background: Whether the prognostic nutritional index (PNI), which is suggested to reflect systemic inflammation and nutritional status of patients, could be used as an effective prognostic factor for small-cell lung cancer (SCLC) has not yet been clarified. The purpose of this study was to verify the prognostic value of the PNI in SCLC patients treated with programmed cell death ligand-1/programmed cell death 1 (PD-L1/PD-1) inhibitors in the alpine region of China.

Methods: SCLC patients treated with PD-L1/PD-1 inhibitors monotherapy or combined with chemotherapy between March 2017 and May 2020 were included. Based on the values of serum albumin and total lymphocyte count, the study population was divided into two groups: high and low PNI. The Kaplan-Meier method was used to compute the median survival time and the log-rank test was used to compare the two groups. To evaluate the prognostic value of the PNI, univariable and multivariable analyses of progression-free survival (PFS) and overall survival (OS) were performed. The correlations between PNI and DCR or ORR were calculated by Point biserial correlation analysis.

Results: One hundred and forty patients were included in this study, of which, 60.0% were high PNI (PNI > 49.43) and 40.0% were low PNI (PNI ≤ 49.43). Results indicated that the high PNI group had better PFS and OS than the low PNI group in the patients who received PD-L1/PD-1 inhibitors monotherapy (median PFS: 11.0 vs. 4.8 months, $p < 0.001$ and median OS: 18.5 vs. 11.0 months, $p = 0.004$). Similarly, better PFS and OS were associated with an increase in PNI level in the patients who accepted PD-L1/PD-1 inhibitors combined with chemotherapy (median PFS: 11.0 vs. 5.3 months, $p < 0.001$ and median OS: 17.9 vs. 12.6 months, $p = 0.005$). Multivariate Cox-regression model showed that high PNI was significantly related to better PFS and OS in patients who accepted PD-L1/PD-1 inhibitors monotherapy or combined with chemotherapy (PD-L1/PD-1 inhibitors monotherapy: PFS: HR = 0.23, 95% CI: 0.10–0.52, $p < 0.001$ and OS:

HR = 0.13, 95% CI: 0.03–0.55, p = 0.006; PD-L1/PD-1 inhibitors combined with chemotherapy: PFS: HR = 0.34, 95% CI: 0.19–0.61, p < 0.001 and OS: HR = 0.53, 95% CI: 0.29–0.97, p = 0.040, respectively). Additionally, Point biserial correlation analysis between PNI and disease control rate (DCR) showed that PNI status was positively correlated with DCR in SCLC patients receiving PD-L1/PD-1 inhibitors or combined with chemotherapy (r = 0.351, p < 0.001; r = 0.285, p < 0.001, respectively).

Concussions: PNI may be a promising biomarker of treatment efficacy and prognosis in SCLC patients treated with PD-L1/PD-1 inhibitors in the alpine region of China.

KEYWORDS

prognostic nutritional index (PNI), small cell lung cancer (SCLC), PD-L1/PD-1 inhibitors, immune-related adverse events (irAEs), prognostic factor

Introduction

Lung cancer continues to be the leading cause of cancer-related deaths worldwide, particularly in the alpine region of China. Small-cell lung cancer (SCLC) accounts for 15% of all lung cancers and has a 5-year survival of 1–2%, as most patients present with late-stage disease (1–4). Immune checkpoint inhibitors (ICIs) such as programmed cell death ligand-1/programmed cell death 1 (PD-L1/PD-1) inhibitors have revolutionized the therapeutic paradigm of SCLC. Gay et al. found that molecular subtypes classified as "SCLC-I" derived a significant overall survival (OS) benefit from immune checkpoint blockade (HR = 0.57, 95% CI: 0.32–1.00), suggesting patients with advanced SCLC may benefit from ICIs (5). Nevertheless, the absolute improvements in progression free survival (PFS) and OS are not occur to all SCLC patients. Therefore, there is an urgent need to determine an appropriate biomarker to identify which SCLC patients may benefit from PD-L1/PD-1 inhibitor treatment (6, 7).

Related studies have shown that PD-L1 expression is low or absent in SCLC patients (8–11). Recently, Iams et al. has demonstrated that even PD-L1 negative patients could responded well to inhibitor treatment (12). Therefore, PD-L1 expression is not used as a predictive biomarker in SCLC patients receiving PD-L1/PD-1 inhibitor treatment. Tumor mutational burden (TMB) has been demonstrated to be related to the efficacy of PD-L1/PD-1 inhibitor treatment in several large clinical trials (13, 14). However, the TMB does not have a clear cut-off value. Therefore, further studies are needed to generate a consensus on utilizing the TMB in clinical practice. Additionally, microsatellite instability (MSI) and tumor-infiltrating lymphocytes (TILs) are promising predictive biomarkers of ICIs response that warrant further evaluation (15, 16). Currently, PD-L1/PD-1 inhibitor treatment in SCLC patients lacks robust indicators for determining which patients will benefit.

Typically, obtaining tumor specimens during the process of treatment is difficult. Therefore, a non-invasive biomarker that can conveniently predict the efficacy and prognosis of immunotherapy is needed. The progression of SCLC is closely associated with systemic inflammation and nutritional status. Several studies have shown that inflammatory and nutritional peripheral blood parameters may be potential biomarkers of the effects of immunotherapy outcomes in patients with melanoma and head and neck squamous cell carcinoma (HNSCC) (17–19).

Prognostic nutritional index (PNI) is obtained by the level of serum albumin and peripheral lymphocytes and is proposed to assess immune-nutritional status (20). A few studies have shown that low PNI status is associated with an unfavorable prognosis in gastrointestinal and colorectal cancer (21, 22).

Owing to the roles played by the PNI, we hypothesized that there is a relationship between treatment response and the PNI in SCLC patients. The aim of the present study was to investigate whether pretreatment PNI is a predictive biomarker in SCLC patients undergoing PD1/PD-L1 treatment in the Chinese Alpine Region.

Patients and methods

Patients selection

We retrospectively enrolled 140 SCLC patients undergoing PD-L1/PD-1 inhibitor treatment at the Harbin Medical University Cancer Hospital between March 2017 and May 2020. The inclusion criteria for SCLC patients treated with PD-L1/PD-1 inhibitors were as follows (1): patients who were diagnosed with SCLC by histopathology or cytopathology; (2) Eastern Cooperative Oncology Group performance status (ECOG PS) ≤ 2 points; (3) at least two cycles of PD-L1/PD-1 inhibitor therapy, PD-L1/PD-1 inhibitors combined with chemotherapy or PD-L1/PD-1 inhibitors as monotherapy. The exclusion criteria for SCLC patients treated with PD-L1/PD-1 inhibitors were as follows: (1) lack of complete clinicopathological information or laboratory data before PD-L1/PD-1 inhibitor treatment;

(2) patients with malignancy in other organs or hematological diseases, autoimmune diseases, and systemic immunosuppression. Basic clinical and pathological data were collected from patients who met these criteria. This study was approved by the Institutional Review Board of the Harbin Medical University Cancer Hospital.

Data collection and definitions

Primary laboratory data from before PD-L1/PD-1 inhibitor treatment and clinicopathologic data were retrieved from SCLC patient medical records. The PNI was calculated as 10 × serum albumin (g/dl) + 0.005 × total lymphocyte count (per mm^3) (20).

Evaluation

Two doctors independently evaluated drug effectiveness based on image examinations every 8–12 weeks, according to the Response Evaluation Criteria in Solid Tumors guidelines version 1.1 (RECIST1.1). Immune-related adverse events (irAEs) were assessed by two doctors independently according to the Common Terminology Criteria for Adverse Events (CTCAE) of the National Cancer Institute (version 4.03). The disease control rate (DCR) was defined as complete plus partial response plus stable disease, and the overall response rate (ORR) was defined as complete plus partial response. OS was defined as the time from the start of treatment with PD-L1/PD-1 inhibitors to death. PFS was defined as the time from the start of treatment with PD-L1/PD-1 inhibitors to disease progression.

Statistical analyses

Statistical analyses were performed using IBM SPSS Statistic 25.0. The clinicopathological data of SCLC patients was compared by Chi-squared or Fisher's exact test. Continuous variables were compared using Student's *t* test. We generated ROC curves and calculated the optimal cut-off values for PNI, lymphocytes and albumin. The Kaplan-Meier method and log-rank test were used to perform survival analyses. Univariate and multivariate Cox regression analyses were used to identify independent prognostic factors associated with OS and PFS. The correlations between PNI and DCR or ORR were calculated by Point biserial correlation analysis. The impact of the PNI on DCR and ORR is represented by a column diagram. A p value < 0.05 was considered statistically significant.

Results

Patient clinical characteristics and outcomes

A total of 140 patients with SCLC treated with PD-L1/PD-1 inhibitors were identified for our analysis. Among these, 60.0% (n = 84) of the patients had high PNI and 40.0% (n = 56) had low PNI. Further, 43.6% (n = 61) of the patients received PD-1 inhibitor treatment and 56.4% (n = 79) received PD-L1 inhibitor treatment. Of all the patients that received PD-L1/PD-1 inhibitor therapy (n = 140), 43.6% (n = 61) received PD-L1/PD-1 inhibitors monotherapy. In addition, ECOG PS was 0–1 in 80.7% (n = 113), and 2 in 19.3% (n = 27) of patients. Moreover, 22.1% (n = 31) of patients had brain metastases, 31.4% (n = 44) had liver metastases, 32.9% (n = 46) had bone metastases, 38.6% (n = 54) had pleural or pericardial metastases, and 11.4% (n = 16) had adrenal metastases. Characteristics for the entire cohort and different PNI groups are summarized in Table 1.

Identification of PNI, albumin, and lymphocyte cut-off values

The PNI, lymphocyte, and albumin values in SCLC patients ranged from 35.15 to 65.10, 0.54 to 2.91, and 30.90 to 55.00, respectively. We generated ROC curves and calculated the optimal cut-off values for PNI, lymphocytes, and albumin (Figure 1). The optimal cut-off values for PNI, lymphocytes, and albumin were 49.43, 1.93, and 43.40, respectively. Based on the optimal cut-off values, we divided patients into two groups for further analysis: Low PNI (PNI ≤ 49.43) and High PNI (PNI > 49.43), Low Lymphocyte (LYM) (LYM ≤ 1.93) and High LYM (LYM > 1.93), or Low Albumin (ALB) (ALB ≤ 43.40) and High ALB (ALB > 43.40).

Association between the PNI and the prognostic utility of SCLC patients received PD-L1/PD-1 inhibitors treatment

As shown in Figure 2, a high PNI was associated with better PFS in the patients who accepted PD-L1/PD-1 inhibitors monotherapy or combined with chemotherapy (median PFS: 11.0 vs. 4.8 months, HR = 0.28, $p < 0.001$ and median PFS: 11.0 vs. 5.3 months, HR = 0.28, $p < 0.001$) (Figures 2A, B). Similarly, improved OS was associated with high PNI relative to low PNI in the patients who accepted PD-L1/PD-1 inhibitors monotherapy or combined with chemotherapy (median OS: 18.5 vs. 11.0 months, HR = 0.33, $p = 0.004$ and median OS: 17.9 vs. 12.6 months, HR = 0.48, $p = 0.005$) (Figures 2C, D). The results revealed that an elevated PNI was associated with significantly lower risk of death in patients who accepted PD-L1/PD-1 inhibitors monotherapy or combined with chemotherapy (all $p < 0.05$).

Association between the PNI and the predictive utility of SCLC patients received PD-L1/PD-1 inhibitors treatment

Of the 61 patients treated with PD-L1/PD-1 inhibitors monotherapy, 7 (25.0%) experienced progressive disease (PD), 11 (39.3%) experienced stable disease (SD), 8 (28.6%) experienced partial response (PR), and 2 (7.1%) experienced complete response (CR) in the High PNI group. This is compared with 23 (69.7%) patients with PD, 7 (21.2%) with SD, and 3 (9.1%) with PR in the low PNI group (Figure 3A). For patients received PD-L1/PD-1

10.3389/fonc.2023.1041140

TABLE 1 Clinical characteristics of the 140 patients received PD-L1/PD-1 inhibitors treatment.

Clinical characteristics	Total [n (%)]	High PNI [n (%)]	Low PNI [n (%)]	*p* value
Total	140 (100)	84 (60.0)	56 (40.0)	
Age				0.015
≤60	60 (42.9)	43 (51.2)	17 (30.4)	
> 60	80 (57.1)	41 (48.8)	39 (69.6)	
Gender				0.032
Male	73 (52.1)	50 (59.5)	23 (41.1)	
Female	67 (47.9)	34 (40.5)	33 (58.9)	
Smoking history				0.227
Yes	112 (80.0)	70 (83.3)	42 (75.0)	
No	28 (20.0)	14 (16.7)	14 (25.0)	
ECOG PS				0.007
0-1	113 (80.7)	74 (88.1)	39 (69.6)	
2	27 (19.3)	10 (11.9)	17 (30.4)	
Stage				
Extended stage	125 (89.3)	72 (85.7)	53 (94.6)	0.094
Limited stage	15 (10.7)	12 (14.3)	3 (5.4)	
Therapy line				0.269
1	68 (48.6)	44 (52.4)	24 (42.9)	
≥2	72 (51.4)	40 (47.6)	32 (57.1)	
Immunotherapy drug				0.210
PD-1	61 (43.6)	33 (39.3)	28 (50.0)	
PD-L1	79 (56.4)	51 (60.7)	28 (50.0)	
Regimen				0.003
Combination therapy	79 (56.4)	56 (66.7)	23 (41.1)	
Monotherapy	61 (43.6)	28 (33.3)	33 (58.9)	
Brain metastases				0.561
Yes	31 (22.1)	20 (23.8)	11 (19.6)	
No	109 (77.9)	64 (76.2)	45 (80.4)	
Liver metastases				0.372
Yes	44 (31.4)	24 (28.6)	20 (35.7)	
No	96 (68.6)	60 (71.4)	36 (64.3)	
Bone metastases				0.883
Yes	46 (32.9)	28 (33.3)	18 (32.1)	
No	94 (67.1)	56 (66.7)	38 (67.9)	
Pleural or pericardial metastases				0.009
Yes	54 (38.6)	25 (29.8)	29 (51.8)	
No	86 (61.4)	59 (70.2)	27 (48.2)	
Adrenal metastases				0.159

(Continued)

TABLE 1 Continued

Clinical characteristics	Total [n (%)]	High PNI [n (%)]	Low PNI [n (%)]	*p* value
Yes	16 (11.4)	7 (8.3)	9 (16.1)	
No	124 (88.6)	77 (91.7)	47 (83.9)	
IrAEs				
Yes	52 (37.1)	38 (45.2)	14 (25.0)	0.015
No	88 (62.9)	46 (54.8)	42 (75.0)	

PD-L1/PD-1, programmed death-ligand 1/programmed death-1; PNI, prognostic nutritional index; ECOG-PS, Eastern Cooperative Oncology Group Performance Status; irAEs, immune-related adverse events.

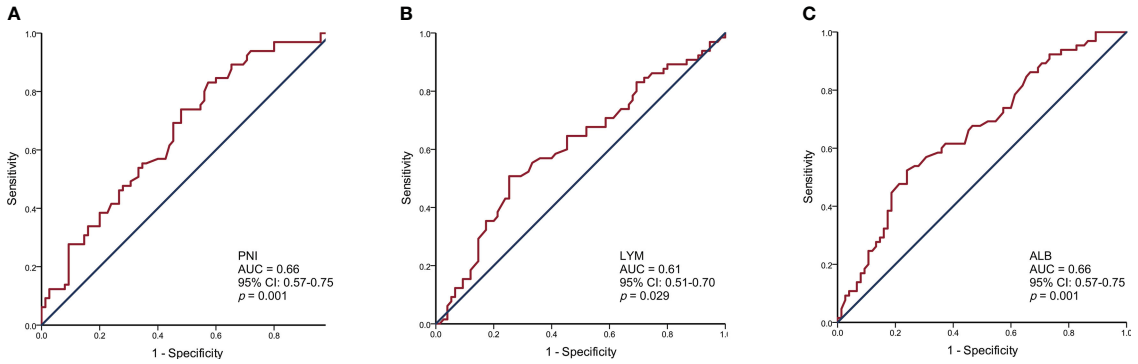


FIGURE 1
Receiver operating characteristic curve analysis for the prognostic nutritional index **(A)**, lymphocytes **(B)**, and albumin **(C)** to predict the prognosis of SCLC patients.

inhibitors combined with chemotherapy, of the 84 patients in the High PNI group, 24 (28.6%) experienced PD, 35 (41.7%) experienced SD, 21 (25.0%) experienced PR, and 4 (4.8%) experienced CR in the High PNI group. Among the 56 patients in the Low PNI group, 31 (55.4%) experienced PD, 13 (23.2%) experienced SD, 10 (17.9%) experienced PR and 2 (3.6%) experienced CR (Figure 3B).

Moreover, the results of Point biserial correlation analysis between PNI and DCR showed that patients who had a higher increase in PNI trend had better DCR compared with those with a higher decrease in SCLC patients treated with PD-L1/PD-1 inhibitors monotherapy or combined with chemotherapy (r = 0.351, $p < 0.001$; r = 0.285, $p < 0.001$, respectively) (Tables 2, 3). In addition, compared with patients with high PNI, those with low PNI experienced worse ORR to PD-L1/PD-1 inhibitors monotherapy or combined with chemotherapy, although a benefit was noted, it was not statistically significant (r = 0.237, p =0.066; r = 0.106, p = 0.211, respectively) (Tables 2, 3).

Univariate and multivariate survival analyses of PFS and OS

For patients received PD-L1/PD-1 inhibitors monotherapy, univariate Cox regression analysis showed that irAEs (p = 0.014) and PNI ($p < 0.001$) were significantly associated with PFS. Similarly, OS was associated with liver metastases (p = 0.027) and PNI (p = 0.007) (Figures 4A, B). Moreover, in patients received PD-L1/PD-1 inhibitors combined with chemotherapy, univariate Cox regression analysis showed that therapy line (p = 0.001), regimen ($p < 0.001$), irAEs (p = 0.004), and PNI ($p < 0.001$) significantly affected PFS. In parallel, OS was significantly associated with stage (p = 0.005), liver metastases (p = 0.001), irAEs (p = 0.014), and PNI (p = 0.006) (Figures 4C, D).

Multivariate Cox-regression model showed that high PNI was significantly related to better PFS and OS in patients who accepted PD-L1/PD-1 inhibitors monotherapy or combined with chemotherapy (PD-L1/PD-1 inhibitors monotherapy: PFS: HR = 0.23, 95% CI: 0.10–0.52, $p < 0.001$ and OS: HR = 0.13, 95% CI: 0.03–0.55, p = 0.006; PD-L1/PD-1 inhibitors combined with chemotherapy: PFS: HR = 0.34, 95% CI: 0.19–0.61, $p < 0.001$ and OS: HR = 0.53, 95% CI: 0.29–0.97, p = 0.040, respectively) (Figures 5A–D). In addition, irAEs was also an independent prognostic factor for better OS in patients who accepted PD-L1/PD-1 inhibitors combined with chemotherapy (HR = 0.38, 95% CI: 0.22–0.67, p = 0.001) (Figure 5D). These results demonstrated that PNI was an independent prognostic factor for PFS and OS in patients who accepted PD-L1/PD-1 inhibitors or combined with chemotherapy.

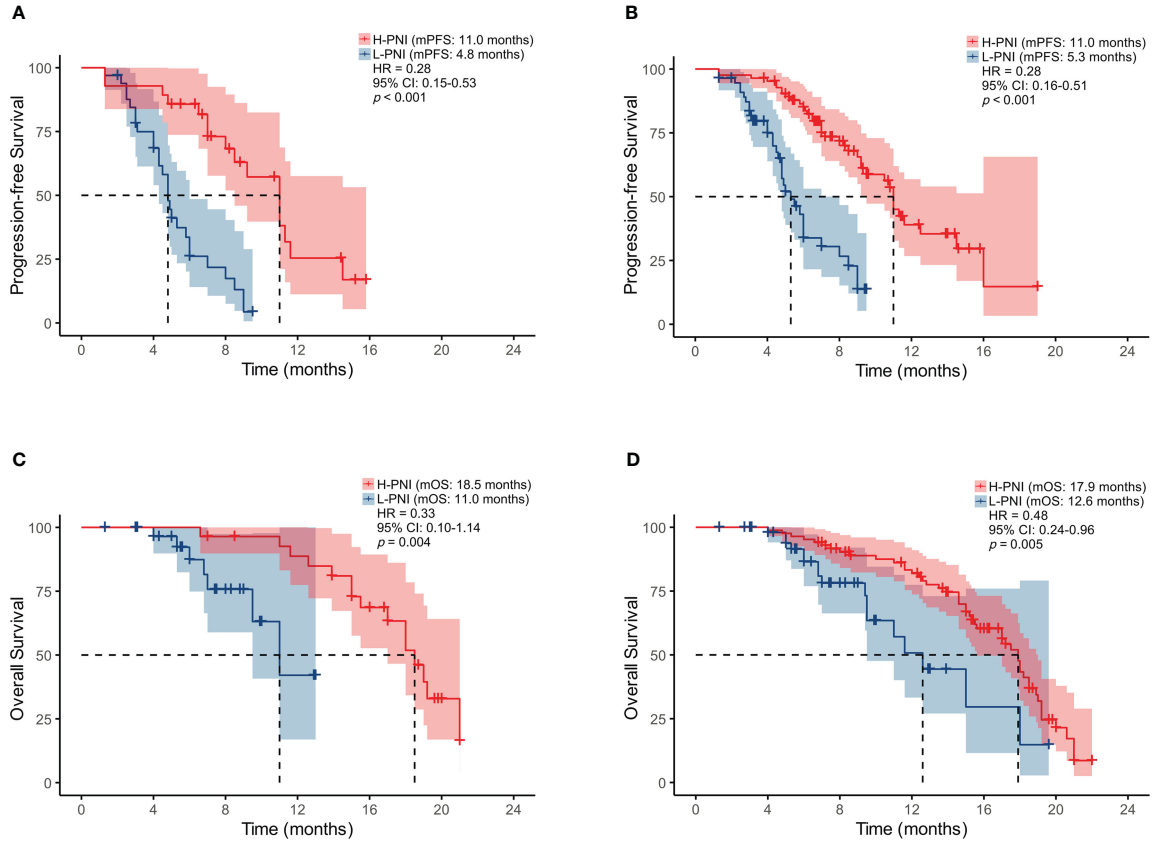


FIGURE 2
Kaplan-Meier survival curves displaying PFS according to PNI group in SCLC patients treated with **(A)** PD-L1/PD-1 inhibitors monotherapy or **(B)** combined with chemotherapy; OS by PNI group in SCLC patients treated with **(C)** PD-L1/PD-1 inhibitors monotherapy or **(D)** combined with chemotherapy.

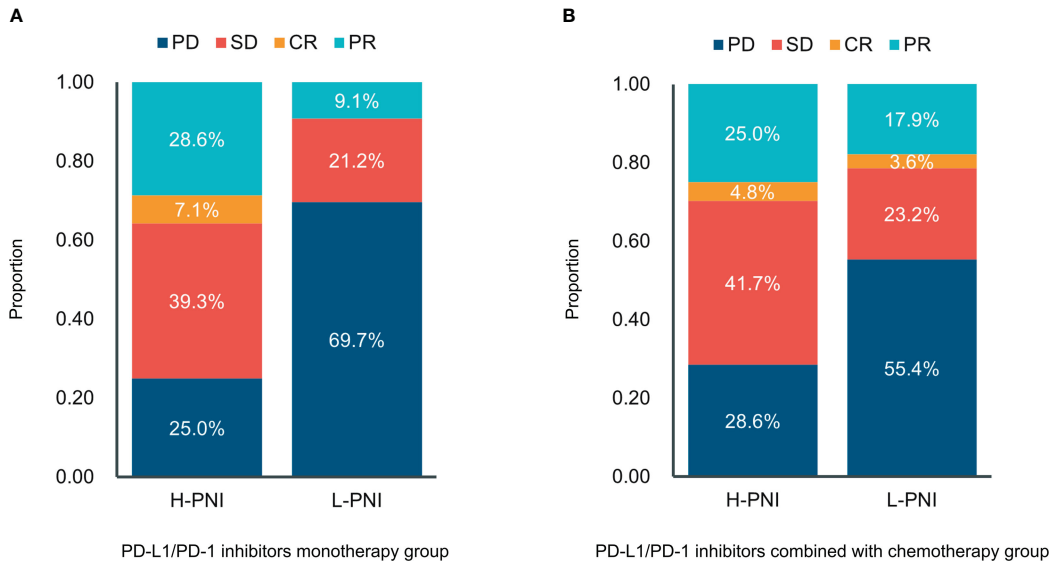


FIGURE 3
Distribution between responses and the PNI groups in patients with **(A)** PD-L1/PD-1 inhibitors monotherapy and **(B)** PD-L1/PD-1 inhibitors combined with chemotherapy. CR, complete response; PR, partial response; SD—stable disease; PD, progressive disease.

TABLE 2 Relationship between clinical response and PNI groups in SCLC patients treated with PD-L1/PD-1 inhibitors monotherapy.

Response	PNI Group [(n%)]		*p* value	Point biserial correlation coefficient PNI level	*p* value
	High PNI	Low PNI			
DCR			0.001	0.351	<0.001
CR+PR+SD	21(75.0)	10(30.3)			
PD	7(25.0)	23(69.7)			
ORR			0.011	0.237	0.066
CR+PR	10(35.7)	3(9.1)			
SD+PD	18(64.3)	30(90.9)			

PNI, prognostic nutritional index; SCLC, small cell lung cancer; CR, complete response; PR, partial response; SD, stable disease; PD, progressive disease; DCR, disease control rate; ORR, overall response rate.

TABLE 3 Relationship between clinical response and PNI groups in SCLC patients treated with PD-L1/PD-1 inhibitors combined with chemotherapy.

Response	PNI Group [(n%)]		*p* value	Point biserial correlation coefficient PNI level	*p* value
	High PNI	Low PNI			
DCR			0.001	0.285	<0.001
CR+PR+SD	60(71.4)	25(44.6)			
PD	24(28.6)	31(55.4)			
ORR			0.273	0.106	0.211
CR+PR	25(29.8)	12(21.4)			
SD+PD	59(70.2)	44(78.6)			

PNI, prognostic nutritional index; SCLC, small cell lung cancer; CR, complete response; PR, partial response; SD, stable disease; PD, progressive disease; DCR, disease control rate; ORR, overall response rate.

Immune-related adverse events (irAEs)

In our study, 37.1% (n = 52) patients experienced six different irAEs of any grade. Among these, 25 (48.1%) experienced rash, 9 (17.3%) experienced hypothyroidism, 7 (13.5%) experienced liver dysfunction, 7 (13.5%) experienced infusion reaction, 1 (1.9%) experienced impaired glucose regulation, and 3 (5.8%) experienced diarrhea. The most common severe irAE (grade ≥ 3) was rash (11.5%, n = 6). The median PFS of the 52 patients with irAEs was significantly better than the 88 patients without irAEs (11.3 vs. 8.0 months, $p = 0.003$) (Figure 6A). Similarly, the median OS of the 52 patients with irAEs was significantly better than the 88 patients without irAEs (18.5 vs. 14.6 months, $p = 0.011$) (Figure 6B).

The High ALB, High LYM, and High PNI groups were composed of 24 (46.2%), 25 (48.1%), and 38 (45.2%) patients, respectively (Table 4). In univariate logistic regression analyses, the High LYM group and the High PNI group were significantly associated with any grade of irAEs ($p = 0.041$, $p = 0.017$). However, the PNI was not an independent prognostic risk factor of the onset of irAEs in the multivariate logistic regression analysis ($p = 0.085$).

Discussion

The emergence of PD-L1/PD-1 inhibitors has brought hope to patients with advanced SCLC, but the 5-year survival rate for patients remains low. Therefore, effective, reliable and easily accessible predictive biomarkers are urgently needed for identifying patients that will benefit from treatment. Currently, more focus has been turned toward the correlation between inflammatory-immune nutritional status and the clinical outcomes of cancer patients who are undergoing PD-L1/PD-1 inhibitor treatment. Systemic inflammation is closely associated with disease promotion and progression in most cancers, including lung cancer (23). PNI is obtained based on serum albumin and peripheral lymphocyte levels, which can reflect the nutritional and immune status of patients. In advanced head and neck cancers, a low PNI has been shown to positively correlate with worse survival and worse response rates to PD-L1/PD-1 inhibitors (19). However, whether the PNI can be used as a strong prognostic factor for SCLC patients has not yet been clarified. The aim of this study was to verify the predictive value of PNI for survival, treatment response rates, and treatment-related toxicity in SCLC patients of the China alpine region undergoing PD-L1/PD-1 inhibitor treatment.

In this retrospective study, the data were collected from the alpine region of China. The characteristics of the alpine region Chinese lung cancer population differ considerably from other lung cancer populations. The burden of lung cancer attributable to tobacco and PM 2.5 concentration in China alpine region remains heavy (24). As shown in Table 1, approximately 80% patients had a history of smoking. Our findings showed that low

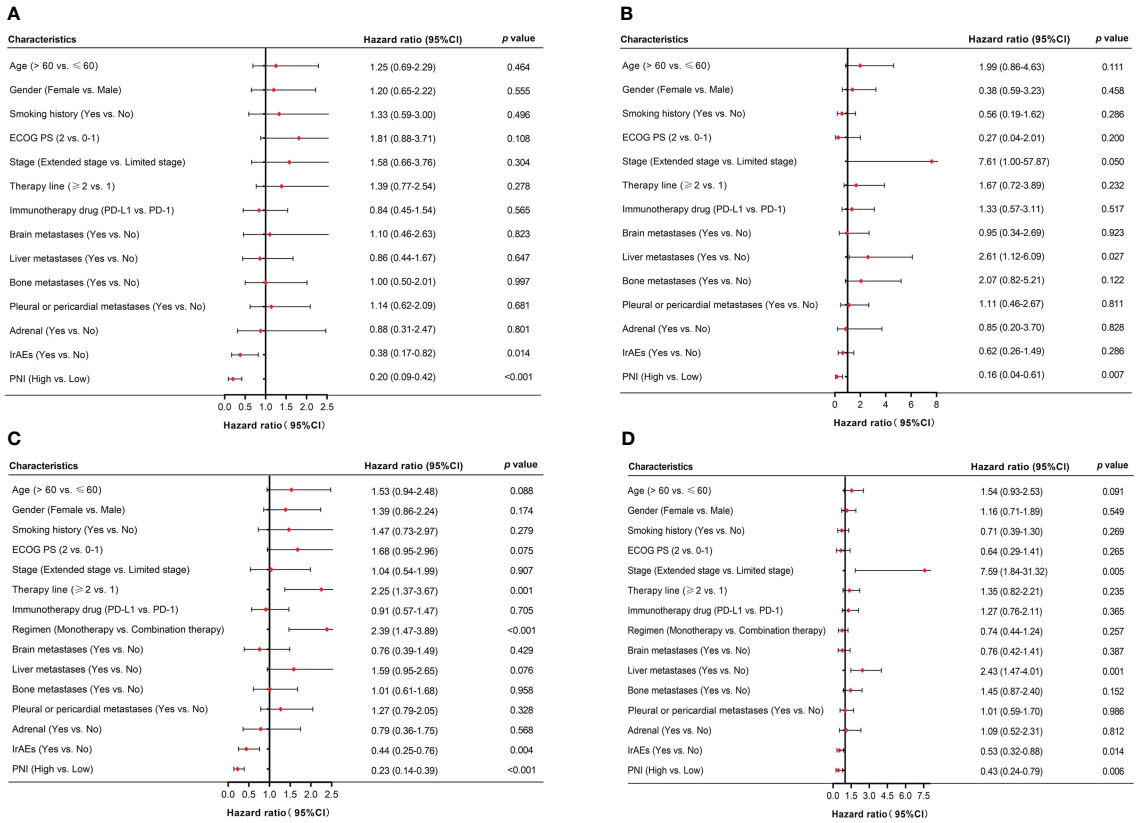


FIGURE 4
Univariate analysis of factors associated with PFS and OS in SCLC patients. **(A, B)** PFS and OS in SCLC patients treated with PD-L1/PD-1 inhibitors monotherapy. **(C, D)** PFS and OS in SCLC patients treated with PD-L1/PD-1 inhibitors combined with chemotherapy.

PNI was independently associated with worse PFS and OS in SCLC patients receiving PD-L1/PD-1 inhibitors monotherapy or combined with chemotherapy. Moreover, the correlation analysis showed that PNI status was positively correlated with DCR in SCLC patients receiving PD-L1/PD-1 inhibitors monotherapy or combined with chemotherapy (r = 0.351, $p < 0.001$; r = 0.285, $p < 0.001$, respectively). Johannet et al. also found that low PNI was associated with worse survival and treatment response rates in

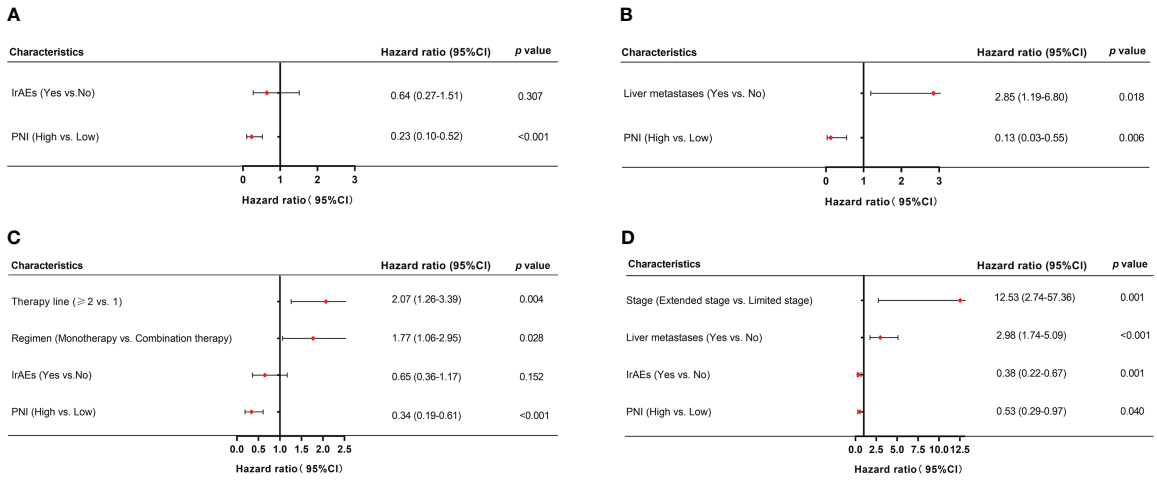


FIGURE 5
Multivariate analysis of factors associated with PFS and OS in SCLC patients. **(A, B)** PFS and OS in SCLC patients treated with PD-L1/PD-1 inhibitors monotherapy. **(C, D)** PFS and OS in SCLC patients treated with PD-L1/PD-1 inhibitors combined with chemotherapy.

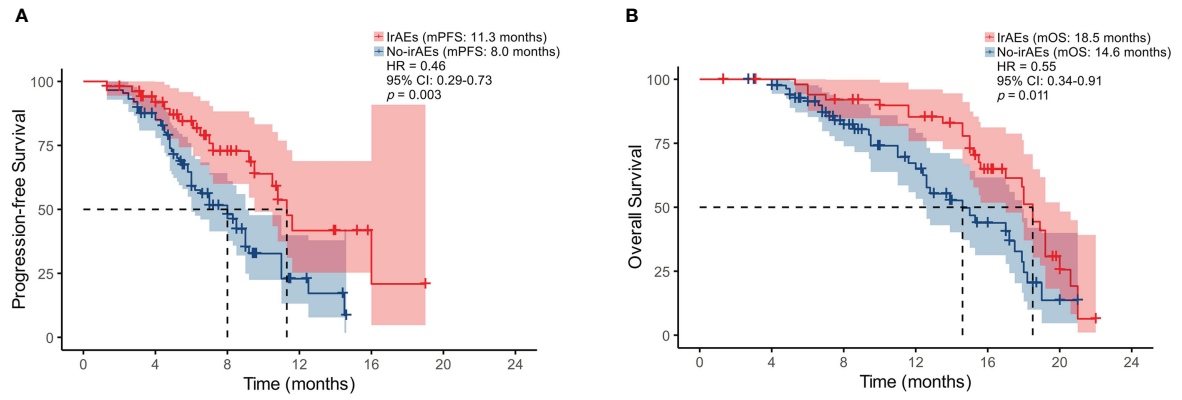


FIGURE 6
Kaplan-Meier analysis of PFS **(A)** and OS **(B)** based on the onset of irAEs.

TABLE 4 Levels of peripheral blood markers by irAEs development.

Blood markers	irAEs, n (%)	Univariate		Multivariate	
		OR (95% CI)	*p* value	OR (95% CI)	*p* value
L-ALB (n=88)	28 (31.8)	1			
H-ALB (n=52)	24 (46.2)	1.84 (0.91–3.72)	0.091		
L-LYM (n=88)	27 (30.7)	1			
H-LYM (n=52)	25 (48.1)	2.09 (1.03–4.25)	0.041	1.54 (0.70–3.38)	0.279
Low PNI (n=56)	14 (25.0)	1		1	
High PNI (n=84)	38 (45.2)	2.48 (1.18–5.20)	0.017	2.05 (0.90–4.65)	0.085

OR, odds ratio; CI, confidence interval; irAEs, immune-related adverse events; ALB, albumin; LYM, lymphocyte; PNI, prognostic nutritional index; L–ALB, Low–ALB; L–LYM, Low–LYM; H–ALB, High ALB; H–LYM, High LYM.

patients with liver cancer, melanoma, and uterine cancer (25). Immune-nutritional status prognosticates a response in SCLC patients treated with PD-L1/PD-1 inhibitors. Chronic inflammation associated with malnutrition inhibited adaptive immune system activation. Heightened levels of the proinflammatory cytokine interleukin-6 (IL-6) could induce endogenous steroid release and could further dampen immune cell functions, consequently reducing the effectiveness of PD-L1/PD-1 inhibitors (26, 27). Furthermore, T cells must acquire adequate nutrients to engage the metabolism which supports their functions. Metabolic competition between T cells and tumor cells in the tumor microenvironment leads to T cell hyporesponsiveness and further impairs PD-L1/PD-1 inhibitor efficacy (28, 29). Poor immune-nutritional status limits a response to PD-L1/PD-1 inhibitor treatment in SCLC patients and further leads to worse prognosis. The present analysis showed low PNI was significantly correlated with worse survival and a lower treatment response rate in SCLC patients treated with PD-L1/PD-1 inhibitors in the China alpine region population.

The occurrence of irAEs limits the use of ICIs. Therefore, early recognition and prompt intervention are particularly important. Wang et al. reported that 66% of patients undergoing PD-L1/PD-1 inhibitor monotherapy developed at least 1 irAEs of any grade in multiple solid tumor types (30). Our current research showed that patients with irAEs had better PFS and OS compared to those without irAEs ($p = 0.003$, $p = 0.011$), and these patients usually had a higher PNI status. Seiwert et al. also explored the association between the development of irAEs and prolonged OS in patients with head and neck cancer receiving ICIs (31). They demonstrated that ORR was higher for patients with irAEs compared to those without irAEs (30.6% vs. 12.3%, $p = 0.020$). Additionally, we explored an association between irAEs and peripheral blood markers and found that high PNI showed a trend towards being a prognostic factor for any grade of irAEs but did not reach the level of statistical significance ($p = 0.085$).

The present study demonstrated that pretreatment PNI is a promising efficacy and prognostic biomarker in SCLC patients treated with PD-L1/PD-1 inhibitors. Monitoring PNI status prior to PD-L1/PD-1 inhibitor treatment may significantly improve survival rate, current preventive and treatment approaches, and enhance accurate personal management of SCLC patients. Furthermore, the PNI can be easily calculated

from peripheral blood counts, avoiding the need to obtain tumor specimens during the treatment process. Further prospective studies with larger sample sizes are necessary to confirm and support our conclusions.

In conclusion, we have demonstrated that low PNI was significantly correlated with worse survival and a lower treatment response rate, supporting its use as an effective biomarker in SCLC patients treated with PD-L1/PD-1 inhibitors. Improving nutrition and immune status by monitoring the PNI status of SCLC patients prior to PD-L1/PD-1 inhibitor treatment may optimize treatment efficacy and improve prognosis.

Data availability statement

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

Author contributions

Conceptualization and methodology: JYW, JY, YXQ, JYL, RZ, TL, LXL, and MW. Formal analysis, investigation, and data curation: JYW, TL, LXL, and MW. Writing—original draft preparation: JYW. Writing—review and editing: JYW, TL, LXL, and MW. Supervision and project administration: MW. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by Grants from the Natural Science Foundation of Heilongjiang Province (No. LH2019H039), and the Haiyan Foundation of Harbin Medical University Cancer Hospital (No. JJZD2020-09).

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

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. Wang S, Zimmermann S, Parikh K, Mansfield AS, Adjei AA. Current diagnosis and management of small-cell lung cancer. *Mayo Clin Proc* (2019) 94(8):1599–622. doi: 10.1016/j.mayocp.2019.01.034

3. Byers LA, Rudin CM. Small cell lung cancer: Where do we go from here? *Cancer* (2015) 121(5):664–72. doi: 10.1002/cncr.29098

4. Rudin CM, Ismaila N, Hann CL, Malhotra N, Movsas B, Norris K, et al. Treatment of small-cell lung cancer: American society of clinical oncology endorsement of the American college of chest physicians guideline. *J Clin Oncol* (2015) 33(34):4106–11. doi: 10.1200/JCO.2015.63.7918

5. Gay CM, Stewart CA, Park EM, Diao L, Groves SM, Heeke S, et al. Patterns of transcription factor programs and immune pathway activation define four major subtypes of SCLC with distinct therapeutic vulnerabilities. *Cancer Cell* (2021) 39 (3):346–360.e7. doi: 10.1016/j.ccell.2020.12.014

6. Liu SV, Reck M, Mansfield AS, Mok T, Scherpereel A, Reinmuth N, et al. Updated overall survival and PD-L1 subgroup analysis of patients with extensive-stage small-cell lung cancer treated with atezolizumab, carboplatin, and etoposide (IMpower133). *J Clin Oncol* (2021) 39(6):619–30. doi: 10.1200/JCO.20.01055

7. Paz-Ares L, Dvorkin M, Chen Y, Reinmuth N, Hotta K, Trukhin D, et al. Durvalumab plus platinum-etoposide versus platinum-etoposide in first-line treatment of extensive-stage small-cell lung cancer (CASPIAN): a randomised, controlled, open-label, phase 3 trial. *Lancet* (2019) 394(10212):1929–39. doi: 10.1016/S0140-6736(19)32222-6

8. Antonia SJ, López-Martin JA, Bendell J, Ott PA, Taylor M, Eder JP, et al. Nivolumab alone and nivolumab plus ipilimumab in recurrent small-cell lung cancer (CheckMate 032): A multicentre, open-label, phase 1/2 trial. *Lancet Oncol* (2016) 17 (7):883–95. doi: 10.1016/S1470-2045(16)30098-5

9. Ott PA, Elez E, Hiret S, Kim DW, Morosky A, Saraf S, et al. Pembrolizumab in patients with extensive-stage small-cell lung cancer: Results from the phase ib KEYNOTE-028 study. *J Clin Oncol* (2017) 35(34):3823–9. doi: 10.1200/JCO.2017.72.5069

10. Schultheis AM, Scheel AH, Ozretić L, George J, Thomas RK, Hagemann T, et al. PD-L1 expression in small cell neuroendocrine carcinomas. *Eur J Cancer* (2015) 51 (3):421–6. doi: 10.1016/j.ejca.2014.12.006

11. Boumber Y. Tumor mutational burden (TMB) as a biomarker of response to immunotherapy in small cell lung cancer. *J Thorac Dis* (2018) 10(8):4689–93. doi: 10.21037/jtd.2018.07.120

12. Iams WT, Porter J, Horn L. Immunotherapeutic approaches for small-cell lung cancer. *Nat Rev Clin Oncol* (2020) 17(5):300–12. doi: 10.1038/s41571-019-0316-z

13. Zheng M. Tumor mutation burden for predicting immune checkpoint blockade response: the more, the better. *J Immunother Cancer* (2022) 10(1):e003087. doi: 10.1136/jitc-2021-003087

14. Chan TA, Yarchoan M, Jaffee E, Swanton C, Quezada SA, Stenzinger A, et al. Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. *Ann Oncol* (2019) 30(1):44–56. doi: 10.1093/annonc/mdy495

15. Wu W, Liu Y, Zeng S, Han Y, Shen H. Intratumor heterogeneity: The hidden barrier to immunotherapy against MSI tumors from the perspective of IFN-γ signaling and tumor-infiltrating lymphocytes. *J Hematol Oncol* (2021) 14(1):160. doi: 10.1186/s13045-021-01166-3

16. Necchi A, Madison R, Raggi D, Jacob JM, Bratslavsky G, Shapiro O, et al. Comprehensive assessment of immuno-oncology biomarkers in adenocarcinoma, urothelial carcinoma, and squamous-cell carcinoma of the bladder. *Eur Urol* (2020) 77(4):548–56. doi: 10.1016/j.eururo.2020.01.003

17. Moreira A, Leisgang W, Schuler G, Heinzerling L. Eosinophilic count as a biomarker for prognosis of melanoma patients and its importance in the response to immunotherapy. *Immunotherapy* (2017) 9(2):115–21. doi: 10.2217/imt-2016-0138

18. Bernard-Tessier A, Jeanville P, Champiat S, Lazarovici J, Voisin AL, Mateus C, et al. Immune-related eosinophilia induced by anti-programmed death 1 or death-ligand 1 antibodies. *Eur J Cancer* (2017) 81:135–7. doi: 10.1016/j.ejca.2017.05.017

19. Guller M, Herberg M, Amin N, Alkhatib H, Maroun C, Wu E, et al. Nutritional status as a predictive biomarker for immunotherapy outcomes in advanced head and neck cancer. *Cancers (Basel)* (2021) 13(22):5772. doi: 10.3390/cancers13225772

20. Hua X, Long ZQ, Huang X, Deng JP, He ZY, Guo L, et al. The value of prognostic nutritional index (PNI) in predicting survival and guiding radiotherapy of patients with T1-2N1 breast cancer. *Front Oncol* (2020) 9:1562. doi: 10.3389/fonc.2019.01562

21. Xie H, Wei L, Yuan G, Liu M, Tang S, Gan J. Prognostic value of prognostic nutritional index in patients with colorectal cancer undergoing surgical treatment. *Front Nutr* (2022) 9:794489. doi: 10.3389/fnut.2022.794489

22. Noh GT, Han J, Cho MS, Hur H, Min BS, Lee KY, et al. Impact of the prognostic nutritional index on the recovery and long-term oncologic outcome of patients with colorectal cancer. *J Cancer Res Clin Oncol* (2017) 143(7):1235–42. doi: 10.1007/s00432-017-2366-x

23. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* (2011) 144(5):646–74. doi: 10.1016/j.cell.2011.02.013

24. Chen W, Xia C, Zheng R, Zhou M, Lin C, Zeng H, et al. Disparities by province, age, and sex in site-specific cancer burden attributable to 23 potentially modifiable risk factors in China: A comparative risk assessment. *Lancet Glob Health* (2019) 7(2):e257–69. doi: 10.1016/S2214-109X(18)30488-1

25. Johannet P, Sawyers A, Qian Y, Kozloff S, Gulati N, Donnelly D, et al. Baseline prognostic nutritional index and changes in pretreatment body mass index associate with immunotherapy response in patients with advanced cancer. *J Immunother Cancer* (2020) 8(2):e001674. doi: 10.1136/jitc-2020-001674

26. Flint TR, Fearon DT, Janowitz T. Connecting the metabolic and immune responses to cancer. *Trends Mol Med* (2017) 23(5):451–64. doi: 10.1016/j.molmed.2017.03.001

27. Giles AJ, Hutchinson MND, Sonnemann HM, Jung J, Fecci PE, Ratnam NM, et al. Dexamethasone-induced immunosuppression: Mechanisms and implications for immunotherapy. *J Immunother Cancer* (2018) 6(1):51. doi: 10.1186/s40425-018-0371-5

28. Chang CH, Qiu J, O'Sullivan D, Buck MD, Noguchi T, Curtis JD, et al. Metabolic competition in the tumor microenvironment is a driver of cancer progression. *Cell* (2015) 162(6):1229–41. doi: 10.1016/j.cell.2015.08.016

29. Kedia-Mehta N, Finlay DK. Competition for nutrients and its role in controlling immune responses. *Nat Commun* (2019) 10(1):2123. doi: 10.1038/s41467-019-10015-4

30. Wang Y, Zhou S, Yang F, Qi X, Wang X, Guan X, et al. Treatment-related adverse events of PD-1 and PD-L1 inhibitors in clinical trials: A systematic review and meta-analysis. *JAMA Oncol* (2019) 5(7):1008–19. doi: 10.1001/jamaoncol.2019.0393

31. Foster CC, Couey MA, Kochanny SE, Khattri A, Acharya RK, Tan YC, et al. Immune-related adverse events are associated with improved response, progression-free survival, and overall survival for patients with head and neck cancer receiving immune checkpoint inhibitors. *Cancer* (2021) 127(24):4565–73. doi: 10.1002/cncr.33780

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 20 March 2023
DOI 10.3389/fonc.2023.1157891

Check for updates

OPEN ACCESS

EDITED BY
Hongda Liu,
Nanjing Medical University, China

REVIEWED BY
Jing Yang,
Chongqing Technology and Business University, China
He Shuishi,
Huazhong University of Science and Technology, China
Fabrizio Minervini,
University of Lucerne, Switzerland
Adam Yongxin Ye,
Boston Children's Hospital and Harvard Medical School, United States
Jing Liu,
University of Pennsylvania, United States

*CORRESPONDENCE
Meiyan Liao
liaomy@whu.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

RECEIVED 03 February 2023
ACCEPTED 06 March 2023
PUBLISHED 20 March 2023

CITATION
Wang J, Zhong F, Xiao F, Dong X, Long Y, Gan T, Li T and Liao M (2023) CT radiomics model combined with clinical and radiographic features for discriminating peripheral small cell lung cancer from peripheral lung adenocarcinoma. *Front. Oncol.* 13:1157891. doi: 10.3389/fonc.2023.1157891

COPYRIGHT
© 2023 Wang, Zhong, Xiao, Dong, Long, Gan, Li and Liao. 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 radiomics model combined with clinical and radiographic features for discriminating peripheral small cell lung cancer from peripheral lung adenocarcinoma

Jingting Wang[1†], Feiyang Zhong[1,2†], Feng Xiao[1], Xinyang Dong[1], Yun Long[1], Tian Gan[1], Ting Li[1] and Meiyan Liao[1*]

[1]Department of Radiology, Zhongnan Hospital of Wuhan University, Wuhan, China, [2]Department of Radiology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China

Purpose: Exploring a non-invasive method to accurately differentiate peripheral small cell lung cancer (PSCLC) and peripheral lung adenocarcinoma (PADC) could improve clinical decision-making and prognosis.

Methods: This retrospective study reviewed the clinicopathological and imaging data of lung cancer patients between October 2017 and March 2022. A total of 240 patients were enrolled in this study, including 80 cases diagnosed with PSCLC and 160 with PADC. All patients were randomized in a seven-to-three ratio into the training and validation datasets (170 vs. 70, respectively). The least absolute shrinkage and selection operator regression was employed to generate radiomics features and univariate analysis, followed by multivariate logistic regression to select significant clinical and radiographic factors to generate four models: clinical, radiomics, clinical-radiographic, and clinical-radiographic-radiomics (comprehensive). The Delong test was to compare areas under the receiver operating characteristic curves (AUCs) in the models.

Results: Five clinical-radiographic features and twenty-three selected radiomics features differed significantly in the identification of PSCLC and PADC. The clinical, radiomics, clinical-radiographic and comprehensive models demonstrated AUCs of 0.8960, 0.8356, 0.9396, and 0.9671 in the validation set, with the comprehensive model having better discernment than the clinical model (P=0.036), the radiomics model (P=0.006) and the clinical–radiographic model (P=0.049).

Conclusions: The proposed model combining clinical data, radiographic characteristics and radiomics features could accurately distinguish PSCLC from PADC, thus providing a potential non-invasive method to help clinicians improve treatment decisions.

KEYWORDS

small cell lung cancer, lung adenocarcinoma, tomography, X-ray computed, radiomics, diagnostic model

1 Introduction

Lung cancer is the leading cause of cancer mortality worldwide in both men and women (1, 2) and is classified into two histological subtypes: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC), with adenocarcinoma (ADC) being one of the most common types of the latter (3). Depending on the location of primary lesions, lung cancers could be divided into central and peripheral types. The central type of lung cancer is mostly squamous-cell carcinoma and SCLC, while the peripheral type is mainly ADC (4–6). Peripheral SCLC (PSCLC) is relatively less common in clinical practice, accounting for 15-30% of all lung cancers (6, 7). PSCLC typically originates in the bronchial submucosa and infiltrates into the peribronchial connective tissue, resulting in bronchial stenosis being less likely to occur in its early stage. PSCLC is easy to overlook due to its non-obvious clinical symptoms and low diagnostic rate (7, 8). The tumors are already extensively metastasized by the time of detection, leading to most patients not benefiting from surgery or local radiotherapy (8). PSCLC is more aggressive and malignant than peripheral ADC (PADC) and they have significant differences in responses to therapy. Surgical resection is preferred for early-stage ADC. The high prevalence of surgery and the rapid development of targeted therapies, such as individualized treatment based on specific histological and molecular subtypes, have greatly improved the survival time of NSCLC, especially ADC (9, 10). However, extreme caution is required in SCLC surgical decision-making. Surgery is only recommended for certain patients with surgically resectable stage I to IIA SCLC (11). SCLC is highly sensitive to initial chemotherapy and radiotherapy. Postoperative adjuvant radiotherapy could significantly improve the 5-year survival rate (9). In addition, although radiotherapy is the most available treatment for patients with middle-late stage SCLC or NSCLC, there are significant differences between the two treatment regimens. Therefore, early, rapid, and accurate differential diagnosis of SCLC and NSCLC plays a crucial role in the treatment decisions and can improve the survival rate and prognosis of patients.

Pathological biopsy is still the gold standard for diagnosis of lung cancer. However, the NCCN Clinical Practice Guidelines in Oncology also suggest that patients with a strong clinical suspicion of stage I or II lung cancer (based on risk factors and radiologic appearance) do not require a biopsy before surgery. The biopsy adds time, costs, and procedural risk and may not be needed for treatment decisions (12). So, the initial determination of the pathological type of lung cancer is needed before treatment. Magnetic resonance imaging and computed tomography (CT), are widely used in clinical settings to investigate the anatomy and function of the body in both health and disease (13). CT images, providing qualitative morphological information, have been the preferred choice for lung cancer screening and diagnosis (14). Many studies have found some morphological CT features differed between NSCLC and SCLC (8, 14–16). Nonetheless, PSCLC, especially early-stage PSCLC, remains difficult to diagnose, some of the imaging features lack specificity, overlapping with peripheral NSCLC and other peripheral tumors. Furthermore, traditional imaging diagnostic accuracy was easily affected by the physician experience and subjective factors. Some research has shown that clinical data, particularly serum tumor markers, could be helpful predictors of the pathological type of lung cancer (17). Radiomics, a potentially non-invasive data-mining method, extracts high-throughput features from routinely acquired radiographic medical images (18). The comprehensive and detailed tumor characterization could be reflected by radiomics, which is cost-effective and non-invasive. Several studies related to radiomics showed very positive and promising results in the detection of lung cancer, the prediction of histology and subtypes, the prediction of prognosis and the assessment of treatment outcome (19). However, there are very few studies in the differential diagnosis between PSCLC and PADC.

This study retrospectively analyzed the clinical, radiographic, and radiomics characteristics between PSCLC and PADC patients. The aim of the study was to develop and validate a CT-based radiomics model to classify PSCLC and PADC and to investigate whether the addition of clinical and radiographic factors could improve the performance of the diagnostic model.

Abbreviations: PSCLC, peripheral small cell lung cancer; PADC, peripheral lung adenocarcinoma; NSCLC, non-small cell lung cancer; NSE, neuron-specific enolase; CEA, carcinoembryonic antigen; CA125, carbohydrate antigen 125; VOI, volume of interest; ICC, intra-class correlation coefficients; DSC, Dice similarity coefficient; mRMR, minimum redundancy–maximum relevance; LASSO, least absolute shrinkage and selection operator; DCA, decision curve analysis; OR, odds ratio; AUC, area under the curve; GGN, ground glass nodules; CT, Computed tomography; PACS, picture archive and communication system.

2 Materials and methods

2.1 Study population

We retrospectively identified participants with pathologic analysis–proven lung cancer from October 2017 to March 2022 and collected clinical and imaging data. The inclusion of this study was based on the following criteria: (1) thin-section (≤1.5 mm) CT scan performed within two weeks before needle biopsy or surgery; (2) tumors were located below the lung segment bronchus; and (3) without anti-cancer treatment before CT scan. Patients were excluded according to the following criteria: (1) incomplete image data or poor image quality; (2) ground-glass nodules (GGN); (3) absence of complete clinical data; and (4) with some other kind of primary malignancy. Due to this research's retrospective nature, a waiver of informed consent for this study was allowed, and the hospital ethics committee approved the study (2021057).

2.2 Clinical data

Seven clinical features, including gender, age, history of smoking, the preoperative clinical stage and serum tumor markers [neuron-specific enolase (NSE), carcinoembryonic antigen (CEA), and carbohydrate antigen 125 (CA125)] were

collected when reviewing the electronic medical record system. Preoperative clinical stage of lung cancer was based on the tumor/node/metastasis (TNM) classification of malignant tumors, 8th edition. The upper limit of the normal value of various serum tumor markers was used as the positivity threshold, with the following positivity criteria: CEA>5ng/mL, NSE>16.3U/ml, and CA125>35U/mL. The above tumor markers were taken as positive when their values were beyond the upper limit of the normal range. Data extraction was performed by two authors (JW and TG) independently and differences were resolved by a third investigator (FZ).

2.3 CT scanning

CT examinations were performed *via* two CT scanners (Somatom Definition 64 and Somatom Sensation 16; Siemens Healthcare, Forchheim, Germany). Technical parameters were: automatic tube current adjustment technology, 100-350 mAs; tube voltage, 120 kV; slice interval, 0 mm; reconstructed section thickness, 1 mm. The original images were reconstructed using a standard soft-tissue algorithm.

2.4 Visual assessment of CT signs

The analysis of the CT images was performed at the lung window setting (width, 1500 HU; level, -700 HU), and the following CT signs were analyzed: maximal lesion size, single/multifocal lesion, the specific lung lobes involved, air bronchogram, cavity, calcification, vessel convergence, margin, contour, lobulation, spiculation, pleural indentation, pleural effusion, peripheral emphysema, and enlarged mediastinal lymph node (see Figure 1 for definitions). Two radiologists (JW and FZ, who had been practicing diagnostic chest imaging for 2 and 5 years, respectively) reviewed CT images separately. Both radiologists were blinded to the clinicopathological information. In case of any discrepancies between the evaluations, a consensus was reached through discussion.

2.5 Clinical and radiographic risk factors

Significant clinical and radiographic features were screened out in univariate logistic regression and subsequently analyzed in multivariate logistic regression analysis in the training cohort. The significant variables in the multivariate logistic regression were identified as potential risk factors for the construction of the clinical-radiographic model. The participant cohort was randomly assigned to 2 sets—a training set (70%) and a validation set (30%). The former was used for feature filtering and model construction, while the latter was used for model validation only.

2.6 CT images segmentation, feature extraction, and model building

To reduce the variation caused by different CT scanners acquisition parameters, all images were resampled to the same

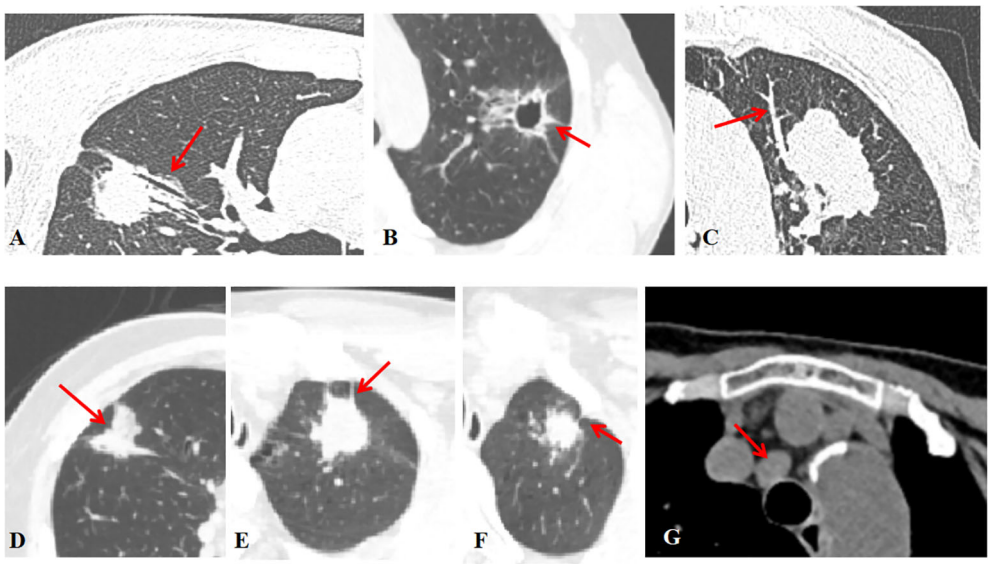


FIGURE 1
Radiographic features **(A)** The air bronchogram is fluid-filled or solid alveoli surrounded by air-filled bronchus. **(B)** The cavity is a gas-filled space >3 mm within a solid lung lesion, mass or nodule. **(C)** Vessel convergence sign was defined as an increase in internal vessels or clustering of vessels. **(D)** Lobulation was defined as a nodal margin showing an uneven lobar outline. **(E)** Spiculation was defined as a small spine-like protrusion from the nodule margin extending into the lung parenchyma without reaching the pleural surface. **(F)** Pleural indentation was defined as a tapered or linear extension of the lesion into the pleura, reflecting pulmonary fibrosis with retraction of the adjacent pleura. **(G)** Hilar and mediastinal lymph nodes were considered positive if the short axis exceeded 10 mm on chest CT images.

voxel size (1mm*1mm*1mm) with the linear interpolation method (20). The volumes of interest (VOIs) of lesion areas were segmented by a radiologist (a 2-year work experience) semi-automatically with ITK-SNAP software (version 3.8.itksnap.org). Care should be taken to avoid areas of the chest wall, lung parenchyma, great vessels, and bronchi around the mass and to outline the contour of the mass as completely as possible.

Intra-class correlation coefficients (ICC) were adopted to assess inter- and intra-observer agreement and to ensure reproducibility and stability of results. Feature reliability and reproducibility were tested using a cohort of twenty randomly selected samples. The consistency of the extracted features obtained between the first observer (JW) and a second observer (YL) (inter-observer reliability) was explored, as well as the consistency of the extracted features obtained by the exact first observer (JW) at different times (intra-observer reproducibility). The ICC was used to determine the consistency of the extracted features, and intra- and inter-observer variability was analyzed *via* calculating the Dice similarity coefficient (DSC). The DSC between two observers (JW and YL) (Table S2A) and between the two observations of the first observer (JW) (Table S2B) was 0.87 ± 0.05 and 0.86 ± 0.06, respectively, which was considered to achieve a good agreement (21). The remaining VOIs were delineated by the first radiologist (JW). Radiomics features were extracted from the pre-processed images with AK Software (Artificial Intelligence Kit; GE Healthcare) and used for subsequent modeling. Supplementary Material provided further details on feature extraction.

Features were ranked for importance by the minimum redundancy–maximum relevance (mRMR) algorithm. The top-ranked radiomics features were then picked out by the least absolute shrinkage and selection operator (LASSO) to obtain an optimal subset of radiomics features, which was employed to build the radiomics score (rad-score) to identify PSCLC and PADC. Each lesion's rad-score is calculated based on the linear combination of selected features, weighted by their coefficients. The study grouped the selected features into four models—the clinical model (only with clinical features), the radiomics model (radiomics features), the clinical-radiographic model (selected clinical and radiographic features), and the comprehensive model (radiomics, clinical and radiographic features). The individualized nomogram incorporated the clinical-radiographic characteristics with rad-score, allowing a simple, visual graphical representation of the risk of PSCLC (22).

2.7 Performance assessment

The discriminative efficacy from the above models was quantified based on the AUC values of the training and validation sets. Calibration curves were plotted with the Hosmer-Lemeshow test as a calibration indicator. The clinical utility of the nomogram was assessed by the decision curve analysis (DCA), which quantified the potential net effect of applying the model at different threshold probabilities.

2.8 Data analysis

All statistical analyses were carried out with SPSS 25.0 and R 4.1 statistical software. For the continuous variables, the Kolmogorov-Smirnov and Levene tests were first performed to verify the normality and cardinality of the samples. Variables conforming to the normal distribution and chi-square measures were expressed as Mean ± SD. Group comparisons were made using the t-test or Wilcoxon rank-sum test. Categorical variables were expressed as frequencies and percentages, and comparisons between groups were made using the χ^2 test. A two-sided $P<0.05$ indicated a statistically significant difference.

The "mRMR" and "glmnet" packages were used to conduct the mRMR algorithm and LASSO regression, respectively. The "pROC" package was used to plot the ROC curves and measure AUCs, which were compared using the Delong test. Nomogram was generated using the "rms" package. "irr" was used to calculate the ICC algorithm. The goodness of fit test was examined using "ResourceSelection" and "ggpubr" was used for data result visualization. The "CalibrationCurves" and "DecisionCurve" packages were used for the calibration curves and DCA analyses.

3 Results

3.1 Patients

885 cases of peripheral lung cancer were collected, including 423 cases of lung ADC, 123 cases of lung SCLC, and 39 cases of lung squamous-cell carcinoma. A total of 240 cases with 218 males and 22 females who met the inclusion/exclusion criteria were identified (mean age 64.14 years, range 34-83 years). Table 1 listed the baseline characteristics of the research population. In the ratio of 7:3 ratio, 170 patients were assigned to the training set (55 PSCLCs and 115 PADCs), while 70 patients were to the validation set (25 PSCLCs and 45 PADCs).

3.2 Clinical and radiographic factors in PSCLCs and PADCs

Univariate analysis of the clinical and radiographic factors showed that serum tumor marker levels, including NSE, CEA, and CA125, as well as the visual assessment of CT signs such as contour, pleural indentation, spiculation, air bronchogram, and lobulation, were significantly related to PSCLC ($P<0.05$), as shown in Table 2.

All statistically significant clinical and radiographic variables in the univariate analysis were further entered into the multivariate logistic regression analysis (Table 3). In the multivariate analysis, five variables, including NSE level, CEA level, spiculation, air bronchogram, and pleural indentation, were ultimately selected as predictors associated with PSCLC, and further included in the clinical-radiographic model.

TABLE 1 Comparison of clinical data in PSCLCs and PADCs.

Variables	SCLC Group (N=80) No. of patient (%)	ADC Group (N=160) No. of patient (%)	P value
Age (years)			0.961
Mean ± SD	64.18 ± 7.86	64.13 ± 7.35	
Gender			0.429
Male	71 (88.75)	147 (91.875)	
Female	9 (11.25)	13 (8.125)	
Smoking history			0.555
Yes	53 (66.25)	112 (70.00)	
No	27 (33.75)	48 (30.00)	
NSE			<0.001
Abnormal	58 (72.50)	20 (12.50)	
Normal	22 (27.50)	140 (87.50)	
CEA			0.014
Abnormal	30 (37.50)	87 (54.375)	
Normal	50 (62.50)	73 (45.673)	
CA125			<0.001
Abnormal	18 (22.50)	75 (46.875)	
Normal	62 (77.50)	85 (53.125)	
Clinical stage			0.521
Early (I, II)	39 (48.75)	71 (44.375)	
Late (III, IV)	41 (51.25)	89 (55.625)	

3.3 Feature extraction

Among the 1316 radiomics features extracted from CT images, the top 100 mRMR-ranked features were selected for training using the LASSO classifier (23). Finally, based on the retained 23 features, the rad-score formula was derived from the LASSO weighting coefficients, as shown in Figure 2, and the rad-score of each lesion was calculated. The rad-score calculation formula was described in the Supplementary Material. The rad-score of SCLC in the training set was lower than the rad-score of ADC (-0.044 ± 0.960 vs. 1.348 ± 0.862, $P < 0.001$), which was confirmed in the validation set (0.393 ± 0.508 vs. 1.348 ± 0.843, $P < 0.001$), as Figure 3 showed.

3.4 Establishment and validation of models

We constructed the radiomics model based on the rad-score and the clinical-radiographic model using five clinical and radiographic features, which were screened by the multivariate analysis, including NSE level, CEA level, spiculation, pleural indentation, and air bronchogram. The clinical model was constructed with the selected clinical features including NSE level and CEA level. The comprehensive model incorporated significant clinical and radiographic factors with the radiomics signature of CT images. The AUCs of the radiomics model, clinical model, clinical-radiographic model, comprehensive model was 0.8579 (95% CI, 0.7988-0.917), 0.851 (95% CI, 0.7956-0.9054), 0.9774 (95% CI, 0.9579-0.9969) and 0.9851 (95% CI, 0.9702-0.9999) in the training sets (Figure 4A) and 0.8356 (95% CI, 0.7419-0.9292), 0.8964 (95% CI, 0.8262-0.9667), 0.9396 (95% CI, 0.8867-0.9924) and 0.9671 (95% CI, 0.9251-0.9999) in the validation sets (Figure 4B), respectively. The comprehensive model outperformed the radiomics model (Z=-2.7472, P=0.006), the clinical model (Z=-2.0916, P=0.036) and the clinical-radiographic model (Z=-1.9639, P=0.049) according to the Delong test, which was statistically significant in the validation cohort.

As shown in Figure 5, the nomogram was developed by combining rad-score, NSE level, CEA level, spiculation, pleural indentation, and air bronchogram. Calibration curves depicted the calibration of the nomogram in terms of the agreement between the predicted risk and the actual probability of PSCLC. The model showed favorable calibration in the validation set, with the Hosmer-Lemeshow test yielding a non-significant P=0.3748. Figure 6 showed the DCA for the four models. The comprehensive model achieved more net clinical benefit than the other three models when differentiating between PSCLCs and PADCs in the training cohort (A) and the validation cohort (B).

TABLE 2 Results of univariate analysis for the classification of PSCLC and PADC in the training set.

Variables	SCLC Group (N=55) No. of patient (%)	ADC Group (N=115) No. of patient (%)	P value
Age (years)			
Mean ± SD	63.73 ± 8.25	64.10 ± 7.44	0.766
Gender			0.306
Male	48 (87.3)	106 (92.2)	
Female	7 (12.7)	9 (7.8)	
History of smoking			0.858
Yes	38 (69.1)	81 (70.4)	
No	17 (30.9)	34 (29.6)	
NSE			<0.001
Abnormal	38 (69.09)	16 (13.91)	
Normal	17 (30.91)	99 (86.09)	
CEA			0.004
Abnormal	18 (32.73)	65 (56.52)	
Normal	37 (67.27)	50 (43.48)	
CA125			<0.001
Abnormal	11 (20.0)	54 (46.96)	
Normal	44 (80.0)	61 (53.04)	
Maximal lesion size (mm)			0.675
Mean ± SD	36.40 ± 16.23	35.37 ± 14.19	
Lesion location			0.115
Right upper lobe	16 (29.1)	49 (42.6)	
Right middle lobe	1 (1.8)	6 (5.2)	
Right lower lobe	17 (30.9)	21 (18.2)	
Left upper lobe	9 (16.4)	24 (20.9)	
Left lower lobe	12 (21.8)	15 (13.0)	
Contour			0.003
Round	38 (69.1)	51 (44.3)	
Irregular	17 (30.9)	64 (55.7)	
Lobulation			0.034
Present	21 (38.2)	26 (22.6)	
Absent	34 (61.8)	89 (77.4)	
Spiculation			<0.001
Present	6 (10.9)	100 (87.0)	
Absent	49 (89.1)	15 (13.0)	
Margin			0.115
Clear	47 (85.5)	86 (74.8)	
Unclear	8 (14.5)	29 (25.2)	
Calcification			0.681

(Continued)

TABLE 2 Continued

Variables	SCLC Group (N=55) No. of patient (%)	ADC Group (N=115) No. of patient (%)	P value
Present	15 (27.3)	28 (24.3)	
Absent	40 (72.7)	87 (75.7)	
Cavity			0.085
Present	15 (27.3)	47 (40.9)	
Absent	40 (72.7)	68 (59.1)	
Air bronchogram			0.001
Present	2 (3.6)	29 (25.2)	
Absent	53 (96.4)	86 (74.8)	
Vessel convergence sign			0.07
Present	21 (38.2)	61 (53.0)	
Absent	34 (61.8)	54 (47.0)	
Pleural indentation			<0.001
Present	12 (21.8)	91 (79.1)	
Absent	43 (78.2)	24 (20.9)	
Peripheral emphysema			0.179
Present	31 (56.4)	77 (67.0)	
Absent	24 (43.6)	38 (33.0)	
Enlarged mediastinal lymph node			0.071
Present	34 (61.8)	54 (47.0)	
Absent	21 (38.2)	61 (53.0)	
Single/multifocal lesion			0.716
Multiple	12 (21.8)	28 (24.3)	
Single	43 (78.2)	87 (75.7)	
Pleural effusion			0.786
Present	11 (20.0)	21 (18.3)	
Absent	44 (80.0)	94 (81.7)	
Clinical stage			0.490
I	11 (20.00)	31 (26.96)	
II	16 (29.09)	22 (19.13)	
III	12 (21.82)	26 (22.61)	
IV	16 (29.09)	36 (31.30)	

4 Discussion

In the present study, we developed and validated a radiomics-based model nomogram that combined clinical data with subjective radiographic characteristics derived from CT images to help preoperatively distinguish PSCLC from PADC. The nomogram showed satisfactory diagnostic performance compared to the radiomics and clinical-radiographic models, with an AUC of 0.9851 for the training set and 0.9671 for the validation set. The comprehensive model incorporated serum tumor markers, CT signs, and the radiomics signature, further improving the diagnostic value.

Serum tumor biomarkers such as CEA, NSE, and CYFRA 21-1 have been revealed to be associated with several clinical events in lung cancer, including diagnosis of tumor subtypes, prognosis, and detection of tumor recurrence, all of which are closely associated with tumor burden (24). As shown in the results of the present study, serum tumor biomarkers, including CEA and NSE, were significant predictors in patients with PSCLC. In patients with an abnormal NSE level, a higher NSE level would indicate an increased

TABLE 3 Results of multivariate logistic regression analysis for classifying PSCLC and PADC in the training set.

Variables	P value	OR (95% CI)
NSE	<0.001	90.372 (8.841-923.732)
CEA	0.019	0.167 (0.037-0.745)
CA125	0.364	0.456 (0.084-2.486)
Contour	0.272	2.462 (0.494-12.272)
Lobulation	0.058	5.524 (0.944-32.343)
Spiculation	<0.001	0.010 (0.001-0.105)
Air bronchogram	0.022	0.016 (0.001-0.556)
Pleural indentation	0.028	0.187 (0.042-0.834)

likelihood of SCLC diagnosis, while the CEA level was higher in ADC than that in SCLC, which is consistent with previous studies (25). This study not only included serum tumor markers to explore the predictive markers but also considered the impact of other clinical factors such as smoking history, gender, and age on SCLC. SCLC is strongly associated with cigarette smoking (26, 27), but it was not found in the present study. This may be because traditional Chinese women smoke less. The inclusion of a small number of Chinese women in this study may lead to the possibility of sample selection bias.

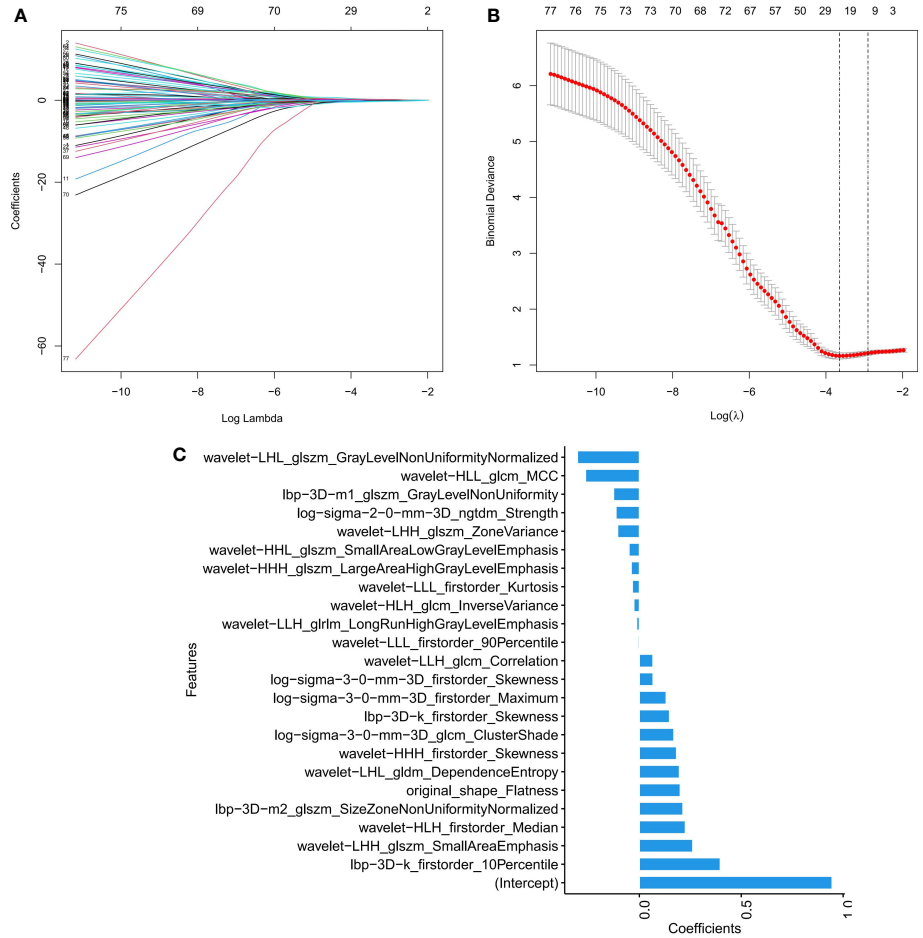


FIGURE 2
Feature selection by LASSO binary logistic regression model. **(A)** Selecting characterization *via* LASSO regression and 10-fold cross-validation method. **(B)** Coefficient curves based on radiomics features with non-zero coefficients are determined by λ. **(C)** Screening the absolute values of 23 radiomics features and their corresponding coefficients.

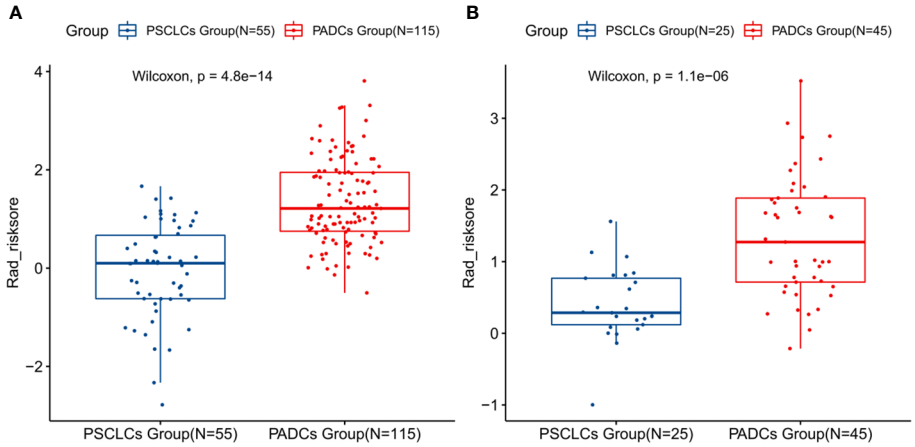


FIGURE 3
Boxplots presenting the statistical differences between PSCLCs and PADCs in the training set **(A)** and validation set **(B)**.

Histopathological biopsy is the gold standard for accurate subclassification of lung cancer. Peripheral lung cancer is mainly obtained by CT-guided percutaneous puncture lung biopsy, which is a complex invasive procedure, and the personal experience is a major determinant of procedural success (28). Sputum cytology is a convenient and non-invasive way to screen and diagnose lung cancer (29). However, this approach often gives negative results in the case of peripherally-placed cancers, although histopathology revealed positive diagnoses. CT screening for lung cancer has relatively high sensitivity and is the imaging modality of choice for diagnosing lung cancer (30, 31). In this study, the models only with clinical features achieved AUCs of 0.851 and 0.896 in the training and validation sets, respectively, which were lower than the comprehensive and clinical-radiographic models, showing the benefits of CT screening. Clinical symptoms of nodular lesions in the lung may appear later than imaging symptoms. Several recent studies have attempted to correlate different CT features of peripheral lung cancer with specific pathological subtypes (32, 33). Ground glass nodules (GGN) on chest CT could be a key imaging biomarker for early primary ADC in diagnosis, which had been observed at our institution (23, 34). Accordingly, mixed and pure GGN were excluded from this study. Both PSCLC and PADC can present as nodular lesions, and some CT signs may also be helpful for differential diagnosis (35). Many previous articles have studied the CT features of PSCLC. PSCLC tumors are often described as round or round-like nodules with lobulation, little spiculated signs, and internal necrosis. Due to its dense tumor cell arrangement with few fibrous tissues (36), air bronchogram were observed less frequently in PSCLC. In addition, pleural indentation was also uncommon in SCLC owing to the weak influence on the surrounding structure. Compared with that in PADC, spiculation was less common in PSCLC; these findings were consistent with previous results (37). SCLC typically presents as a large hilar mass and bulky mediastinal lymphadenopathy. It is uncommon for

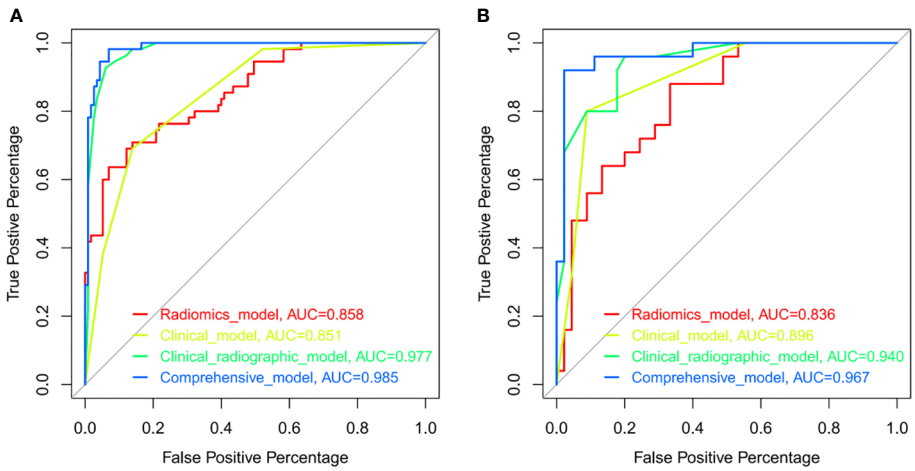


FIGURE 4
ROCs of the radiomics model (red), clinical model (yellow), clinical-radiographic model (green), and comprehensive model (blue) in the training **(A)** and validation **(B)** sets. ROC indicated the differential diagnostic efficacy of each model. The more accurate the diagnostic model, the more the ROC shifts toward the Y-axis, with the AUC approaching a value of 1.

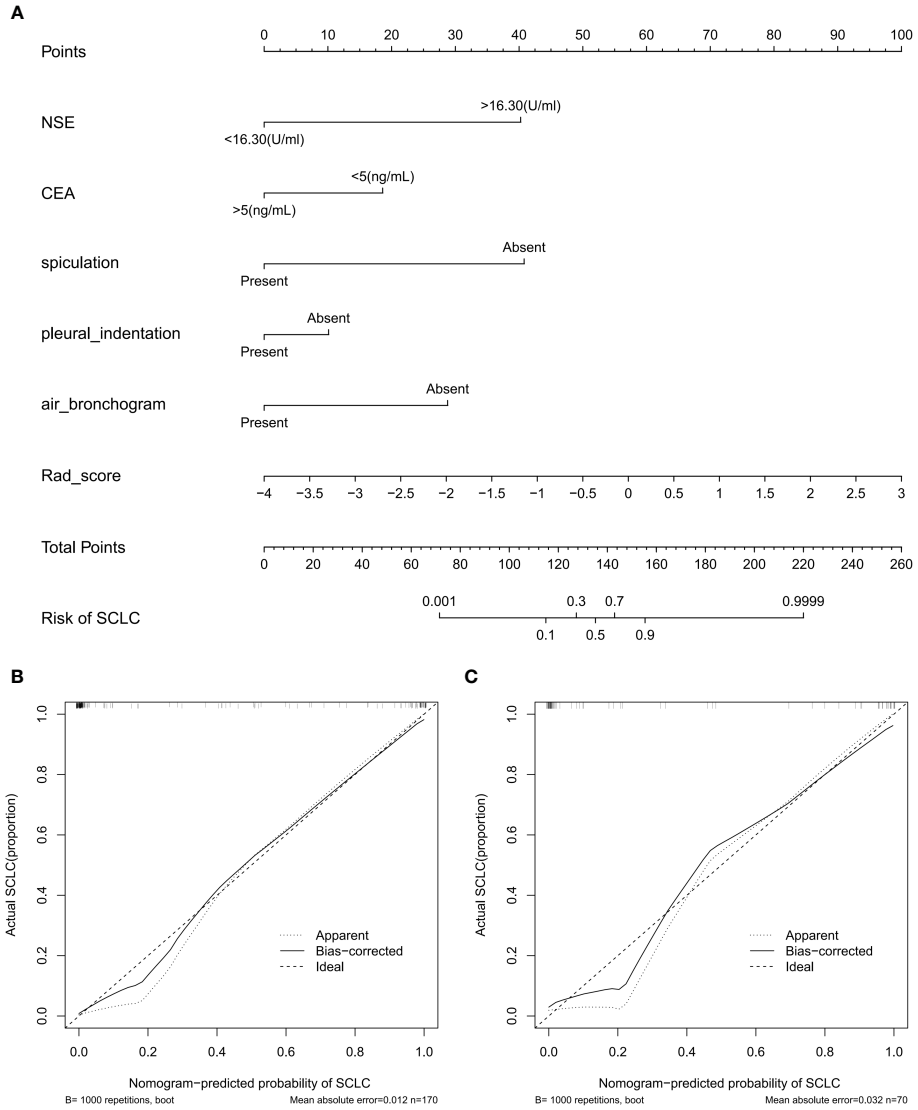


FIGURE 5
Radiomics-based nomogram **(A)** was developed in the training set, including the rad-score, NSE level, CEA level, spiculation, pleural indentation, and air bronchogram. Calibration curve of the nomogram in the training **(B)** and validation **(C)** datasets.

patients to present with a solitary peripheral nodule without central adenopathy (11). However, lymph node enlargement had no significant difference in the present study, which may be due to the selection bias caused by the relatively small sample.

Radiomics performs high-throughput extraction of quantitative image features, converts images into mineable data features, and subsequently analyses this data, providing valuable information for clinical decision-making (18, 21, 38). Junior et al. used radiomics-based CT features as well as machine learning models to differentiate various types of lung cancer, with AUCs yielding 0.97 and 0.71 at testing and validation, respectively, indicating the great potential of the method for differential diagnosis of lung cancer subtypes (39). In exploring the clinical value of radiomics features for histological subtypes of tumors such as SCLC and NSCLC, Liu et al. found that the radiomics-based nomogram combined with clinical factors outperformed the simple application of the radiomics signature (30), confirming that the combination of radiomics and clinical data could improve the predictive performance. The nomogram developed by Liu had good predictive performance (training cohort, AUC: 0.985; validation cohort, AUC: 0.966).

The nomogram model we constructed had higher differential diagnostic efficacy (training cohort, AUC: 0.9851; validation cohort, AUC: 0.9671). One possible explanation could be that this research focused on peripheral solid nodules, making it relatively easy to outline the tumor on CT images, thus avoiding the hilum and main bronchi. Besides, the combination of radiomics, clinical factors, and CT signs may be another reason. This study constructed four models based on radiomics features, clinical data, clinical-radiographic factors, and combined radiomics-clinical-radiographic features. All four had good diagnostic ability in distinguishing peripheral SCLC from ADC, and the

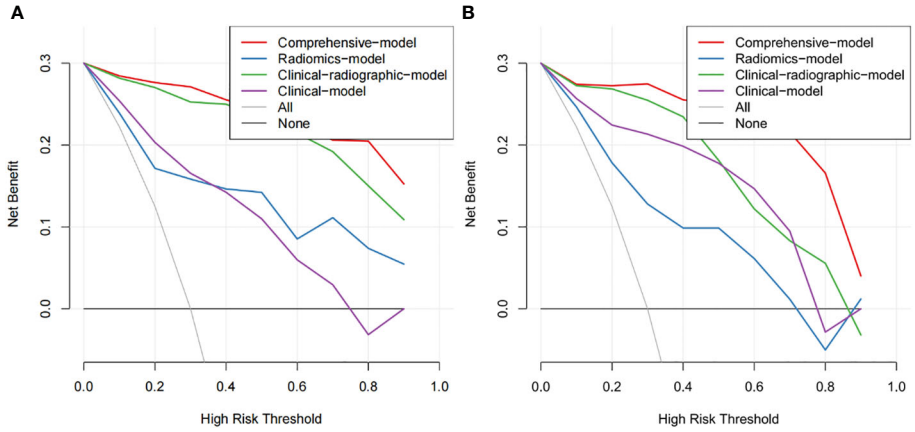


FIGURE 6
The DCA of the developed nomogram in training cohort **(A)** and validation cohort **(B)**. The y-axis and x-axis represent the net benefit and threshold probability, respectively. At the bottom, the black line "None" represented the assumption that none developed PSCLC, while the gray line "All" represented the assumption that all patients suffered PSCLC. The red line represented the net return of using the nomogram to predict PSCLC.

comprehensive model also had the best predictive efficacy compared with the other three models by the Delong test ($P<0.05$) in the validation set. This study developed an individual nomogram model to facilitate clinical decision-making by incorporating clinical data, radiographic characteristics, and radiomics features. The nomogram graphically, visually, and intuitively displayed the extent to which each factor contributes to the diagnosis of PSCLC, providing a non-invasive and individualized analysis method. Mohammadhadi Khorram et al. stated that radiomics was useful for predicting the early-stage NSCLC recurrence, progression, and recurrence free survival (40). Besides, previous study had elucidated the prognostic value of radiomics in SCLC patients scheduled for first-line chemotherapy (41). However, to the best of our knowledge, no previous study has investigated radiomics approaches in the prognosis of PSCLC. This will be an important issue to be explored in our future studies. SCLC typically shows intense uptake on positron emission tomography (PET)/CT scans, reflecting its high metabolic activity (42). Metabolic parameters on pretreatment and posttreatment PET/CT scans can be used as a predictive marker of clinical outcome of SCLC (42). PET images may contain added valuable information, which would be explored in our future work.

One limitation of this study was the sample size of the retrospective single-center study, whose relatively small number of patients may have led to bias in the validation set. All high-resolution CT images in this study were obtained through the hospital's picture archive and communication system (PACS). However, before 2017, the hospital did not save CT thin layer images. Therefore, we only included data between October 2017 to March 2022. Secondly, the paired controls enrolled in the present study were only from peripheral lung adenocarcinoma, while other pathological types of peripheral lung cancer were not included in this study due to their small number. We may need larger cohorts from other centers for further external cross-validation in later studies. Finally, this study only extracted the radiomics features from plain CT images. It may be worthwhile to expand the analysis to include contrast-enhanced CT and PET/CT images as part of the study in future research.

In conclusion, the differentiated clinical-radiographic model constructed in this study performed well in distinguishing PSCLC from PADC. On the basis of this, the nomogram developed based on clinical-radiographic factors combined with CT radiomics further improved the accuracy of diagnosing PSCLC and PADC, which was objective, non-invasive, and reproducible.

Data availability statement

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

Ethics statement

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

JW and FZ: Conception of the project, design of the study, analysis of the data, and writing of the article. YL, TG, and FX: Collection and processing of data. TL, XD and JW: Collection of data. ML: Critical revision of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

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

Conflict of interest

The authors declared 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.2023.1157891/full#supplementary-material

References

1. Schabath MB, Cote ML. Cancer progress and priorities: Lung cancer. *Cancer Epidemiol Biomarkers Prev* (2019) 28(10):1563–79. doi: 10.1158/1055-9965.EPI-19-0221

2. Wang S, Zimmermann S, Parikh K, Mansfield AS, Adjei AA. Current diagnosis and management of small-cell lung cancer. *Mayo Clin Proc* (2019) 94:1599–622. doi: 10.1016/j.mayocp.2019.01.034

3. Planchard D, Popat S, Kerr K, Novello S, Smit EF, Faivre-Finn C, et al. Metastatic non-small cell lung cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol* (2018) 29:iv192–237. doi: 10.1093/annonc/mdy275

4. Liu H, Jing B, Han W, Long Z, Mo X, Li H. A comparative texture analysis based on NECT and CECT images to differentiate lung adenocarcinoma from squamous cell carcinoma. *J Med Syst* (2019) 43(3):1–8. doi: 10.1007/s10916-019-1175-y

5. Wang J, Xie X, Shi J, He W, Chen Q, Chen L, et al. Denoising autoencoder, a deep learning algorithm, aids the identification of a novel molecular signature of lung adenocarcinoma. *Genomics Proteomics Bioinf* (2020) 18(4):468–80. doi: 10.1016/j.gpb.2019.02.003

6. Valadares BN, Stephano MA. Small cell lung cancer: An overview of the targets. *Braz J Pharm Sci* (2022) 58:1–12. doi: 10.1590/s2175-97902022e19114

7. Morabito A, Rolfo C. Small cell lung cancer: A new era is beginning? *Cancers* (2021) 13:2646. doi: 10.3390/cancers13112646

8. Ren Y, Cao Y, Hu W, Wei X, Shen X. Diagnostic accuracy of computed tomography imaging for the detection of differences between peripheral small cell lung cancer and peripheral non-small cell lung cancer. *Int J Clin Oncol* (2017) 22 (5):865–71. doi: 10.1007/s10147-017-1131-0

9. Li C, Lu H. Adenosquamous carcinoma of the lung. *Onco. Targets* (2018) 11:4829–35. doi: 10.2147/OTT.S164574

10. Xu X, Sui X, Zhong W, Xu Y, Wang Z, Jiang J, et al. Clinical utility of quantitative dual-energy CT iodine maps and CT morphological features in distinguishing small-cell from non-small-cell lung cancer. *Clin Radiol* (2019) 74:268–77. doi: 10.1016/j.crad.2018.10.012

11. Ganti AKP, Loo BW, Bassetti M, Blakely C, Chiang A, D'Amico TA, et al. Small cell lung cancer, version 2.2022, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw* (2021) 19(12):1441–64. doi: 10.6004/jnccn.2021.0058

12. Roy V, Ugalde PA, Bourdages-Pageau E, Lacasse Y, Labbé C. Transthoracic needle biopsy versus surgical diagnosis for solid pulmonary nodules. *J Thorac Dis* (2022) 14(7):2472–80. doi: 10.21037/jtd-22-35

13. Qiao Z, Xia C, Shen S, Corwin FD, Liu M, Guan R, et al. Suppression of the PI3K pathway in vivo reduces cystitis-induced bladder hypertrophy and restores bladder capacity examined by magnetic resonance imaging. *PLoS One* (2014) 9(12):e114536. doi: 10.1371/journal.pone.0114536

14. Chamberlin J, Kocher MR, Waltz J, Snoddy M, Stringer NFC, Stephenson J, et al. Automated detection of lung nodules and coronary artery calcium using artificial intelligence on low-dose CT scans for lung cancer screening: Accuracy and prognostic value. *BMC Med* (2021) 19:1–14. doi: 10.1186/s12916-021-01928-3

15. Lee D, Rho JY, Kang S, Yoo KJ, Choi HJ. CT findings of small cell lung carcinoma: Can recognizable features be found? *Med (Baltimore)* (2016) 95:e5426. doi: 10.1097/MD.0000000000005426

16. Thomas A, Pattanayak P, Szabo E, Pinsky P. Characteristics and outcomes of small cell lung cancer detected by CT screening. *Chest* (2018) 154(6):1284–90. doi: 10.1016/j.chest.2018.07.029

17. Liu L, Teng J, Zhang L, Cong P, Yao Y, Sun G, et al. The combination of the tumor markers suggests the histological diagnosis of lung cancer. *BioMed Res Int* (2017) 2017:1–9. doi: 10.1155/2017/2013989

18. Wu S, Zheng J, Li Y, Yu H, Shi S, Xie W, et al. A radiomics nomogram for the preoperative prediction of lymph node metastasis in bladder cancer. *Clin Cancer Res* (2017) 23(22):6904–11. doi: 10.1158/1078-0432.CCR-17-1510

19. Chen BT, Chen Z, Ye N, Mambetsariev I, Fricke J, Daniel E, et al. Differentiating peripherally-located small cell lung cancer from non-small cell lung cancer using a CT radiomic approach. *Front Oncol* (2020) 10:593. doi: 10.3389/fonc.2020.00593

20. Xie Z, Sun H, Wang J, Xu H, Li S, Zhao C, et al. A novel CT-based radiomics in the distinction of severity of coronavirus disease 2019 (COVID-19) pneumonia. *BMC Infect Dis* (2021) 21(1):608. doi: 10.1186/s12879-021-06331-0

21. Liu Q, Jiang P, Jiang Y, Ge H, Li S, Jin H, et al. Development and validation of an institutional nomogram for aiding aneurysm rupture risk stratification. *Sci Rep* (2021) 11(1):1–10. doi: 10.1038/s41598-021-93286-6

22. Zhao M, Li S, Xu Y, Su X, Jiang H. Developing a scoring model to predict the risk of injurious falls in elderly patients: A retrospective case-control study in multicenter acute hospitals. *Clin Interv Aging* (2020) 15:1767–78. doi: 10.2147/CIA.S258171

23. Zhong F, Liu Z, An W, Wang B, Zhang H, Liu Y, et al. Radiomics study for discriminating second primary lung cancers from pulmonary metastases in pulmonary solid lesions. *Front Oncol* (2022) 11:5471. doi: 10.3389/fonc.2021.801213

24. Vos D, Rao S, Pierce JD, Smith DA, Tirumani SH, Yoest JM, et al. The past, present, and future (Liquid biopsy) of serum tumor markers in lung cancer: A primer for the radiologist. *J Comput Assist Tomogr* (2021) 45(6):950–8. doi: 10.1097/RCT.0000000000001204

25. Lu L, Zha Z, Zhang P, Li D, Liu G. NSE, positively regulated by LINC00657-miR-93-5p axis, promotes small cell lung cancer (SCLC) invasion and epithelial-mesenchymal transition (EMT) process. *Int J Med Sci* (2021) 18(16):3768–79. doi: 10.7150/ijms.58415

26. Zugazagoitia J, Paz-Ares L. Extensive-stage small-cell lung cancer: First-line and second-line treatment options. *J Clin Oncol* (2022) 40(6):671–80. doi: 10.1200/JCO.21.01881

27. Gao Q, Tang S, Chen H, Chen H, Li X, Jiang Y, et al. Intratumoral injection of anlotinib hydrogel enhances antitumor effects and reduces toxicity in mouse model of lung cancer. *Drug Delivery* (2020) 27(1):1524–34. doi: 10.1080/10717544.2020.1837292

28. Nath A, Neyaz Z, Hashim Z, Agrawal V, Richa M. Role of percutaneous computed tomography-guided lung biopsy in non-resolving consolidation and identification of clinical and high-resolution computed tomography characteristics predicting outcome. *J Clin Imaging Sci* (2019) 9:48. doi: 10.25259/JCIS_126_2019

29. Liu S, Liu S, Zhang C, Yu H, Liu X, Hu Y, et al. Exploratory study of a CT radiomics model for the classification of small cell lung cancer and non-small-Cell lung cancer. *Front Oncol* (2020) 10:1268. doi: 10.3389/fonc.2020.01268

30. Jazieh K, Khorrami M, Saad A, Gad M, Gupta A, Patil P, et al. Novel imaging biomarkers predict outcomes in stage III unresectable non-small cell lung cancer treated with chemoradiation and durvalumab. *J Immunother Cancer* (2022) 10(3): e003778. doi: 10.1136/jitc-2021-003778

31. Gharraf HS, Mehana SM, ElNagar MA. Role of CT in differentiation between subtypes of lung cancer; is it possible? *Egypt J Bronchol* (2020) 14:1–7. doi: 10.1186/s43168-020-00027-w

32. Inafuku K, Yokose T, Ito H, Eriguchi D, Samejima J, Nagashima T, et al. Two cases of lung neuroendocrine carcinoma with carcinoid morphology. *Diagn Pathol* (2019) 14:1–6. doi: 10.1186/s13000-019-0886-1

33. Xie X, Li X, Tang W, Xie P, Tan X. Primary tumor location in lung cancer: The evaluation and administration. *Chin Med J (Engl)* (2021) 135(2):127–36. doi: 10.1097/CM9.0000000000001802

34. Zhu D, Chen J, Wang Y, Liao M. CT features of peripheral small cell lung cancer smaller than 3.0 cm in diameter. *Med J Wuhan Univ* (2018) 39(3):443–6. doi: 10.14188/j.1671-8852.2017.0924

35. Linning E, Lu L, Li L, Yang H, Schwartz LH, et al. Radiomics for classifying histological subtypes of lung cancer based on multiphasic contrast-enhanced computed tomography. *J Comput Assist Tomogr* (2019) 43(2):300–6. doi: 10.1097/RCT.0000000000000836

36. Zhang X, Lv F, Fu B, Li W, Lin R, Chu Z. Clinical and computed tomography characteristics for early diagnosis of peripheral small-cell lung cancer. *Cancer Manag Res* (2022) 14:589–601. doi: 10.2147/CMAR.S351561

37. Elkilany A, Fehrenbach U, Auer TA, Müller T, Schöning W, Hamm B, et al. A radiomics-based model to classify the etiology of liver cirrhosis using gadoxetic acid-enhanced MRI. *Sci Rep* (2021) 11:1–13. doi: 10.1038/s41598-021-90257-9

38. Ferreira Junior JR, Koenigkam-Santos M, Cipriano FEG, Fabro AT, Azevedo-Marques PM. Radiomics-based features for pattern recognition of lung cancer histopathology and metastases. *Comput Methods Programs BioMed* (2018) 159:23–30. doi: 10.1016/j.cmpb.2018.02.015

39. Polsterl S, Gupta P, Wang L, Conjeti S, Katouzian A, Navab N, et al. Heterogeneous ensembles for predicting survival of metastatic, castrate-resistant prostate cancer patients. *F1000Res* (2016) 5:2676. doi: 10.12688/f1000research.8231.1

40. Wei H, Yang F, Liu Z, Sun S, Xu F, Liu P, et al. Application of computed tomography-based radiomics signature analysis in the prediction of the response of small cell lung cancer patients to first-line chemotherapy. *Exp Ther Med* (2019) 17:3621–9. doi: 10.3892/etm.2019.7357

41. Park H, Tseng SC, Sholl LM, Hatabu H, Awad MM, Nishino M. Molecular characterization and therapeutic approaches to small cell lung cancer: Imaging implications. *Radiology* (2022) 305(3):512–25. doi: 10.1148/radiol.220585

42. Heo JW, Yeo CD, Park CK, Kim SK, Kim JS, Kim JW, et al. Smoking is associated with pneumonia development in lung cancer patients. *BMC Pulm Med* (2020) 20:1–8. doi: 10.1186/s12890-020-1160-8

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 17 April 2023
DOI 10.3389/fonc.2023.1119807

Check for updates

OPEN ACCESS

EDITED BY
Qingqing Zhu,
The First Affiliated Hospital of Soochow University, China

REVIEWED BY
Jing Yang,
Chongqing Technology and Business University, China
Vera Luiza Capelozzi,
University of São Paulo, Brazil

*CORRESPONDENCE
Wentao Li
Liwentootoo@163.com
Ansheng Wang
wangansheng@bbmc.edu.cn

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

SPECIALTY SECTION
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

RECEIVED 09 December 2022
ACCEPTED 23 March 2023
PUBLISHED 17 April 2023

CITATION
Zhou C, Zhao R, Zhao R, Wang A and Li W (2023) Preoperative levels of folate receptor-positive circulating tumor cells in different subtypes of early-stage lung adenocarcinoma: Predictive value for determining extent of surgical resection. *Front. Oncol.* 13:1119807. doi: 10.3389/fonc.2023.1119807

COPYRIGHT
© 2023 Zhou, Zhao, Zhao, Wang and Li. 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.

Preoperative levels of folate receptor-positive circulating tumor cells in different subtypes of early-stage lung adenocarcinoma: Predictive value for determining extent of surgical resection

Chao Zhou[1†], Ran Zhao[2†], Ruiying Zhao[3], Ansheng Wang[4*] and Wentao Li[1*]

[1]Department of Thoracic Surgery, Shanghai Chest Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China, [2]Department of Thoracic Surgery, ShuYang Hospital of Traditional Chinese Medicine, Suqian, China, [3]Department of Pathology, Shanghai Chest Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China, [4]Department of Thoracic Surgery, The First Affiliated Hospital of Bengbu Medical College, Bengbu, China

Background: The objective was to measure the correlations of preoperative levels of folate receptor-positive circulating tumor cells (FR^+CTCs) with clinical characteristics and histologic subtype in early-stage lung adenocarcinoma, and to determine the predictive value of FR^+CTC level in preoperative determination of the extent of surgical resection.

Patients and methods: In this retrospective, single-institution, observational study, preoperative FR^+CTC levels were measured *via* ligand-targeted enzyme-linked polymerization in patients with early-stage lung adenocarcinoma. Receiver operating characteristic (ROC) analysis was used to identify the optimal cutoff value of FR^+CTC level for prediction of various clinical characteristics and histologic subtypes.

Results: No significant difference in FR^+CTC level was observed among patients with adenocarcinoma *in situ* (AIS), minimally invasive adenocarcinoma (MIA), and invasive adenocarcinoma (IAC) (P = 0.813). Within the non-mucinous adenocarcinoma group, no difference was observed among patients with tumors whose predominant growth patterns were lepidic, acinar, papillary, micropapillary, solid, and complex gland (P = 0.053). However, significant differences in FR^+CTC level were observed between patients with and without the micropapillary subtype [11.21 (8.22-13.61) *vs*. 9.85 (7.43-12.63), P = 0.017], between those with and without the solid subtype [12.16 (8.27-14.90) *vs*. 9.87 (7.50-12.49), P = 0.022], and between those with any of the advanced subtypes (micropapillary, solid, or complex glands) vs. none of these [10.48 (7.83-13.67) *vs*. 9.76 (7.42-12.42), P = 0.032]. FR^+CTC level was also correlated with degree of differentiation of lung adenocarcinoma (P = 0.033), presence of visceral pleural

invasion (VPI) of lung carcinoma (P = 0.003), and lymph node metastasis of lung carcinoma (P = 0.035).

Conclusion: FR^+CTC level is of potential predictive value in determining the presence of aggressive histologic patterns (micropapillary, solid, and advanced subtypes), degree of differentiation, and occurrence of VPI and lymph node metastasis in IAC. Measurement of FR^+CTC level combined with intraoperative frozen sections may represent a more effective method of guiding resection strategy in cases of cT1N0M0 IAC with high-risk factors.

KEYWORDS

pathological subtype, intraoperative frozen sections, diagnosis, sensitivity, specificity, lung adenocarcinoma

1 Introduction

Lung cancer is the second-most common cancer worldwide and the leading cause of cancer morbidity and mortality in men (1). Non-small cell lung cancer (NSCLC) accounts for approximately 85% of all lung cancers, and adenocarcinoma is the most common histologic type of NSCLC (2). Lung adenocarcinoma can be further subdivided into preinvasive lesions [atypical adenomatous hyperplasia (AAH) and adenocarcinoma *in situ* (AIS)], minimally invasive adenocarcinoma (MIA), and invasive adenocarcinoma (IAC), according to the degree of invasion (3). The different subtypes of lung adenocarcinoma strongly impact the choice of surgical methods and the prognosis of the patient. Previous studies have reported that in stage I lung adenocarcinoma, patients with the AIS and MIA subtypes have a 5-year disease-free survival (DFS) rate close to 100%, while this rate is only 38%–86% for those with IAC (4, 5).

Lobectomy remains the standard surgical treatment for lung adenocarcinoma. However, non-IAC patients also have the option of limited surgical resection (6). Several prospective, multicenter clinical studies have investigated the effect of lobectomy vs. sublobectomy on the survival rates of AIS and MIA patients. They have found that sublobectomy is the optimal choice, producing similar survival rates along with better quality of life (7, 8). If we could precisely determine the degree of invasiveness of lung adenocarcinoma intraoperatively, the less damaging sublobectomy would be feasible in more cases, based on guaranteed curative resection. However, Yeh et al. have reported that interobserver agreement in terms of discriminating between AIS, MIA, and invasive adenocarcinomas using frozen sections (FSs) is not satisfactory (κ = 0.378, fair agreement) (9). As such, the identification of AIS and MIA by examination of intraoperative FSs is not recommended by the prevailing guidelines.

Non-mucinous invasive lung adenocarcinoma is mainly divided into lepidic, acinar, papillary, micropapillary, and solid predominant patterns; this pattern is increasingly considered to be a powerful predictor of prognosis in lung adenocarcinoma (10). IAC, with a predominant lepidic pattern, has been found to reflect indolent behavior in terms of the biological behavior of the tumor, which is associated with a favorable prognosis (11). In contrast, studies have shown that prognosis is significantly worse in the case of sublobectomy than in the case of lobectomy among IAC patients with micropapillary composition ≥5% (11–13). Furthermore, the presence of micropapillary or solid components has been found to be an independent predictor for unfavorable recurrence-free survival (RFS) rates in cT1N0M0 lung adenocarcinoma patients undergoing lobectomy (14). Preoperatively and intraoperatively, accurate histologic subtyping is helpful to enable thoracic surgeons to decide on the extent of surgical resection. The accuracy of FS for prediction of the predominant histologic subtype is not satisfactory (κ = 0.565, moderate agreement), and the sensitivity of FS analysis, in terms of the presence or absence of micropapillary and solid patterns, is poor (micropapillary: 37%, κ = 0.321, fair agreement; solid: 69%, κ = 0.67, substantial agreement) (9).

To date, studies on intraoperative FS have demonstrated low consistency in the determination of pathological subtype (9, 15–17). Therefore, it is desirable to identify preoperative biomarkers that can distinguish the histologic patterns of early-stage lung adenocarcinoma.

Circulating tumor cell (CTC) level has proven utility in the early diagnosis of cancer, in monitoring recurrence and metastasis, in determining the prognosis of surgical and systemic interventions, and in the selection of postoperative adjuvant therapy (18). Folate receptor-positive CTCs (FR^+CTCs) are a type of circulating rare cell with high expression of FRs on the cell surface; levels of these cells can be reliably quantified using commercially available kits (19). Lung cancer patients of all stages have been found to exhibit significantly higher FR^+CTC levels than patients with benign lung disease and healthy subjects. Studies have also shown that FR^+CTC level can be used to diagnose lung cancer with high sensitivity (79.6%) and specificity (88.2%) (19–21). Furthermore, FR^+CTC level is significantly higher in patients with MIA and IAC than in AIS patients (22). However, the correlation between FR^+CTCs and the histologic patterns of cTis-T1N0M0 lung adenocarcinoma remains to be explored, and the question of whether FR^+CTC

level can help thoracic surgeons to determine the extent of surgical resection required for early-stage lung adenocarcinoma is worthy of investigation.

2 Patients and methods

2.1 Study design

This retrospective, single-institution study was conducted at Shanghai Chest Hospital from April 2021 to August 2022. A total of 1,835 patients with suspected lung cancer who were recommended for radical surgery were included in the study. The inclusion criteria were as follows: 1) age between 18 and 80 years old; 2) clinical stage Tis-T1N0M0 lung adenocarcinoma according to the eighth edition of the Tumor Node Metastasis (TNM) Classification for Lung Cancer (23); and 3) FR^+CTC test conducted prior to the operation. The exclusion criteria were as follows: 1) multiple primary lung cancer (MPLC); 2) a history of lung cancer or other malignancies; 3) prior antitumor therapy (including neoadjuvant therapy); 4) non-lung adenocarcinoma; and 5) incomplete clinical information.

For all patients, pathological evaluation of the diseased tissue was conducted by a professional pathologist according to the IASLC/ATS/ERS classification of lung adenocarcinoma (24). The Ethics Committee of the Shanghai Chest Hospital approved the study.

2.2 FR^+CTC analysis

A sample of 3 ml of peripheral blood was collected from each participant using an ethylenediaminetetraacetic acid (EDTA) anticoagulant vacuum tube, stored at 4°C, and processed within 24 h. FR^+CTC detection was performed using a Folate Receptor-positive Cell Detection Kit (GenoSaber, Shanghai, China). Two steps were followed to enrich CTCs: erythrocytes were first lysed, and leukocytes were subsequently eliminated with reference to the manufacturer's specified protocol. The labeled FR^+CTCs were enumerated *via* quantitative polymerase chain reaction (PCR) using the proprietary ligand-targeted PCR method (21). A series of standards containing oligonucleotides (10^{-14} to 10^{-9} to 2 to 2×10^5 M, corresponding to FU/3 ml blood) were used for quantification of FR^+CTCs. The number of folate-receptor units (FU) per 3 ml of peripheral blood was calculated from the standard curve and was used to determine the FR^+CTC level in each sample.

2.3 Postoperative pathological examination

Pathological sections were fixed by embedding in paraffin and stained with hematoxylin–eosin (HE). All components with a proportion greater than 5% were recorded in 5% increments. The pathological images were independently interpreted by two pathologists with more than 5 years of experience. If there was any dispute, the diagnosis was made by reading the images together under a multihead microscope. Nodules were reclassified according to the 2021 WHO histological classification of lung tumors and divided into AIS, MIA, and IAC (Supplementary Figures 1A–C: A and C, ×200; B, ×50). Cases of IAC were further subdivided into five types according to the component with the highest proportion: lepidic, acinar, papillary, micropapillary, or solid predominant (Supplementary Figures 1D–H, ×200).

2.4 Data collection

Data were collected from clinical records using the standard case report form (CRF). Data including demographic characteristics, clinical symptoms, types of surgery undergone, genetic testing, and CT imaging information.

2.5 Statistical analysis

Descriptive statistics are expressed in the form of medians (interquartile range) for continuous variables and in the form of counts (percentages) for categorical variables. Continuous variables were compared using the Mann–Whitney test for comparisons between two groups or the Kruskal–Wallis test for comparisons among three groups. Categorical variables were compared using the chi-square test. Receiver operating characteristic (ROC) analysis was used to determine the optimal threshold of FR^+CTC level, and the associated specificity and sensitivity, for classifying cases according to growth pattern, degree of differentiation, lymph node metastasis, and visceral pleural invasion (VPI). All statistical analyses were performed using STATA version 16 SE (Stata Corporation, TX, USA). All *P*-values were calculated based on two-sided testing. A *P*-value <0.05 was considered to represent statistical significance.

3 Results

3.1 Characteristics of the patients

A total of 1,835 patients admitted for surgical operation during the study period were reviewed in this study. Of these, 625 patients were excluded: 348 had MPLC, 71 had a benign tumor, 96 had non-tumorous benign lesions, 26 had AAH, 81 had non-lung adenocarcinoma, and 3 had lung adenocarcinoma with intrapulmonary metastasis. Thus, a total of 1,210 patients were included in the final analysis (Figure 1). Among these patients, 438 were men (36.17%), and 519 (42.86%) were older than 60 years of age. Additionally, of the 1,210 patients included, 301 (24.9%) were AIS cases, 284 (23.5%) were MIA, and 625 (51.6%) were IAC. The demographic and clinical characteristics of the participants are summarized in Supplementary Table 1.

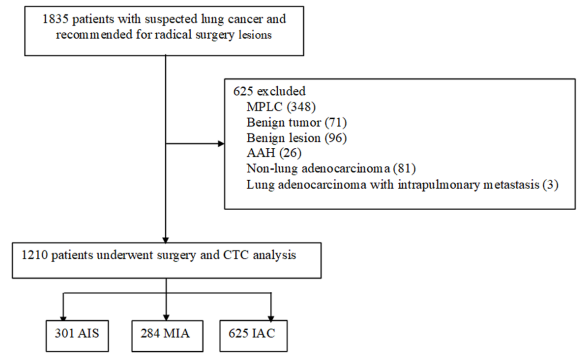


FIGURE 1
Flowchart of patient screening.

3.2 Correlation of FR$^+$CTC level with pathological type

The median (IQR) FR$^+$CTC level in the AIS, MIA, and IAC groups was 9.92 (7.45, 12.74) FU/3 ml, 9.86 (7.36, 13.4) FU/3 ml, and 9.93 (7.59, 12.9) FU/3 ml, respectively. No significant differences in FR$^+$CTC level were observed among the AIS, MIA, and IAC groups (P = 0.813) (Figure 2; Supplementary Table 1).

3.3 Correlation of FR$^+$CTC level with growth pattern subtype in IAC patients

We subsequently performed subgroup analyses of FR$^+$CTC level within the IAC group. There was no significant difference between patients with mucinous adenocarcinoma and those with non-mucinous adenocarcinoma [9.93 (7.59-12.77) *vs.* 10.20 (7.18-20.91), P = 0.622)] (Figure 3A; Table 1). In the non-mucinous adenocarcinoma group, no differences were observed between patients whose tumors exhibited different predominant growth patterns [lepidic: 9.74 (7.28-11.90) *vs.* acinar: 10.11 (7.65-12.93) *vs.* papillary: 9.54 (6.97-11.58) *vs.* micropapillary: 11.49 (8.16-13.11) *vs.* solid: 12.42 (8.87-16.37) *vs.* complex gland: 10.22 (7.17-11.77)), P = 0.053] (Figure 3B; Table 1). However, significant differences were observed between groups with and without a micropapillary component [11.21 (8.22-13.61) *vs.* 9.85 (7.43-12.63), P = 0.017] (Figure 3C; Table 1); between groups with and without a solid component [12.16 (8.27-14.90) *vs.* 9.87 (7.50-12.49), P = 0.022] (Figure 3D; Table 1); between groups with and without a complex gland component [10.74 (7.72-13.57) *vs.* 9.87 (7.51-12.69), P = 0.405] (Figure 3E; Table 1); and between the group with a micropapillary, solid, or complex gland component and the group with none of these components [10.48 (7.83-13.67) *vs.* 9.76 (7.42-12.42), P = 0.032] (Figure 3F; Table 1).

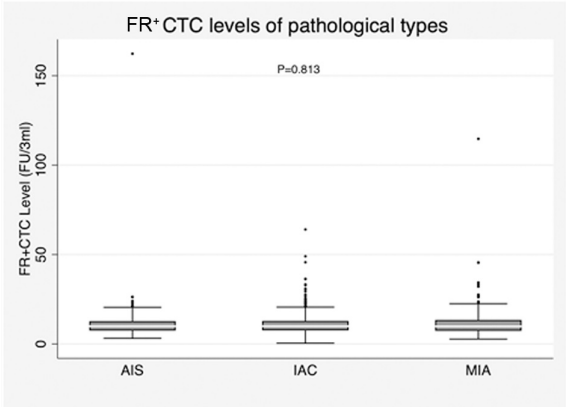


FIGURE 2
Correlation of FR$^+$CTC level with pathological type across all included patients.

3.4 Correlation of FR$^+$CTC level with degree of differentiation, VPI, lymph node metastasis, and spread through air spaces in IAC patients

FR$^+$CTC level was correlated with degree of differentiation in IAC (P = 0.033). Specifically, the median FR$^+$CTC level was higher in patients with poorly differentiated tumors than in those with moderately and well-differentiated tumors [11.17 (8.01-14.21) *vs.* 9.81 (7.49-12.49), P = 0.010; 11.17 (8.01-14.21) *vs.* 9.75 (6.90-12.01), P = 0.013] (Figure 4A; Table 1). FR$^+$CTC level was also correlated with VPI of lung carcinoma [11.53 (8.93-16.56) *vs.* 9.87 (7.43-12.54), P = 0.003] (Figure 4B; Table 1). Additionally, patients with lymph node metastasis had higher FR$^+$CTC levels than those without [11.21 (9.72-13.53) *vs.* 9.86 (7.49-12.69), P = 0.035] (Figure 4C; Table 1). However, FR$^+$CTC level was not correlated with spread through air spaces (STAS) [11.21 (8.22-12.84) *vs.* 9.89 (7.49-12.89), P = 0.224] (Figure 4D; Table 1).

3.5 Correlation of FR$^+$CTC level with genetic mutations

We explored the association between FR$^+$CTC level and commonly reported driver gene mutations in lung cancer. The results suggested that there was no correlation between FR$^+$CTC level and mutations in *EGFR* [9.87 (7.51-12.54) *vs.* 9.82 (7.87-12.69), P = 0.775], *ALK* [9.61 (7.16-14.85) *vs.* 10.00 (7.75-12.71), P = 0.963], *KRAS* [10.21 (7.86-14.01) *vs.* 10.21 (7.86-14.01), P = 0.598], *BRAF* [9.01 (5.67-22.74) *vs.* 9.87 (7.44-13.08), P = 0.867], or *ROS1* [10.96 (9.97-11.51) *vs.* 9.87 (7.49-13.05), P = 0.494] (Supplementary Table 2).

3.6 Results of the ROC analysis

To further explore the potential clinical value of FR$^+$CTC level in distinguishing growth patterns in cases of non-mucinous IAC, an ROC analysis was conducted with two patient groups: all patients with micropapillary component >5% were defined as the case cohort, and patients with micropapillary component ≤5% were

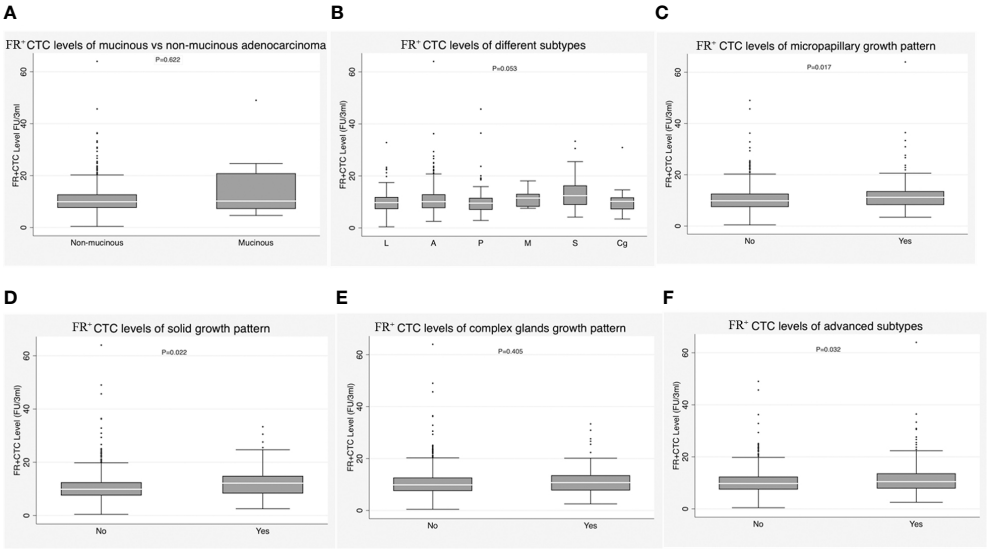


FIGURE 3
(A) Correlation of FR$^+$CTC level with mucinous adenocarcinoma vs. non-mucinous adenocarcinoma in the invasive adenocarcinoma (IAC) group. **(B)** Correlation of FR$^+$CTC level with lepidic subtype, acinar subtype, papillary subtype, micropapillary subtypes, solid subtype, and complex glands in the IAC group. **(C)** Correlation of FR$^+$CTC level with presence vs. absence of micropapillary components in the IAC group. **(D)** Correlation of FR$^+$CTC level with presence vs. absence of solid components in the IAC group. **(E)** Correlation of FR$^+$CTC level with presence vs. absence of complex glands in the IAC group. **(F)** Correlation of FR$^+$CTC level with presence vs. absence of advanced subtypes in the IAC group.

TABLE 1 Baseline characteristics of 625 patients with invasive adenocarcinoma.

Characteristics	*N*	FR$^+$CTCs (FU/3 ml; median, IQR)	*P*-value
Tumor size (cm)			0.524
0 < T1a ≤ 1	91	10.04 (7.82-12.89)	
1 < T1b ≤ 2	356	10.00 (7.71-12.92)	
2 < T1c ≤ 3	178	9.72 (7.23-12.71)	
Invasive adenocarcinoma			0.622
Mucinous	23	10.20 (7.18-20.91)	
Non-mucinous	602	9.93 (7.59-12.77)	
Predominant histologic subtype of non-mucinous invasive adenocarcinoma			0.053
Lepidic	100	9.74 (7.28-11.90)	
Acinar	354	10.11 (7.65-12.93)	
Papillary	81	9.54 (6.97-11.58)	
Micropapillary	14	11.49 (8.16-13.11)	
Solid	29	12.42 (8.87-16.37)	
Complex glands	24	10.22 (7.17-11.77)	
Presence/absence of histologic pattern			
Lepidic			0.160
Yes	395	9.87 (7.5-12.45)	
No	230	10.35 (7.73-13.57)	
Acinar			0.673

(Continued)

TABLE 1 Continued

Characteristics	*N*	FR^+CTCs (FU/3 ml; median, IQR)	*P*-value
Yes	533	9.93 (7.59-12.69)	
No	92	9.94 (7.54-13.05)	
Papillary			0.078
Yes	271	9.73 (7.31-12.49)	
No	354	10.20 (7.8-13.24)	
Micropapillary			0.017
Yes	103	11.21 (8.22-13.61)	
No	522	9.85 (7.43-12.63)	
Solid			0.022
Yes	60	12.16 (8.27-14.90)	
No	565	9.87 (7.50-12.49)	
Complex glands			0.405
Yes	90	10.74 (7.72-13.57)	
No	535	9.87 (7.51-12.69)	
Advanced subtypes (micropapillary, solid, or complex glands)			0.032
Yes	179	10.48 (7.83-13.67)	
No	446	9.76 (7.42-12.42)	
Degree of differentiation			0.033
Well-differentiated	79	9.75 (6.90-12.01)	
Moderately differentiated	401	9.81 (7.49-12.49)	
Poorly differentiated	145	11.17 (8.01-14.21)	
VPI			0.003
Yes	52	11.53 (8.93-16.56)	
No	573	9.87 (7.43-12.54)	
Lymph node metastasis			0.035
Yes	35	11.21 (9.72-13.53)	
No	518	9.86 (7.49-12.69)	
STAS			0.224
Yes	58	11.21 (8.22-12.84)	
No	567	9.89 (7.49-12.89)	

VPI, visceral pleural invasion; STAS, spread through air spaces.

designated as a control cohort. The Youden index was used to determine the optimal cutoff value. The ROC analysis yielded an area under the curve (AUC) of 0.574, with a 95% confidence interval (CI) of 0.514-0.634, at the optimal cutoff value of an FR^+CTC level of 10.40 FU/3 ml. The sensitivity of this test among lung cancer patients was 57.28%, with a specificity of 57.47%, in predicting the presence of micropapillary component >5%, with $P = 0.009$ (Figure 5A; Table 2).

We further explored the use of FR^+CTC level as a predictor of the presence of solid components, advanced subtypes, degree of differentiation, presence of lymph node metastasis, and VPI. The sensitivity and specificity of FR^+CTC level for preoperative prediction of the presence of solid components were 51.67% and 72.74%, respectively, and the AUC in this test was 0.590 (95% CI, 0.506-0.673; $P = 0.011$) (Figure 5B; Table 2). The sensitivity and specificity of FR^+CTC level for prediction of advanced subtypes were 52.51% and 58.07%, respectively, and the AUC in this test was 0.555 (95% CI, 0.504-0.606; $P = 0.016$) (Figure 5C; Table 2). The sensitivity and specificity of FR^+CTC level for distinguishing between well-differentiated and poorly differentiated cases were 39.31% and 77.22%, and the AUC in this test was 0.587 (95% CI, 0.510-0.664; $P = 0.015$) (Figure 5D; Table 2). Additionally, the

10.3389/fonc.2023.1119807

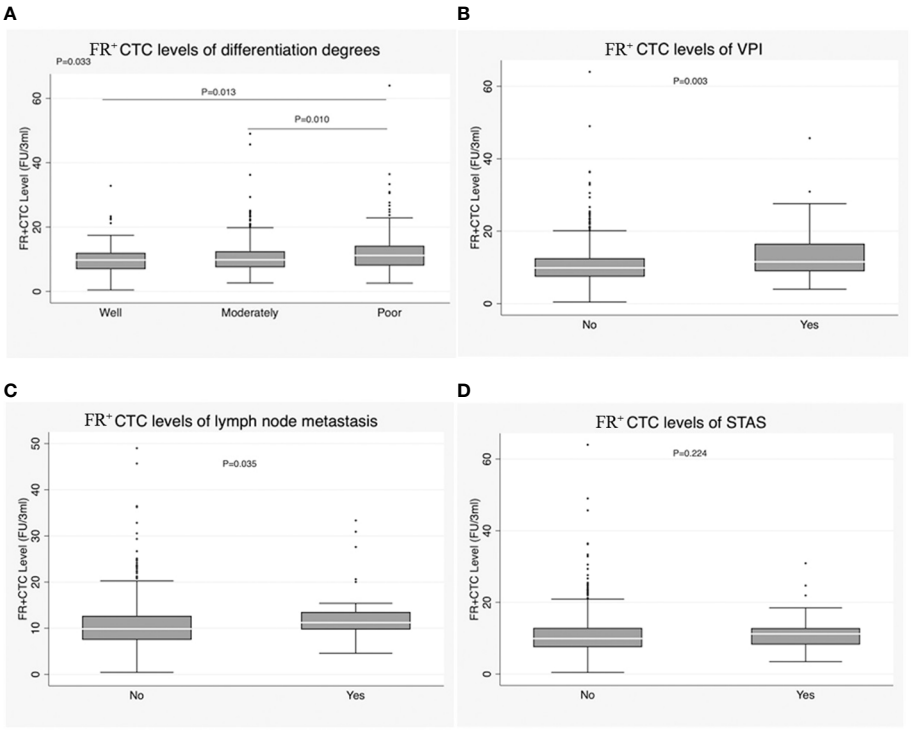


FIGURE 4
(A) Correlation of FR$^+$CTC level with degree of differentiation in the IAC group. **(B)** Correlation of FR$^+$CTC level with visceral pleural invasion (VPI) in the IAC group. **(C)** Correlation of FR$^+$CTC level with lymph node metastasis in the IAC group. **(D)** Correlation of FR$^+$CTC level with STAS in the IAC group.

sensitivity and specificity of FR$^+$CTC level for distinguishing between moderately differentiated and poorly differentiated cases were 51.72% and 62.09%, respectively, and the AUC in this test was 0.565 (95% CI, 0.508-0.622; P = 0.010) (Figure 5E; Table 2). The sensitivity and specificity of FR$^+$CTC level for prediction of VPI were 61.54% and 58.81%, respectively, and the AUC in this test was 0.623 (95% CI, 0.543-0.702; P = 0.002) (Figure 5F; Table 2). Finally, the sensitivity and specificity of FR$^+$CTC level for prediction of lymph node metastasis were 77.14% and 47.88%, respectively, and the AUC in this test was 0.606 (95% CI, 0.523-0.690; P = 0.018) (Figure 5G; Table 2).

4 Discussion

Achieving consistency between intraoperative FS and postoperative final pathology in patients with early-stage lung adenocarcinomas remains a challenging problem. Marchevsky

TABLE 2 Optimal cutoff, sensitivity, specificity, and AUC of ROC curves.

Characteristics	Cutoff	Sensitivity	Specificity	AUC	Lower CI	Upper CI	*P*-value
Micropapillary	10.40	57.28%	57.47%	0.574	0.514	0.634	0.009
Solid	12.13	51.67%	72.74%	0.590	0.506	0.673	0.011
Advanced subtypes	10.40	52.51%	58.07%	0.555	0.504	0.606	0.016
Well differentiated *vs.* poorly differentiated	12.13	39.31%	77.22%	0.587	0.510	0.664	0.015
Moderately differentiated *vs.* poorly differentiated	10.93	51.72%	62.09%	0.565	0.508	0.622	0.010
VPI	10.69	61.54%	58.81%	0.623	0.543	0.702	0.002
Lymph node metastasis	9.72	77.14%	47.88%	0.606	0.523	0.690	0.018

AUC, area under the curve; ROC, receiver operating characteristic; VPI, visceral pleural invasion.

10.3389/fonc.2023.1119807

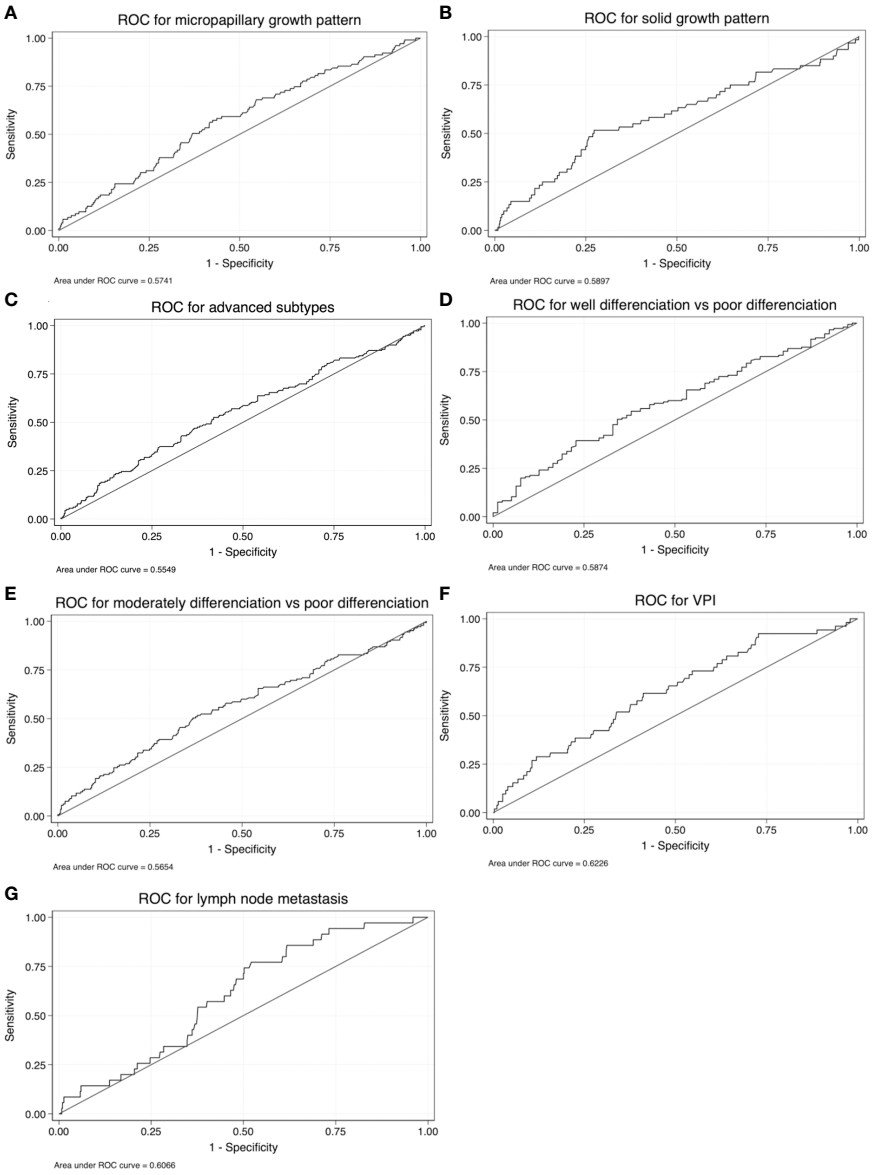


FIGURE 5
Receiver operating characteristic (ROC) curves for FR$^+$CTC level as a test for **(A)** micropapillary components, **(B)** solid components, **(C)** advanced subtypes, **(D)** well-differentiated *vs*. poorly differentiated tumors, **(E)** moderately differentiated *vs*. poorly differentiated tumors, **(F)** VPI, and **(G)** lymph node metastasis.

et al. report that intraoperative FS has lower diagnostic accuracy for AIS and MIA than for IAC (51% *vs*. 97%) (25). He et al. have also found that intraoperative FS pathology offers poor differentiation between AAH and AIS and between AIS and MIA. The possible reasons include the deformation of tissues and cells caused by freezing during the operation and the limitations of FS sampling leading to an incomplete representation of the entire lesion (26). In our study, consistency between intraoperative FS and postoperative pathology was only 62.3%, lower than observed in a previous study (64%) (9). The FR$^+$CTC test was approved by the NMPA to assist in diagnosing pulmonary lesions and has showed high sensitivity and specificity. In our study, FR$^+$CTC level could not be used to distinguish between AIS, MIA, and IAC, which is in line with the results of the study by Ding et al. However, in their study, FR$^+$CTCs could be used to distinguish between AIS and MIA (22); this inconsistency of findings may be related to the fact that the participants included in our study were all in the early stage and the sample size was small. Neither FR$^+$CTC level nor intraoperative FS can accurately distinguish the pathological subtypes of lung adenocarcinoma. However, larger samples in prospective clinical studies are needed to provide validation in the future.

Some studies have shown that pathological subtype, degree of differentiation, and clinical stage are closely correlated with prognosis in lung cancer (27, 28). However, few studies have explored the ability of FR$^+$CTC level to distinguish growth pattern subtypes, as this one has done, and we propose that this measure can be used a reference in determining the extent of surgical resection required in patients with early invasive

adenocarcinoma. Adenocarcinoma patients with tumors exhibiting different growth patterns would ideally receive different forms of treatment, and they have different prognoses. In a study by Tsuta et al., 904 cases of surgically resected adenocarcinomas were investigated; it was found that lepidic, acinar, papillary, micropapillary, and solid predominant adenocarcinomas were associated with different prognoses, and their 5-year OS rates were 93%, 67%, 74%, 62%, and 58%, respectively (29). A recent study consisting of 697 patients with pN0M0/papillar/papillar–acinar predominant lung adenocarcinomas with diameter ≤3 cm who underwent curative resection found that the presence of micropapillary and solid patterns as minor components had a negative impact on prognosis. For MP/S^+ Lep^-, MP/S^+ Lep^+, MP/S^- Lep^-, and MP/S^- Lep^+ types, the 5-year RFS rates were 81.9%, 94.0%, 94.4%, and 98.7%, respectively ($P < 0.001$), and the 5-year OS rates were 87.7%, 96.6%, 94.4%, and 98.4%, respectively ($P < 0.001$). Moreover, a multivariate analysis suggested that the MP/S^+ Lep^- subtype was an independent poor prognostic factor for both RFS and OS (30). The sensitivity and specificity of intraoperative FS for diagnosis of solid primary adenocarcinoma were 79% and 94%, respectively. For the diagnosis of micropapillary primary lung adenocarcinoma, specificity reached 99%, but sensitivity was only 21%. When only the presence or absence of micropapillary subtypes was considered, the diagnostic sensitivity of FS was still only 37% (9).

In our study, we found that there was no difference in level of FR^+CTCs between the acinar, lepidic, papillary, micropapillary, and solid subtypes in the non-mucinous IAC group. Levels were higher in the acinar subtype compared to the lepidic subtype; in the solid subtype compared to the acinar and papillary subtypes; and in the papillary subtype compared to the lepidic subtype. Furthermore, when only the presence or absence of micropapillary or solid subtypes was considered, there was a notable difference in FR^+CTC level between the groups ($P = 0.017$ and $P = 0.022$, respectively). The presence of micropapillary or solid patterns was found to significantly increase the risk of nodal upstaging in the multivariable analysis ($P < 0.001$ and $P = 0.001$, respectively), and these patterns were independently associated with poor RFS ($P = 0.041$ and $P < 0.001$) among patients with cT1N0M0 lung adenocarcinoma (14). Micropapillary or solid components have been proven to be poor prognostic factors. Therefore, these patients should be cautious when considering sublobectomy. According to the guidelines, the standard surgery for IAC is still lobectomy; however, in our study, the proportions of patients who underwent lobectomy in high-risk groups with micropapillary and solid components were only 64.1% and 68.3%, respectively (Supplementary Table 3). Therefore, it is necessary to identify a biomarker that can aid preoperative pathological judgment and improve the accuracy of selection of surgical method. According to the results of our study, FR^+CTCs may be a promising auxiliary biomarker for determination of whether solid and micropapillary components are present. However, the diagnostic efficacy is not very good, and the sensitivity and specificity of FR^+CTC level as a test for the presence of micropapillary components and solid components are relatively low. The AUCs are well below 0.7 (0.574 and 0.590). This means that FR^+CTC level can only function as a reference in distinguishing whether micropapillary or solid components are present, and does not have a strong ability to make this distinction. Prospective clinical studies with larger samples are needed to provide validation of this finding. The proportion of tumors of the complex gland subtype is small among patients with early-stage adenocarcinoma; thus, there is insufficient evidence to draw a conclusion on its role in prognosis after sublobar resection at present.

FR^+CTC level was correlated with lymph node metastasis in IAC, with the level being higher in patients with lymph node metastasis than in patients without metastasis ($P = 0.035$), which was consistent with a previous study (27). The sensitivity and specificity of FR^+CTC level for diagnosis of lymph node metastasis were 77.14% and 47.88%, respectively, and the AUC was 0.606. Thus, at a cutoff value of 9.72 FU/3 ml, this test cannot reliably distinguish whether lymph node metastasis has occurred. Evidence has shown that invasion may occur very early in the tumor process, and CTCs are released into circulation in the early phase of cancer. Lymph node infiltration is related to poor prognosis (28). Many researchers have emphasized the importance of detection of CTCs (31–33). Our results also showed that FR^+CTC level was correlated with degree of differentiation in IAC. Until now, few studies have focused on the prognostic value of biomarkers in lung tumors with different degrees of differentiation, so this finding is of limited significance because a more accurate method of discrimination has already been developed.

VPI is a high-risk factor that affects prognosis in lung adenocarcinoma. Nevertheless, identification of VPI is still reliant on elastic fiber staining, which is time-consuming and difficult to perform in intraoperative FS. The accuracy, sensitivity, and specificity of intraoperative FS for diagnosis of VPI have been found to be 75%, 47.4%, and 97.3%, respectively (34). Furthermore, our results indicated that FR^+CTC level was correlated with VPI ($P = 0.003$). However, when 10.25 FU/3 ml was used as the cutoff value, the sensitivity and specificity of FR^+CTC level for the diagnosis of VPI were 61.54% and 58.81%, respectively. The associated AUC was 0.623, which means that FR^+CTC is almost capable of distinguishing whether VPI has occurred with sufficient accuracy.

Two studies have shown that the sensitivity of STAS detection in intraoperative FS is only 44%-54%. At the same time, specificity is as high as 80%-91%, and multifactorial logistic regression analysis has found that artifacts are the only relevant factor in the misdiagnosis of STAS by FS (35, 36). However, in the present study, there was no significant difference in FR^+CTC level between cases with and without STAS ($P = 0.224$). The possible reasons are a relatively small sample size and the inclusion of patients from a population at a specific stage. Ways to improve the sensitivity of STAS detection preoperatively or intraoperatively are worthy of future research.

In our study, we found that the proportions of patients with advanced subtypes (micropapillary or solid or complex glands), lymph node metastasis, poor differentiation, and VPI who underwent lobectomy were 64.8%, 82.9%, 63.5%, and 75.0%, respectively (Supplementary Table 3); the proportion did not reach 100% in any of these subgroups. In the case of stage I IAC

patients who underwent sublobectomy because high-risk pathological factors (advanced subtypes, VIP, STAS, lymph node metastasis, and so on) could not be identified intraoperatively, postoperative adjuvant therapy or supplementary lobectomy was suggested if the patients agreed (16, 37). FR^+CTCs can provide helpful information in distinguishing aggressive growth patterns of IAC, lymph node metastasis, and VPI, which could help to identify the optimal surgical strategy and avoid infliction of secondary damage on patients. In the meantime, FR^+CTC level can be used as a biomarker and possible prognostic factor in early-stage lung adenocarcinoma.

This study had some limitations. The study was preliminary and retrospective in nature, and patient outcomes were not collected and analyzed. In subsequent studies, we will collect long-term prognostic data and analyze the value of FR^+CTC level in predicting outcomes in early-stage patients. Furthermore, prospective randomized studies are required to truly discern the value of FR^+CTC level in the preoperative differentiation of histological subtypes and as a tool for use in the selection of surgical method.

5 Conclusion

To conclude, FR^+CTC level may be used as an additional biomarker in identifying patients with micropapillary components, solid components, advanced subtypes, poorly differentiated tumors, VPI, and lymph node metastasis. Lobectomy should be selected as the preferred surgical option for these high-risk patients.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by the Shanghai Chest Hospital. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

Conception and design: CZ and WL. Administrative support: AW and WL. Provision of study materials or patients: CZ and WL. Collection and assembly of data: CZ and RZ. Data analysis and interpretation: CZ and RYZ. Manuscript writing and final approval of the manuscript: all authors.

Funding

This study was funded by the Ministry of Finance and National Health Commission of the People's Republic of China, National Key R&D Program of China (Grant number: 2021YFC2500900/2021YFC2500904).

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.2023.1119807/full#supplementary-material

References

1. Sung HJ, Ferlay RL, Siegel M, Laversanne I, Soerjomataram AJ, 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. Boukansa SZ, Benbrahim S, Gamrani S, Bardai L, Bouguenouch AM, et al. Correlation of epidermal growth factor receptor mutation with major histologic subtype of lung adenocarcinoma according to IASLC/ATS/ERS classification. *Cancer Control* (2022) 29:10732748221084930. doi: 10.1177/10732748221084930

3. Travis WDE, Brambilla AG, Nicholson Y, Yatabe JHM, Austin MBB, et al. The 2015 world health organization classification of lung tumors: Impact of genetic, clinical and radiologic advances since the 2004 classification. *J Thorac Oncol* (2015) 10 (9):1243–60. doi: 10.1097/JTO.0000000000000630

4. Hattori AS, Hirayama T, Matsunaga T, Hayashi K, Takamochi SO, et al. Distinct clinicopathologic characteristics and prognosis based on the presence of ground glass opacity component in clinical stage IA lung adenocarcinoma. *J Thorac Oncol* (2019) 14 (2):265–75. doi: 10.1016/j.jtho.2018.09.026

5. Pedersen JHZ, Saghir MM, Wille LH, Thomsen BGS, Ashraf H. Ground-glass opacity lung nodules in the era of lung cancer CT screening: Radiology, pathology, and clinical management. *Oncol (Williston Park)* (2016) 30(3):266–74.

6. Fan LM, Fang Z, Li W, Tu S, Wang WC. Radiomics signature: a biomarker for the preoperative discrimination of lung invasive adenocarcinoma manifesting as a ground-glass nodule. *Eur Radiol* (2019) 29(2):889–97. doi: 10.1007/s00330-018-5530-z

7. Asamura H. Role of limited sublobar resection for early-stage lung cancer: steady progress. *J Clin Oncol* (2014) 32(23):2403–4. doi: 10.1200/JCO.2014.56.4203

8. Dembitzer FR, Flores RM, Parides MK, Beasley MB. Impact of histologic subtyping on outcome in lobar vs sublobar resections for lung cancer: a pilot study. *Chest* (2014) 146(1):175–81. doi: 10.1378/chest.13-2506

9. Yeh YC, Nitadori J, Kadota K, Yoshizawa A, Rekhtman N, Moreira AL, et al. Using frozen section to identify histological patterns in stage I lung adenocarcinoma of </= 3 cm: accuracy and interobserver agreement. *Histopathology* (2015) 66(7):922–38. doi: 10.1111/his.12468

10. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, et al. International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. *J Thorac Oncol* (2011) 6(2):244–85. doi: 10.1097/JTO.0b013e318206a221

11. Trejo Bittar HE, Incharoen P, Althouse AD, Dacic S. Accuracy of the IASLC/ATS/ERS histological subtyping of stage I lung adenocarcinoma on intraoperative frozen sections. *Mod Pathol* (2015) 28(8):1058–63. doi: 10.1038/modpathol.2015.71

12. Nitadori J, Bograd AJ, Kadota K, Sima CS, Rizk NP, Morales EA, et al. Impact of micropapillary histologic subtype in selecting limited resection vs lobectomy for lung adenocarcinoma of 2cm or smaller. *J Natl Cancer Inst* (2013) 105(16):1212–20. doi: 10.1093/jnci/djt166

13. Su H, Xie H, Dai C, Zhao S, Xie D, She Y, et al. Procedure-specific prognostic impact of micropapillary subtype may guide resection strategy in small-sized lung adenocarcinomas: a multicenter study. *Ther Adv Med Oncol* (2020) 12:1758835920937893. doi: 10.1177/1758835920937893

14. Yuan Y, Ma G, Zhang Y, Chen H. Presence of micropapillary and solid patterns are associated with nodal upstaging and unfavorable prognosis among patient with cT1N0M0 lung adenocarcinoma: a large-scale analysis. *J Cancer Res Clin Oncol* (2018) 144(4):743–9. doi: 10.1007/s00432-017-2571-7

15. Liu S, Wang R, Zhang Y, Li Y, Cheng C, Pan Y, et al. Precise diagnosis of intraoperative frozen section is an effective method to guide resection strategy for peripheral small-sized lung adenocarcinoma. *J Clin Oncol* (2016) 34(4):307–13. doi: 10.1200/JCO.2015.63.4907

16. Zhang Y, Deng C, Fu F, Ma Z, Wen Z, Ma X, et al. Excellent prognosis of patients with invasive lung adenocarcinomas during surgery misdiagnosed as atypical adenomatous hyperplasia, adenocarcinoma in situ, or minimally invasive adenocarcinoma by frozen section. *Chest* (2021) 159(3):1265–72. doi: 10.1016/j.chest.2020.10.076

17. Zhu E, Xie H, Dai C, Zhang L, Huang Y, Dong Z, et al. Intraoperatively measured tumor size and frozen section results should be considered jointly to predict the final pathology for lung adenocarcinoma. *Mod Pathol* (2018) 31(9):1391–9. doi: 10.1038/s41379-018-0056-0

18. Mateo J, Gerlinger M, Rodrigues DN, de Bono JS. The promise of circulating tumor cell analysis in cancer management. *Genome Biol* (2014) 15(8):448. doi: 10.1186/s13059-014-0448-5

19. Chen X, Zhou F, Li X, Yang G, Zhang L, Ren S, et al. Folate receptor-positive circulating tumor cell detected by LT-PCR-Based method as a diagnostic biomarker for non-Small-Cell lung cancer. *J Thorac Oncol* (2015) 10(8):1163–71. doi: 10.1097/JTO.0000000000000606

20. Lou J, Ben S, Yang G, Liang X, Wang X, Ni S, et al. Quantification of rare circulating tumor cells in non-small cell lung cancer by ligand-targeted PCR. *PloS One* (2013) 8(12):e80458. doi: 10.1371/journal.pone.0080458

21. Yu Y, Chen Z, Dong J, Wei P, Hu R, Zhou C, et al. Folate receptor-positive circulating tumor cells as a novel diagnostic biomarker in non-small cell lung cancer. *Transl Oncol* (2013) 6(6):697–702. doi: 10.1593/tlo.13535

22. Ding C, X, Xu C, Chen J, Ju S, Chen T, et al. Circulating tumor cell levels and carcinoembryonic antigen: An improved diagnostic method for lung adenocarcinoma. *Thorac Cancer* (2018) 9(11):1413–20. doi: 10.1111/1759-7714.12851

23. Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WE, et al. The IASLC lung cancer staging project: Proposals for revision of the TNM stage groupings in the forthcoming (Eighth) edition of the TNM classification for lung cancer. *J Thorac Oncol* (2016) 11(1):39–51. doi: 10.1016/j.jtho.2015.09.009

24. Board, W.C.o.T.E. WHO classification of tumours. In: *Thoracic tumours, 5thed*. Lyon: IARC Press (2021).

25. Marchevsky AM, Changsri C, Gupta I, Fuller C, Houck W, McKenna RJ, Jr. Frozen section diagnoses of small pulmonary nodules: accuracy and clinical implications. *Ann Thorac Surg* (2004) 78(5):1755–9. doi: 10.1016/j.athoracsur.2004.05.003

26. He P, Yao G, Guan Y, Lin Y, He J. Diagnosis of lung adenocarcinoma in situ and minimally invasive adenocarcinoma from intraoperative frozen sections: an analysis of 136 cases. *J Clin Pathol* (2016) 69(12):1076–80. doi: 10.1136/jclinpath-2016-203619

27. Yin W, Zhu J, Ma B, Jiang G, Zhu Y, He W, et al. Overcoming obstacles in pathological diagnosis of pulmonary nodules through circulating tumor cell enrichment. *Small* (2020) 16(25):e2001695. doi: 10.1002/smll.202001695

28. Liu X, Wang C, Zhang L, Gu H, Yang J, Wang Y, et al. Prognostic and diagnostic values of circulating tumor cells and tumor markers for lung cancer. *Clin Lab* (2021) 67(4). doi: 10.7754/Clin.Lab.2020.200654

29. Tsuta K, Kawago M, Inoue E, Yoshida A, Takahashi F, Sakurai H, et al. The utility of the proposed IASLC/ATS/ERS lung adenocarcinoma subtypes for disease prognosis and correlation of driver gene alterations. *Lung Cancer* (2013) 81(3):371–6. doi: 10.1016/j.lungcan.2013.06.012

30. Hou Y, Song W, Chen M, Zhang J, Luo Q, Um SW, et al. The presence of lepidic and micropapillary/solid pathological patterns as minor components has prognostic value in patients with intermediate-grade invasive lung adenocarcinoma. *Transl Lung Cancer Res* (2022) 11(1):64–74. doi: 10.21037/tlcr-21-934

31. Rhim AD, Mirek ET, Aiello NM, Maitra A, Bailey JM, McAllister F, et al. EMT and dissemination precede pancreatic tumor formation. *Cell* (2012) 148(1-2):349–61. doi: 10.1016/j.cell.2011.11.025

32. Hosseini H, Obradovic MMS, Hoffmann M, Harper KL, Sosa MS, Werner-Klein M, et al. Early dissemination seeds metastasis in breast cancer. *Nature* (2016) 540(7634):552–8. doi: 10.1038/nature20785

33. Gazzaniga P, Gradilone A, de Berardinis E, Busetto GM, Raimondi C, Gandini O, et al. Prognostic value of circulating tumor cells in nonmuscle invasive bladder cancer: a CellSearch analysis. *Ann Oncol* (2012) 23(9):2352–6. doi: 10.1093/annonc/mdr619

34. Xie H, Su H, Chen D, Xie D, Dai C, Ren Y, et al. Use of autofluorescence to intraoperatively diagnose visceral pleural invasion from frozen sections in patients with lung adenocarcinoma 2 cm or less. *Am J Clin Pathol* (2019) 152(5):608–15. doi: 10.1093/ajcp/aqz081

35. Villalba JA, Shih AR, Sayo TMS, Kunitoki K, Hung YP, Ly A, et al. Accuracy and reproducibility of intraoperative assessment on tumor spread through air spaces in stage 1 lung adenocarcinomas. *J Thorac Oncol* (2021) 16(4):619–29. doi: 10.1016/j.jtho.2020.12.005

36. Zhou F, Villalba JA, Sayo TMS, Narula N, Pass H, Mino-Kenudson M, et al. Assessment of the feasibility of frozen sections for the detection of spread through air spaces (STAS) in pulmonary adenocarcinoma. *Mod Pathol* (2022) 35(2):210–7. doi: 10.1038/s41379-021-00875-x

37. Su H, Gu C, She Y, Xu L, Yang P, Xie H, et al. Predictors of upstage and treatment strategies for stage IA lung cancers after sublobar resection for adenocarcinoma in situ and minimally invasive adenocarcinoma. *Transl Lung Cancer Res* (2021) 10(1):32–44. doi: 10.21037/tlcr-20-828

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 03 May 2023
DOI 10.3389/fonc.2023.1131816

Check for updates

OPEN ACCESS

EDITED BY
Hongda Liu,
Nanjing Medical University, China

REVIEWED BY
Nika Guberina,
University Duisburg-Essen, Germany
Burcin Celik,
Ondokuz Mayıs University, Türkiye

*CORRESPONDENCE
Linyou Zhang
lyzhang@hrbmu.edu.cn

†These authors have contributed equally to this work

RECEIVED 26 December 2022
ACCEPTED 24 April 2023
PUBLISHED 03 May 2023

CITATION
Peng B, Wang K, Xu R, Guo C, Lu T, Li Y, Wang Y, Wang C, Chang X, Shen Z, Shi J, Xu C and Zhang L (2023) Preoperative computed tomography-based tumoral radiomic features prediction for overall survival in resectable non-small cell lung cancer.
Front. Oncol. 13:1131816.
doi: 10.3389/fonc.2023.1131816

COPYRIGHT
© 2023 Peng, Wang, Xu, Guo, Lu, Li, Wang, Wang, Chang, Shen, Shi, Xu 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.

Preoperative computed tomography-based tumoral radiomic features prediction for overall survival in resectable non-small cell lung cancer

Bo Peng[1†], Kaiyu Wang[1†], Ran Xu[1†], Congying Guo[1], Tong Lu[1], Yongchao Li[1], Yiqiao Wang[2], Chenghao Wang[1], Xiaoyan Chang[1], Zhiping Shen[1], Jiaxin Shi[1], Chengyu Xu[1] and Linyou Zhang[1*]

[1]Department of Thoracic Surgery, The Second Affiliated Hospital of Harbin Medical University, Harbin, China, [2]Department of Radiology, The Second Affiliated Hospital of Harbin Medical University, Harbin, China

Objectives: The purpose of this study was to evaluate whether preoperative radiomics features could meliorate risk stratification for the overall survival (OS) of non-small cell lung cancer (NSCLC) patients.

Methods: After rigorous screening, the 208 NSCLC patients without any pre-operative adjuvant therapy were eventually enrolled. We segmented the 3D volume of interest (VOI) based on malignant lesion of computed tomography (CT) imaging and extracted 1542 radiomics features. Interclass correlation coefficients (ICC) and LASSO Cox regression analysis were utilized to perform feature selection and radiomics model building. In the model evaluation phase, we carried out stratified analysis, receiver operating characteristic (ROC) curve, concordance index (C-index), and decision curve analysis (DCA). In addition, integrating the clinicopathological trait and radiomics score, we developed a nomogram to predict the OS at 1 year, 2 years, and 3 years, respectively.

Results: Six radiomics features, including gradient_glcm_InverseVariance, logarithm_firstorder_Median, logarithm_firstorder_RobustMeanAbsoluteDeviation, square_gldm_LargeDependenceEmphasis, wavelet_HLL_firstorder_Kurtosis, and wavelet_LLL_firstorder_Maximum, were selected to construct the radiomics signature, whose areas under the curve (AUCs) for 3-year prediction reached 0.857 in the training set (n=146) and 0.871 in the testing set (n=62). The results of multivariate analysis revealed that the radiomics score, radiological sign, and N stage were independent prognostic factors in NSCLC. Moreover, compared with clinical factors and the separate radiomics model, the established nomogram exhibited a better performance in predicting 3-year OS.

Conclusions: Our radiomics model may provide a promising non-invasive approach for preoperative risk stratification and personalized postoperative surveillance for resectable NSCLC patients.

KEYWORDS

computed tomography, radiomics, non-small cell lung cancer, prognosis, LASSO

Introduction

Non-small cell lung cancer (NSCLC), the most essential subtype of lung cancer, represents a prevalent malignant tumor with an unsatisfactory prognosis (1). In recent years, an expanding availabilities of targeted therapy and immune checkpoint inhibitors (ICIs), has been approved for lung cancer and improved the long-term prognosis of NSCLC patients. However, due to the secondary mutations and low response rate to ICIs, only a limited number of patients can benefit from those therapeutic approaches (2, 3). Thus, precision diagnose and comprehensive prognostic evaluation are essential steps when dealing with resectable NSCLC patients, in order to select the most appropriate treatment. The tumor node metastasis (TNM) staging system is still a classic evaluation approach which can assist in adjuvant therapy choices and predict the outcome of NSCLC (4). Nonetheless, patients with the same TNM stage typically manifest different clinical outcomes, which is largely attributed to tumor heterogeneity and anatomical factors (5). With the advent of multi-omics, evaluation of prognostic features based on multidisciplinary methods make personalized medicine possible.

Medical imaging has been long regarded as standard procedure for early screening, treatment decision-making, and postoperative surveillance of cancer patients. Computed tomography (CT) imaging, commonly stored in the form of Digital Imaging and Communications in Medicine (DICOM), can be conveniently obtained and utilized for quantitative assessment. Over the last decade, radiomics has emerged as a hot research field that provides massive high-dimensional feature space derived from raw imaging data by automatically high-throughput algorithm (6). There is growing evidence that quantitative parameters and features mined from functional and morphological images offer a new perspective for tumor phenotypes and microenvironment, which also have a significant complementary interrelation with other omics approaches such as genomics, hematology, and proteomics (7–9). Radiomics analyses based on intratumoral and peritumoral regions have been extensively used for exploring underlying biological process, predicting pathological characteristics, evaluating the drug treatment response, and assisting therapeutic decision-making in several human carcinomas (10–14).

As a novelly emerging tool, radiomics provides a new direction of exploring intratumor heterogeneity and predictive markers using a noninvasive evaluation in lung cancer. Using three machine-learning (ML) classifiers derived from radiomics, Liu et al. estimated the benefit from Nivolumab treatment and progression probability in patients with stage IIIB/IV NSCLC (15). In terms of histological subtypes, an automatic deep-learning radiomics model showed satisfactory performance to distinguish lung squamous cell carcinoma (LUSC), lung adenocarcinoma (LUAD), and small cell lung cancer (SCLC) on CT images (16). However, there are limited studies investigating the contingent value of radiomics in improving prognosis stratification of resectable NSCLC patients.

Accordingly, the current study aimed to analyze the association of radiomics features with 3-year overall survival (OS) in enrolled patient cohort. Furthermore, the integration of radiomics model and clinicopathological traits was conducted to establish a comprehensive nomogram, which strengthen the prediction ability and may provide assistance in improving follow-up plans and individualized treatment in clinical practice.

Materials and methods

Patient data and study design

This study was approved by the Ethics and Scientific Committees of the Second Affiliated Hospital of Harbin Medical University (Approval Number: KY2022-144), due to the retrospective nature of this study, written informed consent for CT images was waived. The study design is illustrated in Figure 1.

493 consecutive NSCLC patients who underwent radical surgery (segmental resection, wedge resection, and lobectomy) at our institute from March 2015 to May 2019 were preliminarily included. All patients fulfilled the following inclusion criteria: (1) CT performed within 2 weeks before surgery; (2) Available CT images stored in DICOM format; and (3) Primary NSCLC confirmed by histopathology. The subsequent patient selection and exclusion criteria were visualized in Figure 1A. Consequently, we finally recruited 208 consecutive patients, including 167 cases with LUAD, 35 cases with LUSC, 3 cases with large cell carcinoma (LCC), 2 cases with adenosquamous carcinoma (ASC), and 1 case with carcinoid. According to the random allocation scheme and a ratio of 7:3, all patients were separated into two individual cohorts: 146 for training and 62 for testing. Patient grouping and corresponding clinicopathological traits are recapitulated in Table 1.

Follow-up

After radical surgery, all enrolled subjects were followed up by outpatient review or telephone every 3 months for the first year and every 6 months thereafter. We applied 3-year OS as the primary study endpoint, which is construed as the time between the operation and the date of all-cause death.

CT image acquisition and pre-processing

Helical CT images of all enrolled subjects were acquired by 64-channel CT (Discovery 750, GE Healthcare, Milwaukee, USA) and 256-channel CT (Revolution CT, GE Healthcare, Waukesha, WI, USA). Detailed scanning parameters were as follows: tube voltage, 120 kV; tube current, 100-250 mAs; slice thickness, 0.625-5 mm; field of view (FOV), 350-400mm; 512 x 512 matrix; and reconstructed slice thickness, 0.625-3mm. Filtered back projection (FBP) and adaptive statistical iterative reconstruction (ASIR) level 40% were utilized to reconstruct all enrolled CT images. A standard kernel was also used in the reconstruction procedure.

Due to the potential differences in specifications caused by distinct reconstruction slice thickness and voxel spacing, we performed image pre-processing before radiomics feature extracting. Specifically, we chose to resample the raw data to a standard voxel spacing of 1x1x1 mm^3 by near interpolation algorithm (17). Moreover, we adopted 25 HU as the fixed bin width to perform gray level discretization (18).

10.3389/fonc.2023.1131816

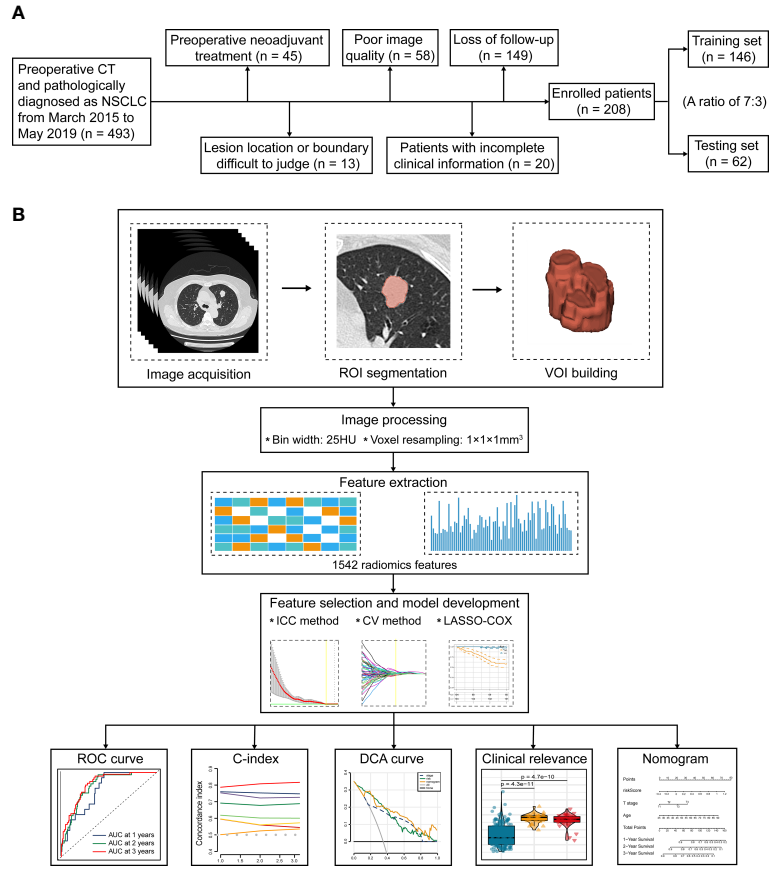


FIGURE 1
The flowchart of this study. **(A)** Patient selection and exclusion criteria of this work. **(B)** Tumor segmentation, radiomics feature extraction, feature selection, and radiomics model construction.

Tumor segmentation and radiomics feature extraction

CT images stored in DICOM format were loaded to 3D slicer (Version 4.1.1) software to perform tumor segmentation. As an open source platform, 3D slicer software has superiority in image interactive segmentation and medical raw data processing (19). A thoracic surgeon with 5 years of experience first manually delineated all lesions slice by slice in the lung window of CT images, the volumes of interest (VOIs) were then inspected and revised by an experienced radiologist. They were blinded to the clinical outcome and medical records of enrolled patients, in parallel, intratumoral or tumor-adjacent vessels, bronchi, and air spaces were carefully avoided during delineation of the tumor volume. To improve the reproducibility of radiomics features, the VOIs of 50 randomly selected CT images were repetitively delineated by another experienced thoracic surgeon, which were used for subsequent interclass correlation coefficient (ICC) analysis.

In our study, 1542 radiomics features were extracted from each manually-defined VOI using "pyradiomics" python package, the detailed definitions of which are congruity with Imaging Biomarker Standardization Initiative (IBSI). The extracted radiomics features can be assigned to the following categories: (1) Gray level dependence matrix-based features (GLDM); (2) neighboring gray tone difference matrix-based features (NGTDM); (3) gray level co-occurrence matrix-based features (GLCM); (4) gray level run-length matrix-based features (GLRLM); (5) gray level size zone matrix-based features (GLSZM); (6) first-order statistics features; (7) shape-based features, and (8) transformed features: features extracted from images pre-processed with several built-in filters including laplacian of gaussian (LoG), wavelet, logarithm, square, square root, and gradient. Detailed mathematic definitions and feature explanations can be acquired from the previous literature (20, 21).

Feature selection and radiomics model construction

To assess the consistency and robustness of extracted radiomics features between the two thoracic surgeons, we calculated the interclass correlation coefficient (ICC) value for each radiomics feature. The radiomics features with ICC values > 0.75 were retained for deeper analysis. Next, we performed data pre-processing using Z-score transformation by "caret" R package. Least absolute shrinkage and selection operator (LASSO) Cox regression analysis were utilized for further feature selection and model construction.

TABLE 1 Patient grouping and corresponding clinicopathological traits.

Covariates		Training set (n=146)	Testing set (n=62)	Total set (n=208)	P-value
Age, no (%)	≤ 65	99(67.81)	48(77.42)	147(70.67)	0.220
	> 65	47(32.19)	14(22.58)	61(29.33)	
Gender, no (%)	Female	77(52.74)	35(56.45)	112(53.85)	0.735
	Male	69(47.26)	27(43.55)	96(46.15)	
Smoking history, no (%)	No	89(60.96)	41(66.13)	130(62.50)	0.584
	Yes	57(39.04)	21(33.87)	78(37.50)	
T stage, no (%)	T1-2	131(89.73)	56(90.32)	187(89.90)	1.000
	T3-4	15(10.27)	6(9.68)	21(10.10)	
N stage, no (%)	N0	115(78.77)	48(77.42)	163(78.37)	0.975
	N1-2	31(21.23)	14(22.58)	45(21.63)	
Pathologic TNM stage, no (%)	Stage I	101(69.18)	46(74.19)	147(70.67)	0.754
	Stage II	21(14.38)	7(11.29)	28(13.46)	
	Stage III	24(16.44)	9(14.52)	33(15.87)	
Lateral location, no (%)	Left	60(41.10)	30(48.39)	90(43.27)	0.413
	Right	86(58.90)	32(51.61)	118(56.73)	
Lobe location, no (%)	Mid-lower	65(44.52)	32(51.61)	97(46.63)	0.432
	Upper	81(55.48)	30(48.39)	111(53.37)	
Location classification, no (%)	Central	29(19.86)	7(11.29)	36(17.31)	0.196
	Peripheral	117(80.14)	55(88.71)	172(82.69)	
Max diameter, no (%)	≤3cm	112(76.71)	49(79.03)	161(77.40)	0.854
	>3cm	34(23.29)	13(20.97)	47(22.60)	
Pathological type, no (%)	LUAD	112(76.71)	55(88.71)	167(80.29)	0.138
	LUSC	29(19.86)	6(9.68)	35(16.83)	
	Other	5(3.42)	1(1.61)	6(2.88)	
Histological grade, no (%)	Well	24(16.44)	8(12.9)	32(15.38)	0.644
	Moderate	44(30.14)	15(24.19)	59(28.37)	
	Poor	50(34.25)	26(41.94)	76(36.54)	
	Unknown	28(19.18)	13(20.97)	41(19.71)	
Radiological sign, no (%)	Cavitary appearance	30(20.55)	21(33.87)	51(24.52)	0.091
	Part-solid appearance	40(27.4)	12(19.35)	52(25)	
	Pure ground-glass appearance	16(10.96)	10(16.13)	26(12.5)	
	Pure solid appearance	60(41.1)	19(30.65)	79(37.98)	

LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma.

In the training set, LASSO with ten-fold cross-validation effectively avoided over-fitting and identified optimal features with nonzero coefficients by applying a constraint on the model hyperparameter (λ). Finally, we obtained a linear combination of optimal radiomics features weighted with the regression coefficients during multiple computing. All the selected features for model construction are detailed in Table 2.

Radiomics model validation and evaluation

The prognostic efficacy of radiomics model established in the training set was subsequently verified in the testing set and the total enrolled subjects. Patients in these three sets were respectively dichotomized into low- and high-risk groups based on the median radiomics score threshold. To appraise diversities in the

TABLE 2 Six selected radiomics features.

	Selected radiomics feature	Radiomics group	Filter associated	Regression coefficient
R1	InverseVariance	GLCM	gradient	0.068
R2	Median	First order	logarithm	0.096
R3	RobustMeanAbsoluteDeviation	First order	logarithm	0.181
R4	LargeDependenceEmphasis	GLDM	square	0.375
R5	Kurtosis	First order	wavelet_HLL	0.016
R6	Maximum	First order	wavelet_LLL	0.269

GLCM, gray level co-occurrence matrix; GLDM, Gray level dependence matrix.

overall survival between two cohorts with different radiomics score, R packages survival and survMiner were used to perform Kaplan-Meier survival analysis. We utilized receiver operating characteristic (ROC) curve analysis and calculated the area under the curve (AUC) to access the sensitivity and specificity of the prognostic model. The R packages survminer, survival, and timeROC were employed to enable this process. Moreover, Harrell's concordance index (C-index) was applied to estimate the predictive accuracy of the signature. AUC and C-index both ranges from 0.5 (poor predictive performance) to 1 (perfect predictive performance).

A nomogram comprising of radiomics score and clinicopathological parameters (age, gender, smoking history, pathological type, lobe location, location classification, lateral location, max diameter, T stage, N stage, and TNM stage) was set up to predict the 1-, 2-, and 3-year OS of resectable NSCLC patients. Correction curve analysis, ROC curve analysis, and C-index were applied to evaluate the predictive performance of the nomogram. In addition, we performed decision curve analysis (DCA) to compare the clinical benefit of the classical TNM stage, established radiomics model, and the nomogram by quantifying the net benefits (22).

Statistical analyses

All statistical analyses were carried out by R software (version 4.1.2) and Python software (version 3.4.3). The statistically significant threshold was set to p value < 0.05 (*p < 0.05, **p < 0.01, ***p < 0.001).

Results

Construction of the radiomics model

All enrolled 208 subjects were randomly divided into training set (n=146) and testing set (n=62) with a ratio of 7:3, and there was no statistical difference in terms of clinical features (Table 1). The training set was applied to create the prognostic radiomics model, we then used the testing set and the total set to validate the constructed model.

According to the 1542 radiomics features extracted by two experienced thoracic surgeons, The ICC analysis indicated good consistency in 1178 (76.39%) radiomics features, moderate consistency in 266 (17.25%) radiomics features, and poor consistency in 98 (6.36%) radiomics features. We ultimately included the 1178 most stable features in the subsequent model construction. LASSO Cox regression analysis is widely used in multiple regression analysis, which not only optimize the selection of characteristics with a deficient correlation and prominent predicted value from high-dimensional data, but also improve the forecast accuracy. We performed tenfold cross-validation to select the minimal penalty term (λ) (Figures 2A, B).

Afterwards, we established a prognostic radiomics model implicating six selected features for resectable NSCLC patients. The formula was constructed as follows: Radiomics score = 0.068 × gradient_glcm_InverseVariance + 0.096 × logarithm_firstorder_Median + 0.181 × logarithm_firstorder_RobustMeanAbsoluteDeviation + 0.375 × square_gldm_LargeDependenceEmphasis + 0.016 × wavelet_HLL_firstorder_Kurtosis + 0.269 × wavelet_LLL_firstorder_Maximum.

Survival analyses in the training set, testing set, and the entire NSCLC set

We first dichotomize 146 NSCLC patients of training set into low- and high-risk groups based on the median radiomics score. Kaplan-Meier analysis showed significantly distinct prognoses between the two risk groups (Figure 2C). In addition, the distributions of the radiomics score, survival time, and survival status of each NSCLC patient in the training set were shown in Figure 2D. To further verify the prognostic performance of the constructed radiomics model, we acquired the radiomics score of each NSCLC patient in the testing set and the total NSCLC set using the radiomics score formula. Employing similar segmentation method, we divided the testing set and the total NSCLC set into subgroups with different risk level. The Kaplan-Meier analysis demonstrated that the patients in the high-risk group had significantly poorer outcome (Figures 2F, 3A). The distribution patterns of the radiomics score, survival time, and survival status of both sets were illustrated in Figures 2G, 3B.

Performance evaluation of the radiomics model

ROC curve analysis indicated that the AUC of established radiomics model reached 0.857 and 0.871 at 3 years in the

10.3389/fonc.2023.1131816

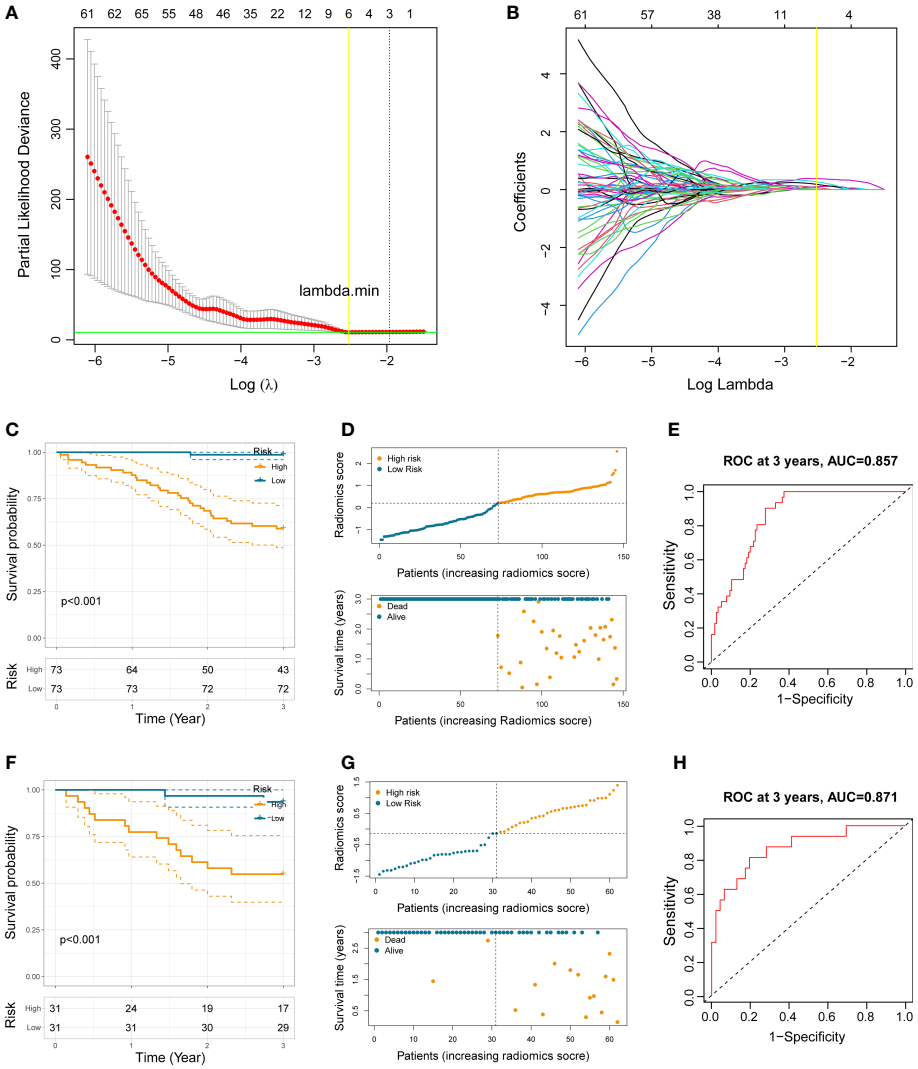


FIGURE 2
Construction of the prognostic radiomics model. **(A)** Selection of the tuning parameter (log λ) based on minimum criteria in the LASSO analysis. **(B)** LASSO coefficient profiles. **(C, F)** Kaplan-Meier analysis in the training set and the testing set. **(D, G)** The distribution patterns of radiomics score, survival time, and survival status of NSCLC patients in the training set and the testing set. **(E, H)** ROC curves of the radiomics model at 3 years in the training set and the testing set.

training set and the testing set (Figures 2E, H). Notably, the AUC of our model in total NSCLC set reached 0.792, 0.839, and 0.861 at 1 year, 2 years, and 3 years, respectively (Figure 3C). When compared with other clinicopathological variables in terms of ROC analysis, the radiomics model still had an advantage in evaluating precision and sensitivity (Figure 3D). Furthermore, The C-index curve showed promising predictive accuracy of the radiomics model from another dimension (Figure 3E).

To assess the independent prognostic value of the radiomics model, we incorporated age, gender, smoking history, pathological type, lobe location, location classification, lateral location, max diameter, T stage, N stage, TNM stage, radiological sign, and radiomics score into the univariate and multivariate Cox regression analyses. The results suggested that the established radiomics model, N stage, and radiological sign could act as independent predictors for overall survival in resectable NSCLC patients (Figure 4, $p < 0.001$, $p < 0.05$, and $p < 0.05$).

To examine the applicability of our radiomics model in subgroups stratified by different clinicopathological traits, the total 208 NSCLC subjects were dichotomized into disparate subgroups by age, gender, smoking history, histological grade, lobe location, location classification, lateral location, max diameter, T stage, N stage, TNM stage, and radiological sign. In the overwhelming majority of subgroups, our radiomics model could accurately distinguish those high-risk patients with worse outcome (Figure 5 and Supplementary Figure S1).

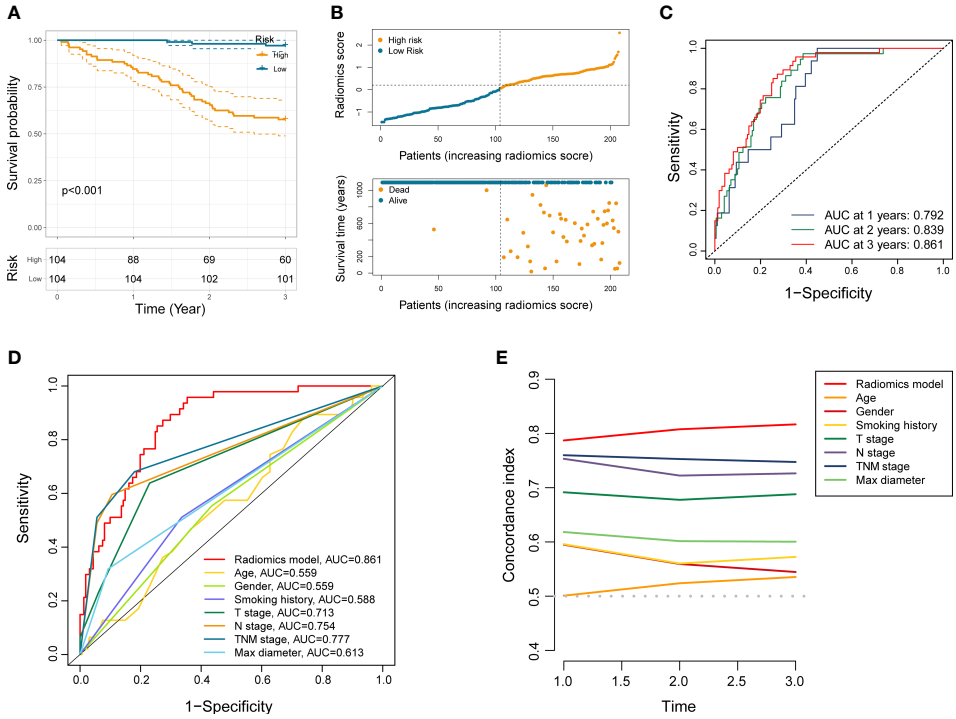


FIGURE 3
Performance evaluation of the radiomics model. **(A)** Kaplan-Meier analysis in entire NSCLC set. **(B)** The distribution patterns of radiomics score, survival time, and survival status of NSCLC patients in entire NSCLC set. **(C)** ROC analysis of the radiomics model in entire NSCLC set at 1 year, 2 years, and 3 years, respectively. **(D, E)** ROC and C-index curves comparing the radiomics model and other clinical characteristics.

Correlation between constructed radiomics model and clinicopathological features

We next performed correlation analysis to dig deeper connections between radiomics score and the clinicopathological traits obtained from the electronic medical record. The results revealed that higher radiomics score with worse survival was significantly associated with age > 65 ($p < 0.05$), male ($p < 0.001$), smoking history ($p < 0.001$), T2-4 ($p < 0.001$), N1-2 ($p < 0.001$), TNM stage II-III ($p < 0.001$), max diameter > 3cm ($p < 0.001$), poor differentiation grade ($p < 0.001$), central-type NSCLC ($p < 0.001$), LUSC and other pathological types ($p < 0.01$), middle and lower lobes ($p < 0.01$), and pure solid appearance ($p < 0.001$), which is similar to previous experiences acquired from clinical practice and may provide additional clues to clinical management of NSCLC (Figure 6A and Supplementary Figure S2).

Development and assessment of the prognostic nomogram

In order to furnish a comprehensive prognostic tool to predict the survivability of resectable NSCLC patients at 1, 2, and 3 years, we constructed a nomogram integrating the radiomics model and clinicopathological parameters (Figure 6B). In terms of predicting OS at 1 year, 2 years, and 3 years, calibration plot displayed that there is a decent consistency between the prediction curve and the ideal curve (Figure 7A). The AUC values of the prognostic nomogram were 0.863, 0.870, and 0.898 at 1-, 2–, and 3–year, respectively, which exhibited better performance than radiomics model and other clinicopathological parameters (Figures 7B, C). Moreover, the C-index values of the comprehensive nomogram and established radiomics model reached 0.854 and 0.814 at 3 years, respectively. Clinical utility is commonly utilized to measure the practical clinical value of prognostic models. Subsequent DCA curves revealed that the appropriate combination of radiomics model and the comprehensive nomogram may bring significantly more benefit than TNM staging system in clinical work (Figure 7D).

Discussion

In this retrospective study, we explored the potential value of radiomics features in predicting 3-year OS for resectable NSCLC patients varied from pathologic stage IA to stage IIIA. The established radiomics model exhibited good prediction accuracy with an AUC of 0.857 in the training set (n = 146) and an AUC of 0.871 in the testing set (n = 62). The stratified analysis indicated that our radiomics model could distinguish those patients with worse prognosis in the vast majority of subgroups. Subsequently, a comprehensive nomogram incorporating radiomics model and clinical parameters further enhanced the prognostic performance of the single radiomics model with an AUC of 0.898. Thus, this radiomics classifier could be an advantageous noninvasive biomarker in the whole-course clinical management of resectable NSCLC patients.

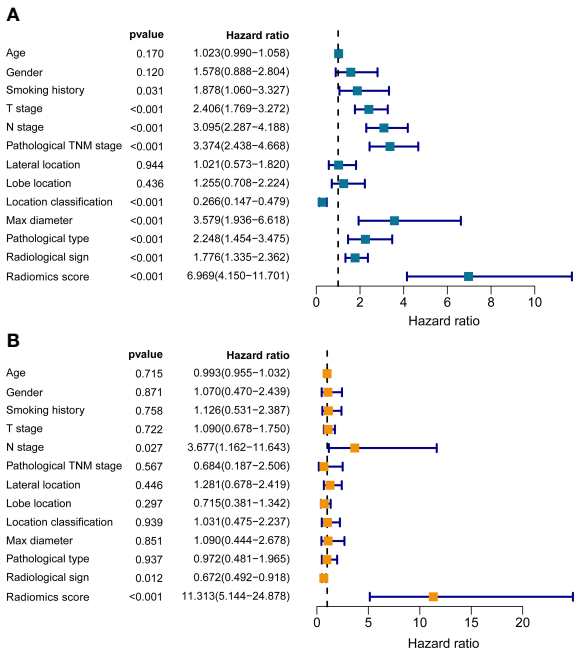


FIGURE 4
The univariate **(A)** and multivariate **(B)** Cox regression analyses of radiomics score and other clinical traits.

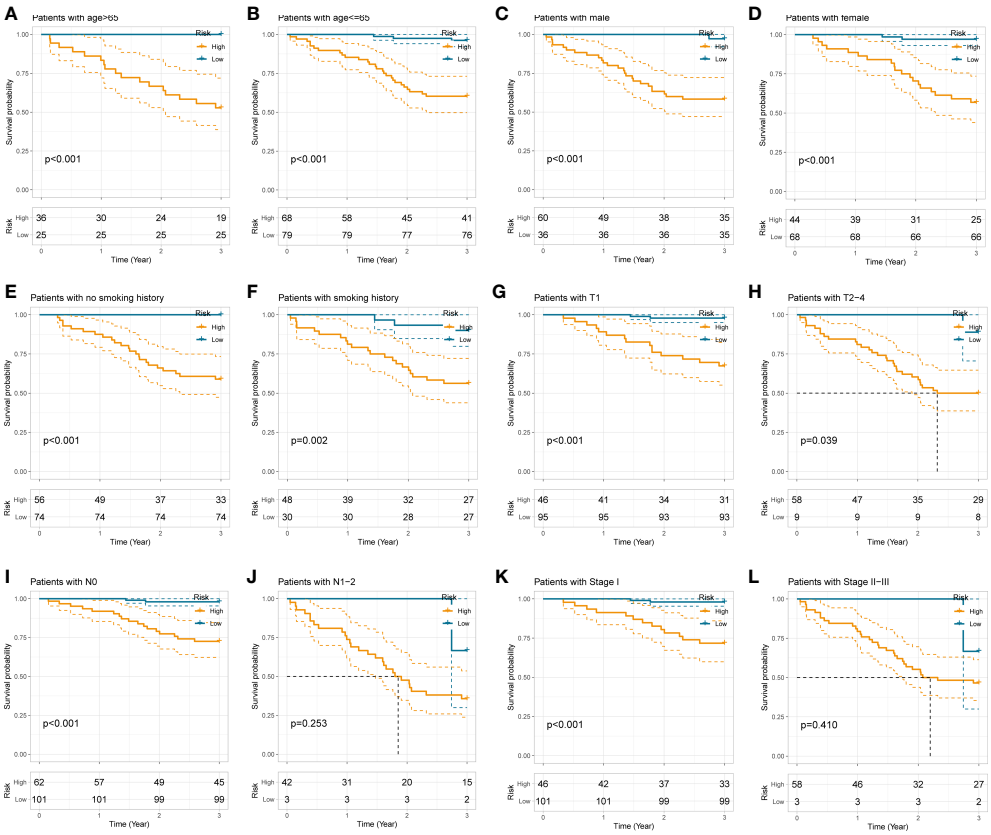


FIGURE 5
Stratification analyses of the radiomics model in different subgroups stratified by age **(A, B)**, gender **(C, D)**, smoking history **(E, F)**, T stage **(G, H)**, N stage **(I, J)**, and pathological TNM stage **(K, L)**.

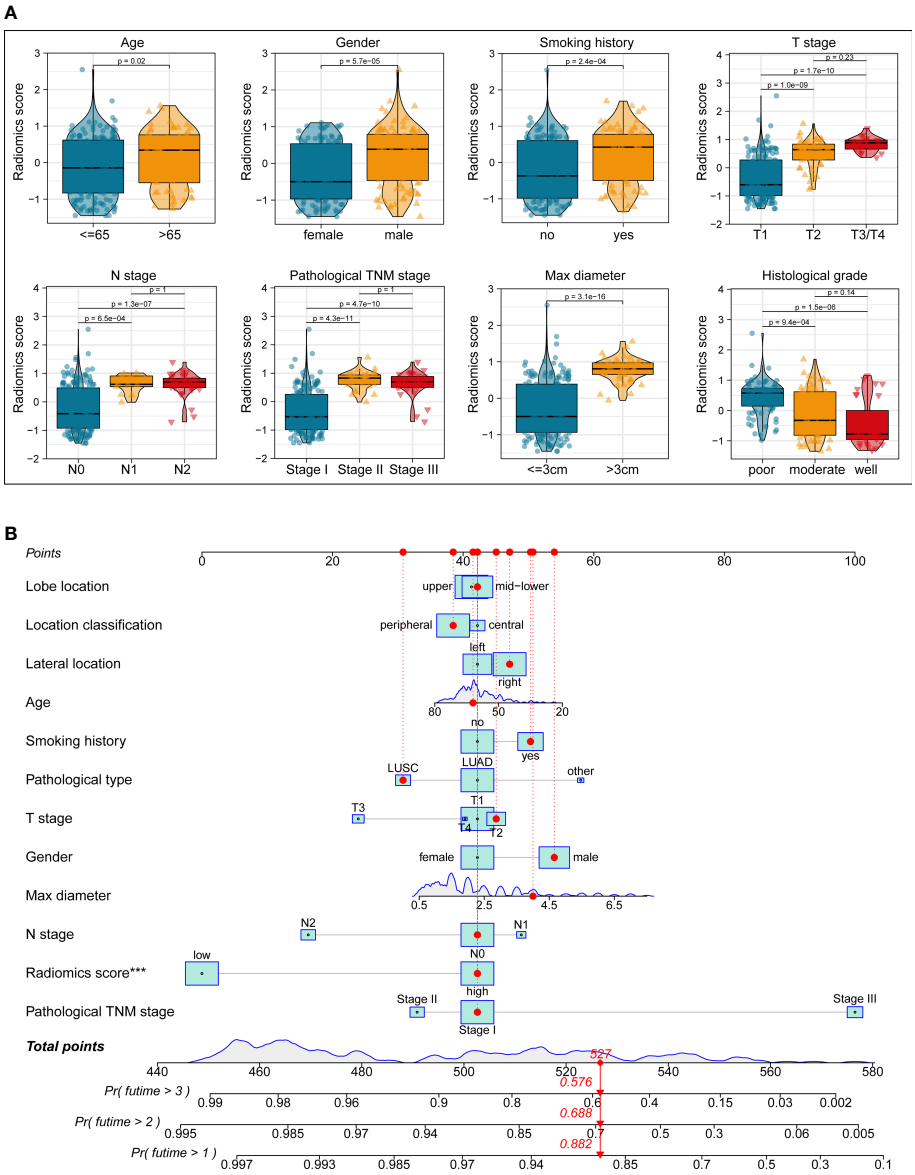


FIGURE 6
Correlation analyses **(A)** between radiomics score and clinicopathological features. A comprehensive nomogram **(B)** integrating the radiomics model and clinicopathological parameters to predict the 1-, 2-, and 3-year OS of resectable NSCLC patients.

The generic workflow of NSCLC biomarkers comprises various prognostic models reliant on clinical elements including lserum tumor markers, specific gene expression, next-generation sequencing, and circulating tumor DNA (ctDNA) (23–25). For instance, lu et al. constructed the tumor mutation index (TMI) model based on ctDNA sequencing to predict OS and recognize NSCLC patients who may respond well to monotherapy with atezolizumab or docetaxel (26). To improve NSCLC treatment in terms of chemotherapy and immunotherapy, Guo et al. developed a 7-gene predictive signature using qRT-PCR assays based on 337 snap-frozen NSCLC tissues. Patients receiving adjuvant chemotherapy were precisely identified with significantly better disease-specific survival in the predicted benefit group *via* the multi-gene prognostic signature (27). However, the few available predictive models based on quantitative gene expression levels in the clinical practice have usual limitations such as invasive procedures, time-consuming, cost-effectiveness, and some degree of interference to clinical workflow.

In contrast, radiomics have exhibited a bright prospect for prognosis, diagnosis, and treatment response prediction, as well as long-term health surveillance of NSCLC treatment in a non-invasive modality. Several studies focused on exploring potential combined signature based on radiomics features for NSCLC patients at specific pathologic stage. Huang et al. extracted 132 texture features from CT images of early stage NSCLC (stage I or II) and obtained a better performance for disease-free survival (DFS) prediction (C-index = 0.72) when incorporating the radiomics signature into a comprehensive nomogram (28). Xie et al. enrolled 554 candidates with resected stage I LUAD from three multicenter cohorts and further recognized potential subjects who

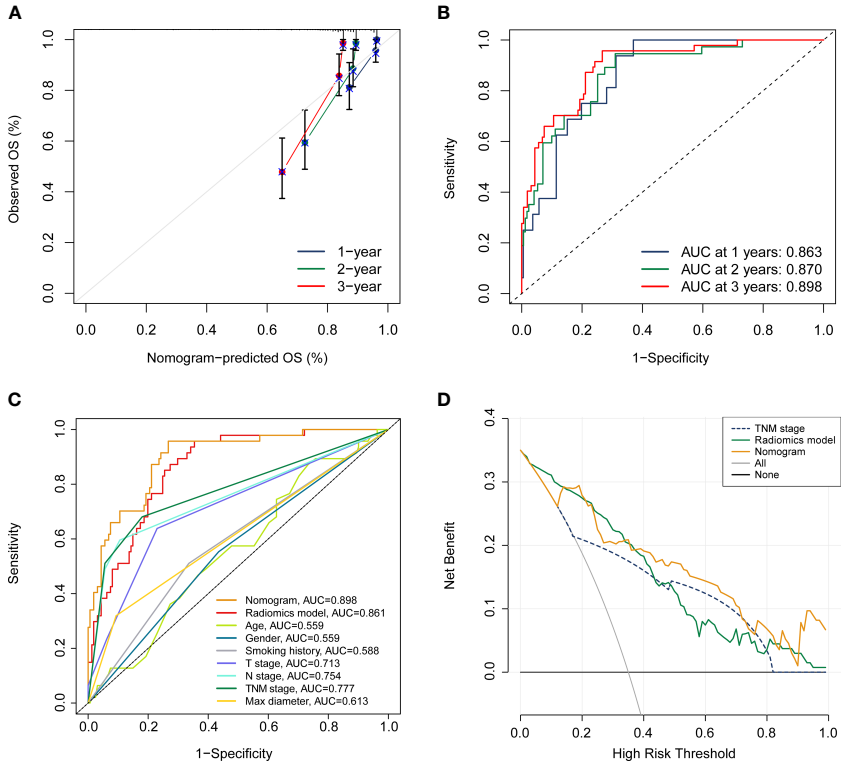


FIGURE 7
Assessment of the prognostic nomogram. **(A, B)** calibration curve and ROC analyses of the established nomogram at 1 year, 2 years, and 3 years, respectively. **(C)** Comparison of the comprehensive nomogram, radiomics model, and other clinical characteristics by ROC curves. **(D)** DCA curves of the nomogram, radiomics model, and pathological TNM stage.

may benefit from adjuvant chemotherapy (29). Moreover, a number of studies mined high-dimensional clues from functional and metabolic images of ^{18}F-FDG-PET/CT, aiming to improve clinical decision of epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors (TKIs) or immune checkpoint inhibitors (ICIs) treatments for specific NSCLC patient populations (30–34). In this study, we restricted the entrance criteria to preoperative CT scans, since CT scans serve as the major approaches for lung cancer screening and whole-course monitoring in a real-world clinical environment, especially in relatively early stages of NSCLC.

Six screened radiomics features used in our prediction model consisted of four first-order-based features (Maximum, Median, Kurtosis, and RobustMeanAbsoluteDeviation), one GLDM-based feature (LargeDependenceEmphasis), and one GLCM-based feature (InverseVariance), which were extracted from CT images pre-processed with built-in filters including wavelet_HLL, wavelet_LLL, logarithm, square, and gradient. First-order statistics features are generally acquired by measuring the gray values of region of interest (ROI) cropping, which reflects the intratumoral distribution of grayscale intensity (20). Previous literature indicated that two First-order statistics features may serve as radiomics predictors for identifying invasive phenotype of solitary pulmonary nodule (35). Moreover, GLDM-based features manifest the intrinsic grayscale associations of central voxel with neighboring voxels, which may be a reflection of heterogeneity and homogeneity of tumors. Padmakumari et al. demonstrated that LargeDependenceEmphasis, a radiomics feature from GLDM, can exhibit robust performance in discriminating lung cancer from tuberculosis with an AUC of 0.92 (36). GLCM describes the integrated information about spatial correlation characteristics of pixel pairs in terms of the pattern of grayscale arrangement, direction, distance, and gray value (37). Notably, textural features derived from GLCM have been demonstrated to have pathological association and can be applied to the diagnosis of malignant lesion in breast cancer (38). Granata et al. identified Correlation from GLCM as a reliable predictor for recognizing tumor recurrence in colorectal liver metastases patients (39).

In our study, we identified the established radiomics model, N stage, and radiological sign as independent predicting indicators for OS in in resectable NSCLC patients. Indeed, several studies have confirmed the accuracy and reliability of radiomics signature in predicting prognosis in NSCLC (40–42). Yang et al. incorporated the radiomics signatures and four clinicopathological features (N stage, T stage, age, and sex) to construct a comprehensive nomogram for survival prediction in NSCLC patients at stage I/II, the performance of which was measured by a C-index of 0.710 (41). The radiomics model and corresponding nomogram in this study showed more satisfactory performance with C-index values of 0.814 and 0.854, respectively. Subsequent DCA and calibration analyses further supported their clinical utility. In the current clinical practice, adjuvant chemotherapy after surgery is not recommended for patients with pathologic stage IA and there is a

controversial debate regarding its potential benefit for stage IB (43). In the model evaluation section, our radiomics model seemed to be more robust than traditional TNM staging system. In stratified analyses, the established radiomics model could dichotomize participants with pathologic stage I into high- and low-risk groups by Kaplan-Meier method, which would be helpful to performed personalized treatment interventions on these high-risk patients with worse prognosis.

Some limitations of this study have to be acknowledged. First, due to the relatively small sample size from single center and the retrospective nature, potential selection bias may obstruct the robustness and generalizability of our radiomics model. Therefore, it is necessary to recruit more subjects and perform multicentric external verification in future research. Second, the 3-year follow-up period of included subjects was relatively short, we will conduct the remaining follow-up until 5 years in the next work. Third, we delineated all lesions manually, which was laborious and time-consuming. Automatic delineation based on deep learning method is worth further study to improve the workflow of radiomics in busy clinical practice.

In summary, the current study proposed a novel non-invasive approach based on preoperative CT scans that can predict OS in patients with NSCLC after radical surgery, which may provide clues to help clinicians improve clinical decisions and guide personalized treatment. However, further external validation is warranted before its widespread application in clinical practice.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by The Ethics and Scientific Committees of the Second Affiliated Hospital of Harbin Medical University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

BP conceived the study, conducted radiomics analysis, and wrote the preliminary draft. KW and RX participated in the radiomics analysis and revised the manuscript. CG, TL, YL, CW, and XC contributed to study design and data preprocess. YW contributed to tumor segmentation and provided professional suggestions for statistical analysis. ZS, JS, and CX contributed to data collection. The manuscript was revised and approved by LZ. All authors contributed to the article and approved the submitted version.

Acknowledgments

We sincerely thank professor Huijie Jiang (Department of Radiology, The Second Affiliated Hospital of Harbin Medical University, China), for her precious suggestions to the work. We also thank all participants for their contribution to the study.

Conflict of interest

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

Publisher's note

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.2023.1131816/full#supplementary-material

References

1. Muller DC, Johansson M, Brennan P. Lung cancer risk prediction model incorporating lung function: development and validation in the UK biobank prospective cohort study. *J Clin Oncol* (2017) 35(8):861–9. doi: 10.1200/JCO.2016.69.2467

2. He J, Hu Y, Hu M, Li B. Development of PD-1/PD-L1 pathway in tumor immune microenvironment and treatment for non-small cell lung cancer. *Sci Rep* (2015) 5:13110. doi: 10.1038/srep13110

3. Pasche B, Grant S. Non-small cell lung cancer and precision medicine: a model for the incorporation of genomic features into clinical trial design. *JAMA* (2014) 311 (19):1975–6. doi: 10.1001/jama.2014.3742

4. Groome PA, Bolejack V, Crowley JJ, Kennedy C, Krasnik M, Sobin LH, et al. The IASLC lung cancer staging project: validation of the proposals for revision of the T, n, and m descriptors and consequent stage groupings in the forthcoming (seventh) edition of the TNM classification of malignant tumours. *J Thorac Oncol* (2007) 2 (8):694–705. doi: 10.1097/JTO.0b013e31812d05d5

5. Rami-Porta R, Asamura H, Goldstraw P. Predicting the prognosis of lung cancer: the evolution of tumor, node and metastasis in the molecular age-challenges and opportunities. *Transl Lung Cancer Res* (2015) 4(4):415–23. doi: 10.3978/j.issn.2218-6751.2015.07.11

6. Tunali I, Gillies RJ, Schabath MB. Application of radiomics and artificial intelligence for lung cancer precision medicine. *Cold Spring Harb Perspect Med* (2021) 11(8):a039537. doi: 10.1101/cshperspect.a039537

7. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* (2016) 278(2):563–77. doi: 10.1148/radiol.2015151169

8. Wei W, Wang K, Liu Z, Tian K, Wang L, Du J, et al. Radiomic signature: a novel magnetic resonance imaging-based prognostic biomarker in patients with skull base chordoma. *Radiother Oncol* (2019) 141:239–46. doi: 10.1016/j.radonc.2019.10.002

9. Thawani R, McLane M, Beig N, Ghose S, Prasanna P, Velcheti V, et al. Radiomics and radiogenomics in lung cancer: a review for the clinician. *Lung Cancer* (2018) 115:34–41. doi: 10.1016/j.lungcan.2017.10.015

10. 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(6):1785–94. doi: 10.1007/s00259-020-05142-w

11. Zhong Y, She Y, Deng J, Chen S, Wang T, Yang M, et al. Deep learning for prediction of N2 metastasis and survival for clinical stage I non-small cell lung cancer. *Radiology* (2022) 302(1):200–11. doi: 10.1148/radiol.2021210902

12. Li Y, Zhang Y, Fang Q, Zhang X, Hou P, Wu H, et al. Radiomics analysis of [18F]FDG PET/CT for microvascular invasion and prognosis prediction in very-early- and early-stage hepatocellular carcinoma. *Eur J Nucl Med Mol Imaging* (2021) 48 (8):2599–614. doi: 10.1007/s00259-020-05119-9

13. Fang J, Zhang B, Wang S, Jin Y, Wang F, Ding Y, et al. Association of MRI-derived radiomic biomarker with disease-free survival in patients with early-stage cervical cancer. *Theranostics* (2020) 10(5):2284–92. doi: 10.7150/thno.37429

14. Kirienko M, Sollini M, Corbetta M, Voulaz E, Gozzi N, Interlenghi M, et al. Radiomics and gene expression profile to characterise the disease and predict outcome in patients with lung cancer. *Eur J Nucl Med Mol Imaging* (2021) 48(11):3643–55. doi: 10.1007/s00259-021-05371-7

15. Liu C, Gong J, Yu H, Liu Q, Wang S, Wang J. A CT-based radiomics approach to predict nivolumab response in advanced non-Small-Cell lung cancer. *Front Oncol* (2021) 11:544339. doi: 10.3389/fonc.2021.544339

16. Guo Y, Song Q, Jiang M, Guo Y, Xu P, Zhang Y, et al. Histological subtypes classification of lung cancers on CT images using 3D deep learning and radiomics. *Acad Radiol* (2021) 28(9):e258–66. doi: 10.1016/j.acra.2020.06.010

17. Zhao H, Li W, Lyu P, Zhang X, Liu H, Liang P, et al. TCGA-TCIA-Based CT radiomics study for noninvasively predicting Epstein-Barr virus status in gastric cancer. *AJR Am J Roentgenol* (2021) 217(1):124–34. doi: 10.2214/AJR.20.23534

18. Larue RTHM, van Timmeren JE, de Jong EEC, Feliciani G, Leijenaar RTH, Schreurs WMJ, et al. Influence of gray level discretization on radiomic feature stability for different CT scanners, tube currents and slice thicknesses: a comprehensive phantom study. *Acta Oncol* (2017) 56(11):1544–53. doi: 10.1080/0284186X.2017.1351624

19. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, et al. 3D slicer as an image computing platform for the quantitative imaging network. *Magn Reson Imaging* (2012) 30(9):1323–41. doi: 10.1016/j.mri.2012.05.001

20. Mayerhoefer ME, Materka A, Langs G, Häggström I, Szczypiński P, Gibbs P, et al. Introduction to radiomics. *J Nucl Med* (2020) 61(4):488–95. doi: 10.2967/jnumed.118.222893

21. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res* (2017) 77(21):e104–7. doi: 10.1158/0008-5472.CAN-17-0339

22. Fitzgerald M, Saville BR, Lewis RJ. Decision curve analysis. *JAMA* (2015) 313 (4):409–10. doi: 10.1001/jama.2015.37

23. Vaidya P, Bera K, Gupta A, Wang X, Corredor G, Fu P, et al. CT derived radiomic score for predicting the added benefit of adjuvant chemotherapy following surgery in stage I, II resectable non-small cell lung cancer: a retrospective multi-cohort study for outcome prediction. *Lancet Digit Health* (2020) 2(3):e116–28. doi: 10.1016/s2589-7500(20)30002-9

24. Li B, Cui Y, Diehn M, Li R. Development and validation of an individualized immune prognostic signature in early-stage nonsquamous non-small cell lung cancer. *JAMA Oncol* (2017) 3(11):1529–37. doi: 10.1001/jamaoncol.2017.1609

25. Devarakonda S, Rotolo F, Tsao MS, Lanc I, Brambilla E, Masood A, et al. Tumor mutation burden as a biomarker in resected non-Small-Cell lung cancer. *J Clin Oncol* (2018) 36(30):2995–3006. doi: 10.1200/JCO.2018.78.1963

26. Lu J, Wu J, Lou Y, Shi Q, Xu J, Zhang L, et al. Blood-based tumour mutation index act as prognostic predictor for immunotherapy and chemotherapy in non-small cell lung cancer patients. *biomark Res* (2022) 10(1):55. doi: 10.1186/s40364-022-00400-5

27. Guo NL, Dowlati A, Raese RA, Dong C, Chen G, Beer DG, et al. A predictive 7-gene assay and prognostic protein biomarkers for non-small cell lung cancer. *EBioMedicine* (2018) 32:102–10. doi: 10.1016/j.ebiom.2018.05.025

28. Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z, et al. Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell lung cancer. *Radiology* (2016) 281(3):947–57. doi: 10.1148/radiol.2016152234

29. Xie D, Wang TT, Huang SJ, Deng JJ, Ren YJ, Yang Y, et al. Radiomics nomogram for prediction disease-free survival and adjuvant chemotherapy benefits in patients with resected stage I lung adenocarcinoma. *Transl Lung Cancer Res* (2020) 9 (4):1112–23. doi: 10.21037/tlcr-19-577

30. Mu W, Jiang L, Zhang J, Shi Y, Gray JE, Tunali I, et al. Non-invasive decision support for NSCLC treatment using PET/CT radiomics. *Nat Commun* (2020) 11 (1):5228. doi: 10.1038/s41467-020-19116-x

31. Mu W, Tunali I, Gray JE, Qi J, Schabath MB, Gillies RJ. Radiomics of 18F-FDG PET/CT images predicts clinical benefit of advanced NSCLC patients to checkpoint blockade immunotherapy. *Eur J Nucl Med Mol Imaging* (2020) 47(5):1168–82. doi: 10.1007/s00259-019-04625-9

32. Zhang J, Zhao X, Zhao Y, Zhang J, Zhang Z, Wang J, et al. Value of pre-therapy 18F-FDG PET/CT radiomics in predicting EGFR mutation status in patients with non-small cell lung cancer. *Eur J Nucl Med Mol Imaging* (2020) 47(5):1137–46. doi: 10.1007/s00259-019-04592-1

33. Mu W, Jiang L, Shi Y, Tunali I, Gray JE, Katsoulakis E, et al. Non-invasive measurement of PD-L1 status and prediction of immunotherapy response using deep learning of PET/CT images. *J Immunother Cancer* (2021) 9(6):e002118. doi: 10.1136/jitc-2020-002118

34. Monaco L, De Bernardi E, Bono F, Cortinovis D, Crivellaro C, Elisei F, et al. The "digital biopsy" in non-small cell lung cancer (NSCLC): a pilot study to predict the PD-L1 status from radiomics features of [18F]FDG PET/CT. *Eur J Nucl Med Mol Imaging* (2022) 49(10):3401–11. doi: 10.1007/s00259-022-05783-z

35. Weng Q, Zhou L, Wang H, Hui J, Chen M, Pang P, et al. A radiomics model for determining the invasiveness of solitary pulmonary nodules that manifest as part-solid nodules. *Clin Radiol* (2019) 74(12):933–43. doi: 10.1016/j.crad.2019.07.026

36. Thattaamuriyil Padmakumari L, Guido G, Caruso D, Nacci I, Del Gaudio A, Zerunian M, et al. The role of chest CT radiomics in diagnosis of lung cancer or tuberculosis: a pilot study. *Diagnostics (Basel)* (2022) 12(3):739. doi: 10.3390/diagnostics12030739

37. Qiu L, Zhang X, Mao H, Fang X, Ding W, Zhao L, et al. Comparison of comprehensive morphological and radiomics features of subsolid pulmonary nodules to distinguish minimally invasive adenocarcinomas and invasive adenocarcinomas in CT scan. *Front Oncol* (2022) 11:691112. doi: 10.3389/fonc.2021.691112

38. Li X, Guindani M, Ng CS, Hobbs BP. Spatial Bayesian modeling of GLCM with application to malignant lesion characterization. *J Appl Stat* (2018) 46(2):230–46. doi: 10.1080/02664763.2018.1473348

39. Granata V, Fusco R, De Muzio F, Cutolo C, Setola SV, Grassi R, et al. Radiomics textural features by MR imaging to assess clinical outcomes following liver resection in colorectal liver metastases. *Radiol Med* (2022) 127(5):461–70. doi: 10.1007/s11547-022-01477-6

40. Wang T, Deng J, She Y, Zhang L, Wang B, Ren Y, et al. Radiomics signature predicts the recurrence-free survival in stage I non-small cell lung cancer. *Ann Thorac Surg* (2020) 109(6):1741–9. doi: 10.1016/j.athoracsur.2020.01.010

41. Yang L, Yang J, Zhou X, Huang L, Zhao W, Wang T, et al. Development of a radiomics nomogram based on the 2D and 3D CT features to predict the survival of non-small cell lung cancer patients. *Eur Radiol* (2019) 29(5):2196–206. doi: 10.1007/s00330-018-5770-y

42. Wang X, Duan H, Li X, Ye X, Huang G, Nie S. A prognostic analysis method for non-small cell lung cancer based on the computed tomography radiomics. *Phys Med Biol* (2020) 65(4):045006. doi: 10.1088/1361-6560/ab6e51

43. Artal Cortés Á, Calera Urquizu L, Hernando Cubero J. Adjuvant chemotherapy in non-small cell lung cancer: state-of-the-art. *Transl Lung Cancer Res* (2015) 4(2):191–7. doi: 10.3978/j.issn.2218-6751.2014.06.01

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 19 May 2023
DOI 10.3389/fonc.2023.1103169

Check for updates

OPEN ACCESS

EDITED BY
Miao Liu,
Harvard Medical School, United States

REVIEWED BY
Xu Chen,
Shandong University, China
Qing Li,
Jackson Laboratory, United States
Xianwei Ye,
Guizhou Provincial People's Hospital, China

*CORRESPONDENCE
Weiping Xie
wpxie@njmu.edu.cn
Hui Kong
konghui@njmu.edu.cn

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

RECEIVED 20 November 2022
ACCEPTED 17 April 2023
PUBLISHED 19 May 2023

CITATION
Zhang Q, Dai Y, Jin L, Shi S, Liu C, Rong R, Sun W, Dai S, Kong H and Xie W (2023) Clinicopathological characteristics and cancer-specific prognosis of primary pulmonary lymphoepithelioma-like carcinoma: a population study of the US SEER database and a Chinese hospital. *Front. Oncol.* 13:1103169. doi: 10.3389/fonc.2023.1103169

COPYRIGHT
© 2023 Zhang, Dai, Jin, Shi, Liu, Rong, Sun, Dai, Kong and Xie. 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.

Clinicopathological characteristics and cancer-specific prognosis of primary pulmonary lymphoepithelioma-like carcinoma: a population study of the US SEER database and a Chinese hospital

Qun Zhang[1†], Yuan Dai[1†], Linling Jin[1†], Shuangshuang Shi[1], Chang Liu[1], Rong Rong[2], Wenkui Sun[1], Shanlin Dai[1], Hui Kong[1*] and Weiping Xie[1*]

[1]Department of Respiratory and Critical Care Medicine, The First Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu, China, [2]Department of Pathology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China

Introduction: Primary pulmonary lymphoepithelioma-like carcinoma (PPLELC) is a rare histological type of non-small cell lung cancer (NSCLC), which accounts for less than 1% of NSCLC. Currently, there is no well-recognized treatment guideline for PPLELC.

Methods: We identified PPLELC patients from the Surveillance, Epidemiology, and End Results (SEER) dataset between 2000 and 2015 (n = 72) as well as from our medical center between 2014 and 2020 (n = 16). All diagnoses were confirmed by pathological testing, and the clinicopathological characteristics of patients were retrieved and summarized. Survival analyses were conducted using the Kaplan–Meier analysis and log-rank tests. Multivariate survival analysis was performed with the Cox regression hazards model.

Results: The median age at diagnosis of the PPLELC cohort was 64 years, ranging from 15 to 86 years. The percentages of patients with TNM stages I, II, III, and IV were 52.3%, 10.2%, 20.5%, and 17.0%, respectively. Among the 88 cases, lesion resection was performed in 69 cases (78.4%), 16 cases (18.1%) received beam radiation, and 40 cases (45.5%) underwent chemotherapy. In the SEER dataset of lung cancer, the percentage of PPLELC in the Asian race (0.528‰) was almost 10 times higher than that in the white (0.065‰) and black (0.056‰) races. Patients with TNM stage III–IV exhibited a worse prognosis than those with TNM stage I–II (p = 0.008), with a 5-year cancer-specific survival (CSS) rate of 81.8% for TNM stage I–II and 56.2% for TNM stage III–IV. Specifically, the N stage and M stage were the leading prognostic factors, not the T stage and tumor size. Moreover, patients who underwent surgery had significantly better outcomes than those who did not (p = 0.014). Additional multivariate analysis indicated that the TNM stage was an independent prognosis factor for CSS (HR, 3.31; 95% CI, 1.08–10.14).

Conclusion: PPLELC is a rare tumor with Asian susceptibility. Although the prognosis of PPLELC is better than that of other subtypes of NSCLC, it remains unsatisfactory for advanced-stage disease. The current treatment options for PPLELC include surgical resection, chemotherapy, radiotherapy, and immune therapy. Among these options, patients with surgical resection have better survival rates in this study. However, large-scale clinical research trials will be necessary to develop effective treatment guidelines for PPLELC.

KEYWORDS

primary pulmonary lymphoepithelioma-like carcinoma (PPLELC), cancer-specific survival, incidence, SEER, NSCLC

Introduction

Lung cancer accounts for the leading cancer-related death worldwide. The disease can be broadly categorized into non-small-cell lung cancer (NSCLC) and small-cell lung cancer (SCLC). Though subtype-specific morbidity and mortality had been extensively reported, some rare pathological types of lung cancer, such as primary pulmonary lymphoepithelioma-like carcinoma (PPLELC), have not been fully characterized in terms of incidence and prognosis (1). PPLELC is a rare histological subtype of NSCLC, and further research is needed to understand its clinical features.

PPLELC was first reported by Begin et al. in 1987 on a 40-year-old woman of Southeast Asian descent in Canada (2). This epithelial tumor, which is associated with Epstein–Barr (EB) virus infection, is histologically similar to nasopharyngeal carcinoma (NPC) (3). In 2004, the World Health Organization (WHO) initially classified PPLELC as a subtype of large cell carcinoma, but in the 2005 classification, it was re-categorized as "other and unclassified carcinomas" of NSCLC (4). Studies have indicated that PPLELC was typically diagnosed at an early stage and has a better prognosis than other NSCLC subtypes (5, 6). However, the majority of reported PPLELC cases were from Asia, and limited studies have focused on its incidence and prognosis in Western countries (7, 8).

Over the past decade, our center has treated numerous cases of PPLELC. Our study aimed to examine the clinicopathological characteristics, treatment methods, and cancer-specific prognosis of PPLELC in both the United States and our own medical center.

Patients and methods

Patient selection

We identified lung cancer patients in Surveillance, Epidemiology, and End Results (SEER) dataset between 2000 and 2015 who had been diagnosed with ICD-O-3 code as "8082/3: lymphoepithelial carcinoma". All cases were confirmed as primary tumors with positive histology and no previous history of malignancy.

Another additional retrospective cohort of 18 cases was collected from Jiangsu People's Hospital (JSPH, Nanjing, China) during 2014–2020. All diagnoses in this cohort were based on pathological testing. Two cases were excluded from this cohort due to a history of gastric cancer at the time of their PPLELC diagnoses.

DNA sequencing

The DNA extraction kit (Kai Shuo) was used to extract genomic DNA (gDNA) from the tissue, following the manufacturer's instructions. Library construction was carried out using a probe hybridization capture method, with commercial reagents and customized probes. Fragmentation enzymes were employed to shear 15–200 ng of gDNA into 200–350 bp. The constructed DNA libraries were loaded on the NovaSeq 6000 platform (Illumina, San Diego, CA, USA), and the sequences were generated as 150-bp paired-end reads. The resulting base calls were converted to FASTQ files. Adapter trimming and filtering of low-quality bases were performed using the fastp tool (v.2.20.0), and duplicate reads from PCR were eliminated using Dedup with Error Correct. Variants were filtered against common single-nucleotide polymorphisms (SNPs) in public databases, while copy number variations (CNVs) and fusions were conducted with CNVkit (dx1.1) and factera (v1.4.4), respectively (9, 10).

Statistics

SPSS Version 24.0 software was used for data analysis. The Kaplan–Meier method and log-rank test were used for univariate survival analyses. The Cox regression hazards model was used for multivariate survival analysis. SPSS Version 23.0 software was applied for statistical analysis. $p < 0.05$ was considered statistically significant.

Ethics

This study was approved by the Institutional Review Board of The Ethic Committee of Nanjing Medical University First Affiliated Hospital.

Results

Incidence of PPLELC in the SEER dataset

As shown in Figure 1, there were only two to eight cases diagnosed as PPLELC in the SEER dataset every year, accounting for less than 0.01% of all lung cancers. This highlighted the rarity of the PPLELC subtype. According to the SEER dataset, there were 816,436 cases of diagnosed lung cancer during 2000–2015. Among them, 0.065‰ (44/679,436) were white patients diagnosed with PPLELC, 0.056‰ (5/88,611) were black patients, and 0.528‰ (23/43,571) cases were Asian–Pacific Islanders (APIs) (Table 1). This indicated a higher proportion of the PPLELC subtype in lung cancer among APIs.

Clinical and demographic characteristics of PPLELC cases

A total of 72 patients in the SEER dataset and 16 cases in the JSPH cohort were included in our study. All cases were diagnosed with pathological tests, and their basic characteristics are listed in Table 2. The median age of diagnosis for the two PPLELC cohorts was 64 years, ranging from 15 to 86 years. The incidence was similar between women (48.9%) and men (51.1%). More cases exhibited right lung laterality (47/88, 53.4%) than left lung (39/88, 44.3%). All 88 cases were pathologically classified as poorly differentiated or undifferentiated, except for two that were moderately differentiated.

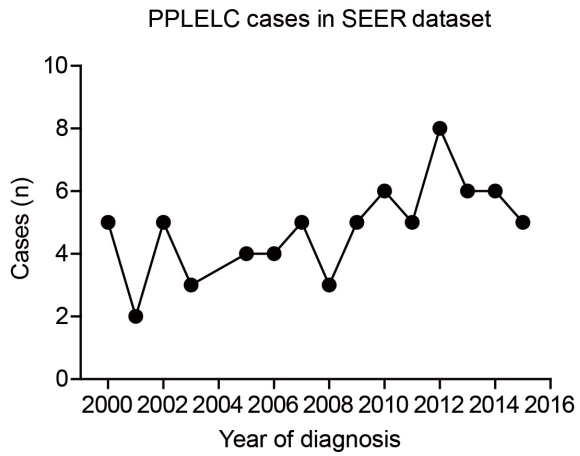


FIGURE 1
Case numbers of PPLELC in SEER dataset according to diagnostic year. PPLELC, primary pulmonary lymphoepithelioma-like carcinoma; SEER, Surveillance, Epidemiology, and End Results.

More than half of the cases were diagnosed as TX/T1 stage (55.7%), while the percentages were 26.1%, 10.2%, and 8.0% for the T2, T3, and T4 stages, respectively. Most cases (54/88, 61.4%) exhibited negative lymph node metastasis, while the other 34 cases (38.6%) showed positive lymph node metastases. Among the 34 cases above, 11 cases were classified as N1 stage, 21 cases as N2 stage, and 2 cases as N3 stage. The distant metastasis rate was 17.0% in the PPLELC cohort, as reflected by the M1 stage. Accordingly, the percentages of patients with TNM stage I, II, III, and IV were 55.7%, 10.2%, 20.5%, and 17.0%, respectively. As for the disease treatment, we retrieved the therapeutic information including surgery resection, radiotherapy, and chemotherapy. Among the 88 cases, 69 cases (78.4%) underwent lesion resection, 16 cases (18.1%) received beam radiation, and 40 cases (45.5%) accepted chemotherapy.

Among the JSPH cohort, only two cases (2/16, 12.5%) were diagnosed at the age of 65 or older, which was significantly different compared with the SEER dataset (37/72, 51.4%). Interestingly, 43.8% (7/16) of cases showed tumor location within the middle lobe of the right lung. In contrast, only 6.9% (5/72) of SEER cases showed this tumor location, and all of them were APIs. This difference seems to be related to race since none of the white or black patients exhibited right middle lobe lung tumor location. However, further validation will be necessary due to the limited number of cases in this study. Another difference was the smaller tumor size in our cohort. In the JSPH cohort, 62.5% (10/16) of cases were diagnosed with tumor size less than or equal to 2.5 cm in diameter, while it was only 38.9% (28/72) in the SEER cohort. Considering the different diagnostic eras in the SEER cohort (2000–2015) and JSPH cohort (2014–2020), this may be caused by more sensitive CT scans in the recent decade rather than racial differences. Although the percentages of M1 stage cases were comparable in the SEER cohort (13/72, 18.1%) and JSPH cohort (2/16, 12.5%), the treatment selection seemed different. There were more cases that underwent surgery resection (87.5% *vs.* 75.0%) and fewer cases accepted radiotherapy (6.3% *vs.* 20.8%) in the JSPH cohort (Table 2).

Survival analysis of CSS in PPLELC cases

The median follow-up time was 51 months, ranging from 1 to 226 months. By the end of the follow-up period, there were 24 cases (24/88, 27.3%) dead due to PPLELC (Supplementary Table 1). The 1-, 3-, and 5-year cancer-specific survival (CSS) rates were 90.8%, 79.5%, and 72.6%, respectively (Figure 2A). The mean CSS time was 148.7 ± 12.6 months. Figure 2A shows a significant difference between the OS and CSS of PPLELC (p = 0.02).

We next conducted Kaplan–Meier survival analyses by dividing patients into different subgroups according to clinical and demographic characteristics (Figures 2B–N). The M stage of patients significantly affected the CSS of PPLELC (p = 0.001, Figure 2B). The 5-year CSS rate of patients with M0 stage was 78.5%, while it decreased to 40.7% for patients with M1 stage. The mean CSS time of patients with M0 stage was 166.5 ± 12.8 months, while it decreased to 54.4 ± 11.4 months for patients with M1 stage (Table 3). Additionally, the N stage of patients also affected

TABLE 1 PPLELC proportion in lung cancer of different races in SEER dataset.

Races	PPLELC	Lung cancer (LC)	Percentage (PPLELC/LC)
White	44	679,436	0.065‰
Black	5	88,611	0.056‰
Asian–Pacific Islander	23	43,571	0.528‰

PPLELC, primary pulmonary lymphoepithelioma-like carcinoma; SEER, Surveillance, Epidemiology, and End Results.

prognosis with a statistically significant difference (p = 0.032, Figure 2C). The 5-year CSS rate of patients with negative lymph nodes was 80.7%, while it decreased to 59.4% for patients with positive lymph node metastases (Table 3). Consistently, patients with TNM stage III–IV exhibited worse prognoses than those in TNM stage I–II (p = 0.008, Figure 2D). The 5-year CSS rate of patients in TNM stage I–II was 81.8%, while it decreased to 56.2% for patients in TNM stage III–IV. The mean CSS time of cases in

TABLE 2 Clinical and demographic characteristics of PPLELC cases.

Variables	Total no. (%)	JSPH no. (%)	SEER no. (%)
Total	88 (100%)	16 (100%)	72 (100%)
Age (years)			
Median (range)	64 (15–86) years	55 (39–74) years	66 (15–86) years
≤65 years	49 (55.7%)	14 (87.5%)	35 (48.6%)
>65 years	39 (44.3%)	2 (12.5%)	37 (51.4%)
Sex			
Female	43 (48.9%)	8 (50.0%)	35 (48.6%)
Male	45 (51.1%)	8 (50.0%)	37 (51.4%)
Race			
White	44 (50.0%)	0 (0%)	44 (61.1%)
Black	5 (5.7%)	0 (0%)	5 (6.9%)
Asian–Pacific Islander	39 (44.3%)	16 (100%)	23 (31.9%)
Primary site			
Upper lobe	40 (45.5%)	3 (18.8%)	37 (51.4%)
Lower lobe	29 (33.0%)	6 (37.5%)	23 (31.9%)
Others[#]	19 (21.6%)	7 (43.8%)	12 (16.7%)
Laterality			
Left lung	39 (44.3%)	7 (43.8%)	32 (44.4%)
Right lung	47 (53.4%)	9 (56.3%)	38 (52.8%)
Paired laterality	2 (2.3%)	0 (0%)	2 (2.8%)
Tumor size (cm)			
Median (range)	2.4 (0.5–9.5)	2.2 (0.8–5.6)	2.5 (0.5–9.5)
≤2.5 cm	38 (43.2%)	10 (62.5%)	28 (38.9%)
>2.5 cm	28 (31.8%)	4 (25.0%)	24 (33.3%)
Unknown	22 (25.0%)	2 (12.5%)	20 (27.8%)
T stage			
TX/T1	49 (55.7%)	12 (75.0%)	37 (51.4%)
T2	23 (26.1%)	1 (6.3%)	22 (30.6%)

(Continued)

TABLE 2 Continued

Variables	Total no. (%)	JSPH no. (%)	SEER no. (%)
T3	9 (10.2%)	2 (12.5%)	7 (9.7%)
T4	7 (8.0%)	1 (6.3%)	6 (8.3%)
N stage			
N0	54 (61.4%)	11 (68.8%)	43 (59.7%)
N1	11 (12.5%)	1 (6.3%)	10 (13.9%)
N2	21 (23.9%)	4 (25.0%)	17 (23.6%)
N3	2 (2.3%)	0 (0%)	2 (2.8%)
M stage			
M0	73 (83.0%)	14 (87.5%)	59 (81.9%)
M1	15 (17.0%)	2 (12.5%)	13 (18.1%)
TNM stage*			
I	46 (55.7%)	11 (68.8%)	35 (48.6%)
II	9 (10.2%)	0 (0%)	9 (12.5%)
III	18 (20.5%)	3 (18.8%)	15 (20.8%)
IV	15 (17.0%)	2 (12.5%)	13 (18.1%)
Surgery resection			
Yes	69 (78.4%)	14 (87.5%)	54 (75.0%)
No/unknown	20 (22.7%)	2 (12.5%)	18 (25.0%)
Radiotherapy			
Beam radiation	16 (18.1%)	1 (6.3%)	15 (20.8%)
No/unknown	72 (81.8%)	15 (93.8%)	57 (79.2%)
Chemotherapy			
Yes	40 (45.5%)	10 (62.5%)	30 (41.7%)
No/unknown	48 (54.5%)	6 (37.5%)	42 (58.3%)

PPLELC, primary pulmonary lymphoepithelioma-like carcinoma; SEER, Surveillance, Epidemiology, and End Results; JSPH, Jiangsu People's Hospital.
[#] Included 2 main bronchus localization, 5 middle lobe localization, 1 overlapping localization, and 3 unspecified localization.
* $p < 0.05$ by log-rank test.

TNM stage I–II was 176.3 ± 13.5 months, while it decreased to 80.2 ± 10.5 months for patients in TNM stage III–IV (Table 3). Patients who underwent surgical resection showed a 5-year CSS rate of 78.8%, while it was only 48.7% in those without surgical intervention (p = 0.0014, Figure 2E). The mean CSS time of patients who underwent surgical resection was 164.1 ± 13.9 months, while it decreased to 72.6 ± 13.2 months for patients without surgical treatment (Table 3). In contrast, other factors seemed to have no significant effect on PPLELC survival (Figures 2F–N, all $p > 0.05$).

Multivariate analysis was performed using a Cox regression hazards model. The variables with $p < 0.3$ from univariate analyses, including patients' age, sex, TNM stage, surgery resection, radiotherapy, and chemotherapy (Table 3), were subjected to this model. We excluded N and M stages during multivariate analysis due to their inseparable correlation with the TNM stage. As a result, the TNM stage was the only factor that had an independent prognostic effect on PPLELC survival (p = 0.036, HR, 3.31; 95% CI, 1.08–10.14). The hazard ratio (95% confidence interval) for age > 65, sex male, no surgery treatment, no radiotherapy, and no chemotherapy was 2.13 (0.94–4.83), 0.54 (0.23–1.27), 1.41 (0.44–4.59), 0.99 (0.33–3.02), and 1.71 (0.47–6.16), respectively.

Discussion

PPLELC is a rare subtype of unclassified NSCLC, and the presence of the EB virus in tumor cells is critical for disease diagnosis. The clinical characteristics, prognostic factors, and therapies of PPLELC remain unclear. The major reports on PPLELC were primarily from Eastern countries, such as Southern China and Taiwan (8). Our study of the US SEER dataset also showed a high proportion of the PPLELC subtype among Asians with lung cancer, indicating a prevalence of PPLELC among East

10.3389/fonc.2023.1103169

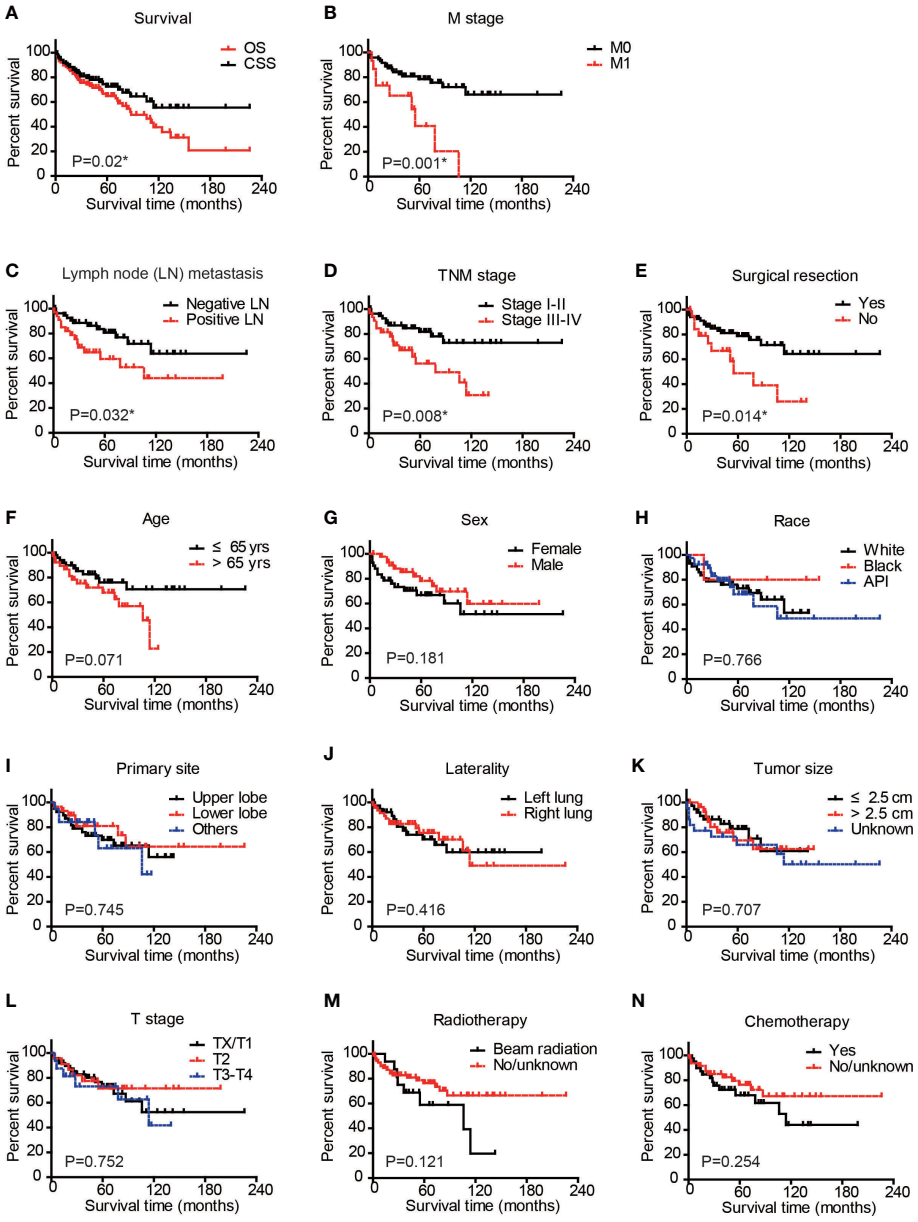


FIGURE 2
Cancer-specific survival analysis of PPLELC patients. Cancer-specific survival (CSS) and overall survival (OS) of entire cohort **(A)** CSS of patients based on M stage **(B)**, Lymph node metastasis **(C)**, TNM stage **(D)**, surgical resection **(E)**, age **(F)**, sex **(G)**, race **(H)**, primary tumor site **(I)**, tumor laterality **(J)**, tumor size **(K)**, T stage **(L)**, radiotherapy **(M)**, or chemotherapy **(N)**. Data were analyzed by Kaplan–Meier method and log-rank test. * $p < 0.05$. PPLELC, primary pulmonary lymphoepithelioma-like carcinoma.

Asians. It is similar to another EB-associated disease, NPC, which occurs frequently in Southeast Asian countries (11).

Previous research has reported that PPLELC showed a better prognosis than other common subtypes of NSCLC (7). For instance, Jiang et al. reported a 5-year disease-free survival (DFS) rate of 47% from their medical center (12). Moreover, He et al. indicated that the median overall survival time was 107 months using the SEER dataset of 62 PPLELC cases during 1973–2011, and univariate analysis demonstrated that only age > 65 was associated with poor prognosis (13). However, non-lung cancer-related mortality hazard contributed significantly to the overall survival of long-term cancer survivors, such as PPLELC patients (14). Supplementary Table 1 indicates almost 38.5% of PPLELC death attributed to causes other than lung cancer. Therefore, CSS would have more advantages in predicting the prognosis of PPLELC than overall survival. In this study, CSS was the main outcome used to evaluate the prognosis of the SEER and JSPH cohorts. Our study suggested a 5-year CSS rate of 81.8% for stage I–II cases and 56.2% for stage III–IV patients. In contrast, the 5-year survival rates of lung adenocarcinoma were 14.0% and 3.7% for stage III and IV cases, respectively (14). In addition, the SEER and JSPH datasets presented that the TNM stage served as a significant independent prognostic factor for CSS of PPLELC in this study. Consistently, a previous meta-analysis revealed that early-stage diagnosis was a

10.3389/fonc.2023.1103169

TABLE 3 Cancer-specific survival (CSS) of PPLELC cases.

Variables	Case (n)	5-year CSS (%)	Mean CSS (months)	p-Value[a]	HR	95% CI	p-Value[b]
Age (years)				0.071			0.070
≤65 years	49	76.0	170.83 ± 14.6		Reference		
>65 years	39	67.9	81.77 ± 8.0		2.13	0.94–4.83	
Sex				0.181			0.154
Female	43	66.8	137.4 ± 18.3		Reference		
Male	45	78.2	143.5 ± 13.9		0.54	0.23–1.27	
Race				0.766			
White	44	73.0	101.0 ± 9.0		–		
Black	5	80.0	128.0 ± 24.2				
Asian–Pacific Islander	39	68.4	138.6 ± 21.2				
Primary site				0.745			
Upper lobe	40	69.9	101.0 ± 9.4		–		
Lower lobe	29	81.0	163.0 ± 20.3				
Others	19	63.2	83.9 ± 11.0				
Laterality[#]				0.416			
Left lung	39	70.2	135.3 ± 14.6		–		
Right lung	47	75.5	144.1 ± 19.2				
Tumor size (cm)				0.707			
≤2.5 cm	38	78.8	106.5 ± 10.1		–		
>2.5 cm	28	69.4	108.0 ± 11.8				
Unknown	22	65.9	136.9 ± 22.2				
T stage				0.752			
TX/T1	49	72.8	144.6 ± 17.9		–		
T2	23	71.5	149.2 ± 17.0				
T3–T4	16	73.1	94.2 ± 14.4				
N stage				0.032*			
N0	54	80.7	165.9 ± 15.3		–		
N1–N3	34	59.4	110.7 ± 17.0				
M stage				0.001**			
M0	73	78.5	166.5 ± 12.8		–		
M1	15	40.7	54.4 ± 11.4				
TNM stage				0.008**			0.036*
I–II	55	81.8	176.3 ± 13.5		Reference		
III–IV	33	56.2	80.2 ± 10.5		3.31	1.08–10.14	
Surgery resection				0.014*			0.414
Yes	68	78.8	164.1 ± 13.9		Reference		
No/unknown	20	48.7	72.6 ± 13.2		1.412	0.44–4.59	

(Continued)

TABLE 3 Continued

Variables	Case (n)	5-year CSS (%)	Mean CSS (months)	p-Value[a]	HR	95% CI	p-Value[b]
Radiotherapy				0.121			0.987
Beam radiation	16	58.9	89.0 ± 13.0		Reference		
No/unknown	72	76.4	154.1 ± 13.8		0.99	0.33–3.02	
Chemotherapy				0.254			0.566
Yes	40	67.9	119.1 ± 15.8		Reference		
No/unknown	48	76.4	165.4 ± 15.0		1.71	0.47–6.16	

PPLELC, primary pulmonary lymphoepithelioma-like carcinoma.
[#] Two cases were recorded with paired laterality and were not shown here due to limited case numbers.
[a]Data significance was compared by log-rank test.
[b]Data significance was compared by Cox regression test.
* $p < 0.05$, ** $p < 0.01$.

favorable prognosis factor for both overall survival and DFS. Moreover, their systematic review concluded that a better outcome of PPLELC was detected in men and patients who underwent radiotherapy (15).

For NSCLC patients in the early stages, surgery is the only curative treatment. This study confirmed that patients with PPLELC who received surgery had a better prognosis than those who did not. However, most PPLELC cases have no specific clinical manifestations at diagnosis, resulting in a common diagnosis of advanced tumor stages and a missed opportunity for curative resection. Literature reviews suggested effective chemotherapy and radiotherapy for late-stage disease (15, 16). However, our research revealed that chemotherapy and radiotherapy were not correlated with the CSS of PPLELC. Moreover, previous studies have implicated that advanced PPLELC patients benefited little from targeted therapy (17). Unlike classical NSCLC, rare driver genes that were commonly mutated (e.g., EGFR) were detected in PPLELC according to recent reports, which was consistent with our study (18). Eleven tumor samples were tested for gene mutations by NSG, and only three of them carried mutated genes: one had a CYP2D6*10 homozygous mutation T/T and a UGT1A1*6 heterozygous mutation G/A, one had a PAK3 mutation, and one had a TP53 mutation (Supplementary Table 2). Ying et al. reported that TP53 (43%) and CYLD (43%) were the two most commonly mutated genes in PPLELC (18), while the frequency of the TP53 mutation rate was revealed to be 6.5% in another cohort (19). The discrepancy may be attributed to the diverse detection methods. Nevertheless, a combination of anti-angiogenic therapy with chemotherapy was reported to be superior to chemotherapy alone in PPLELC of advanced stages (20). According to Bao et al., the median progression-free survival (PFS) was 11.2 months in the former group and only 7.0 months in the latter group ($p = 0.008$). Consistently, the 1-year PFS rates were 41.9% and 17.6%, respectively (20).

The therapeutic efficiency of immune therapy had also been recently reported in several studies. According to a meta-analysis by Tang et al., which included 13 retrospective studies of 1,294 PPLELC cases, positive PD-L1 was observed in 63%–76% of cases. Interestingly, high PD-L1 was more frequent in younger patients ($p = 0.01$) and was correlated with poor DFS ($p = 0.02$) (15). A subsequent study indicated the median PD-L1 expression as 40% in a retrospective cohort from a single medical center (n = 5) (21). Consistently, Fan et al. reported 42.9% (3/7) of PPLELC cases exhibited high PD-L1 expression (18), highlighting that PD-L1 expression was observed in a large proportion of PPLELC cases. Of note, a high proportion of PD-L1 in PPLELC implied immunotherapy as a potential direction in disease treatment. For example, in Wu et al.'s cohort, five PPLELC patients at advanced stages who failed with surgery and chemo-radiotherapy were subjected to immune checkpoint blockade therapy. Three of the patients (3/5, 60%) responded favorably, with the best overall response being partial remission (21). Another case report from Thailand described a patient who received pembrolizumab, a PD-1 inhibitor, with the best response being a stable disease, although the patient died 28 months after diagnosis (22). In the JSPH cohort, one PPLELC patient received chemotherapy combined with immunotherapy and had a stable disease response for 21 months at the end of the follow-up. Nevertheless, prospective clinical studies are necessary to obtain more evidence for PPLELC immunological treatment.

In conclusion, our study described the clinicopathological features, current treatment, and prognosis of PPLELC patients in both the SEER dataset and our medical center. PPLELC is a rare subtype of NSCLC with a higher incidence among Asians and a favorable prognosis. Consistent with other NSCLC subtypes, the TNM stage serves as an independent prognostic factor of PPLELC.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by Institutional Review Board of The Ethic Committee of Nanjing Medical University First Affiliated Hospital. The patients/

participants provided their written informed consent to participate in this study.

Author contributions

QZ and YD contributed to data curation and writing preparation. LJ, SS, and CL contributed to methodology and formal analysis. RR, WS, and SD contributed to software application and resources provision. WX and HK contributed to project administration and supervision. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the National Natural Science Foundation of China (82002438), the Fellowship of China Postdoctoral Science Foundation (2020M671396), and the Fellowship of Jiangsu Postdoctoral Science Foundation (2020Z239).

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.2023.1103169/full#supplementary-material

References

1. Sathirareuangchai S, Hirata K. Pulmonary lymphoepithelioma-like carcinoma. *Arch Pathol Lab Med* (2019) 143(8):1027–30. doi: 10.5858/arpa.2018-0149-RS

2. Bégin LR, Eskandari J, Joncas J, Panasci L. Epstein-Barr Virus related lymphoepithelioma-like carcinoma of lung. *J Surg Oncol* (1987) 36(4):280–3. doi: 10.1002/jso.2930360413

3. Iezzoni JC, Gaffey MJ, Weiss LM. The role of Epstein-Barr virus in lymphoepithelioma-like carcinomas. *Am J Clin patho* (1995) 103(3):308–15. doi: 10.1093/ajcp/103.3.308

4. William D, Travis E, Allen P, Burke A, Nicholson AG. *WHO classification of tumors of the lung, pleura, thymus and heart*. Lyon: International Agency for Research on Cancer (IARC (2015).

5. Liang Y, Wang L, Zhu Y, Lin Y, Liu H, Rao H, et al. Primary pulmonary lymphoepithelioma-like carcinoma: fifty-two patients with long-term follow-up. *Cancer* (2012) 118(19):4748–58. doi: 10.1002/cncr.27452

6. Sun YH, Lin SW, Hsieh CC, Yeh YC, Tu CC, Chen KJ. Treatment outcomes of patients with different subtypes of large cell carcinoma of the lung. *Ann Thorac surg* (2014) 98(3):1013–9. doi: 10.1016/j.athoracsur.2014.05.012

7. Han AJ, Xiong M, Gu YY, Lin SX, Xiong M. Lymphoepithelioma-like carcinoma of the lung with a better prognosis. a clinicopathologic study of 32 cases. *Am J Clin pathol* (2001) 115(6):841–50. doi: 10.1309/BUAN-BGFW-69U9-C3H8

8. Ho JC, Wong MP, Lam WK. Lymphoepithelioma-like carcinoma of the lung. *Respirology (Carlton Vic).* (2006) 11(5):539–45. doi: 10.1111/j.1440-1843.2006.00910.x

9. Chen HB, Gong XY, Li W, Chen DS, Zhao LL, Li S, et al. Genomic landscape and tumor mutation burden analysis of Chinese patients with sarcomatoid carcinoma of the head and neck. *Oral Oncol* (2021) 121:105436. doi: 10.1016/j.oraloncology.2021.105436

10. Guan Y, Wang Y, Li H, Meng J, You X, Zhu X, et al. Molecular and clinicopathological characteristics of ERBB2 gene fusions in 32,131 Chinese patients with solid tumors. *Front Oncol* (2022) 12:986674. doi: 10.3389/fonc.2022.986674

11. Guan S, Wei J, Huang L, Wu L. Chemotherapy and chemo-resistance in nasopharyngeal carcinoma. *Eur J Med Chem* (2020) 207:112758. doi: 10.1016/j.ejmech.2020.112758

12. Jiang WY, Wang R, Pan XF, Shen YZ, Chen TX, Yang YH, et al. Clinicopathological features and prognosis of primary pulmonary lymphoepithelioma-like carcinoma. *J Thorac dis* (2016) 8(9):2610–6. doi: 10.21037/jtd.2016.08.40

13. He J, Shen J, Pan H, Huang J, Liang W, He J. Pulmonary lymphoepithelioma-like carcinoma: a surveillance, epidemiology, and end results database analysis. *J Thorac dis* (2015) 7(12):2330–8. doi: 10.3978/j.issn.2072-1439.2015.12.62

14. Zhang Q, Dai Y, Liu H, Sun W, Huang Y, Gong Z, et al. Causes of death and conditional survival estimates of long-term lung cancer survivors. *Front Immunol* (2022) 13:1012247. doi: 10.3389/fimmu.2022.1012247

15. Tang L, Chen N, He W, Zhou J, Zhang J, Lin Z, et al. The clinicopathological features and prognosis of primary pulmonary lymphoepithelioma-like carcinoma: a systematic review and meta-analysis. *PloS One* (2020) 15(10):e0240729. doi: 10.1371/journal.pone.0240729

16. Zhang L, Hao T, Wei Y, Dong M, Xiong Y. Primary pulmonary lymphoepithelioma-like carcinoma: a case report of pathological complete response (pCR) by neoadjuvant treatment. *Medicine* (2021) 100(11):e24987. doi: 10.1097/MD.0000000000024987

17. Liu Q, Ma G, Yang H, Wen J, Li M, Yang H, et al. Lack of epidermal growth factor receptor gene mutations in exons 19 and 21 in primary lymphoepithelioma-like carcinoma of the lung. *Thorac Cancer* (2014) 5(1):63–7. doi: 10.1111/1759-7714.12060

18. Fan Y, Shan Q, Gong J, Qin J, Lu H. Molecular and clinical characteristics of primary pulmonary lymphoepithelioma-like carcinoma. *Front Mol Biosc* (2021) 8. doi: 10.3389/fmolb.2021.736940

19. Chang YL, Wu CT, Shih JY, Lee YC. Unique p53 and epidermal growth factor receptor gene mutation status in 46 pulmonary lymphoepithelioma-like carcinomas. *Cancer Sci* (2011) 102(1):282–7. doi: 10.1111/j.1349-7006.2010.01768.x

20. Bao H, Ma LZ, Zhao C, Yu M, Zhang B, Zhang J, et al. Anti-angiogenic therapy for advanced primary pulmonary lymphoepithelioma-like carcinoma: a retrospective multicenter study. *J Cancer Res Clin Oncol* (2023). 149(3):1185-93 doi: 10.1007/s00432-022-03935-0

21. Wu Z, Xian X, Wang K, Cheng D, Li W, Chen B. Immune checkpoint blockade therapy may be a feasible option for primary pulmonary lymphoepithelioma-like carcinoma. *Front Oncol* (2021) 11:626566. doi: 10.3389/fonc.2021.626566

22. Archwamety A, Ruangchira-Urai R, Akewanlop C, Korphaisarn K. Primary pulmonary lymphoepithelioma-like carcinoma treated with immunotherapy: a case report and literature review. *Thoracic Cancer* (2022) 13(17):2539–41. doi: 10.1111/1759-7714.14580

frontiers | Frontiers in Oncology

TYPE Brief Research Report
PUBLISHED 02 June 2023
DOI 10.3389/fonc.2023.1204435

Check for updates

OPEN ACCESS

EDITED BY
Qingqing Zhu,
The First Affiliated Hospital of Soochow University, China

REVIEWED BY
Xi Yang,
Fudan University, China
Marion Tonneau,
University of Montreal Hospital Centre (CRCHUM), Canada

*CORRESPONDENCE
Mantang Qiu
qiumantang@163.com
Nan Li
rainbow6283@sina.com
Zhi Yang
pekyz@163.com

RECEIVED 12 April 2023
ACCEPTED 17 May 2023
PUBLISHED 02 June 2023

CITATION
Ding X, Lin G, Wang P, Chen H, Li N, Yang Z and Qiu M (2023) Diagnosis of primary lung cancer and benign pulmonary nodules: a comparison of the breath test and 18F-FDG PET-CT. *Front. Oncol.* 13:1204435. doi: 10.3389/fonc.2023.1204435

COPYRIGHT
© 2023 Ding, Lin, Wang, Chen, Li, Yang and Qiu. 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.

Diagnosis of primary lung cancer and benign pulmonary nodules: a comparison of the breath test and 18F-FDG PET-CT

Xiangxiang Ding[1,2], Guihu Lin[3], Peiyu Wang[4,5], Haibin Chen[6], Nan Li[1,2]*, Zhi Yang[1,2]* and Mantang Qiu[4,5,6]*

[1]Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Department of Nuclear Medicine, Peking University Cancer Hospital & Institute, Beijing, China, [2]Key Laboratory for Research and Evaluation of Radiopharmaceuticals (National Medical Products Administration), Department of Nuclear Medicine, Peking University Cancer Hospital & Institute, Beijing, China, [3]Department of Thoracic Surgery, Aerospace 731 Hospital, Beijing, China, [4]Department of Thoracic Surgery, Peking University People's Hospital, Beijing, China, [5]Thoracic Oncology Institute, Peking University People's Hospital, Beijing, China, [6]Breax Laboratory, PCAB Research Center of Breath and Metabolism, Beijing, China

With the application of low-dose computed tomography in lung cancer screening, pulmonary nodules have become increasingly detected. Accurate discrimination between primary lung cancer and benign nodules poses a significant clinical challenge. This study aimed to investigate the viability of exhaled breath as a diagnostic tool for pulmonary nodules and compare the breath test with 18F-fluorodeoxyglucose (18F-FDG) positron emission tomography (PET)–computed tomography (CT). Exhaled breath was collected by Tedlar bags and analyzed by high-pressure photon ionization time-of-flight mass spectrometry (HPPI-TOFMS). A retrospective cohort (n = 100) and a prospective cohort (n = 63) of patients with pulmonary nodules were established. In the validation cohort, the breath test achieved an area under the receiver operating characteristic curve (AUC) of 0.872 (95% CI 0.760–0.983) and a combination of 16 volatile organic compounds achieved an AUC of 0.744 (95% CI 0.7586–0.901). For PET-CT, the SUVmax alone had an AUC of 0.608 (95% CI 0.433–0.784) while after combining with CT image features, 18F-FDG PET-CT had an AUC of 0.821 (95% CI 0.662–0.979). Overall, the study demonstrated the efficacy of a breath test utilizing HPPI-TOFMS for discriminating lung cancer from benign pulmonary nodules. Furthermore, the accuracy achieved by the exhaled breath test was comparable with 18F-FDG PET-CT.

KEYWORDS

lung cancer, breath test, pulmonary nodule, PET-CT, early detection

Introduction

The National Lung Screening Trial (NLST) has confirmed that low-dose computed tomography (LDCT) lung cancer screening can decrease lung cancer mortality by 20% in high-risk populations compared to X-ray (1). Since then, LDCT has been recommended for lung cancer screening and many pulmonary nodules have been detected along with lung cancer screening (2–4). Pulmonary nodules are defined as pulmonary lesions less than 3 cm in CT images. The pathology of pulmonary nodules includes malignant diseases, such as primary lung cancer, distant metastases, or rarer lymphoma, as well as benign causes, such as tuberculosis, pneumonia, fungi infections, and primary benign tumors (hamartoma, angioma, etc.) (5). Discriminating between primary lung cancer and benign nodules is a clinical challenge for radiologists, thoracic surgeons, and physicians practicing with LDCT-based lung cancer screening (6, 7).

Several clinical associations have made guidelines or recommendations to manage pulmonary nodules, such as the American College of Chest Physicians and The Fleischner Society in 2013 and 2017, respectively (3, 8). Rather than biopsy or surgery, high-risk pulmonary nodules are often recommended for positron emission tomography (PET) first, since it is non-invasive and can provide information for differential diagnosis and staging at the same time (9, 10). However, PET-CT is very expensive and not sensitive enough for small pulmonary nodules.

Human breath includes thousands of volatile organic compounds (VOCs) (11, 12), and mounting evidence has proved that testing VOCs of exhaled breath can precisely detect lung cancer (13–15). Cancer cells have altered metabolism and generate a variety of aberrant metabolites, and some of these aberrant metabolites could be exhaled outside and detected by mass spectrometry or nano-sensors. Breath test is totally non-invasive and easy to collect, which is a promising tool for lung cancer early detection and screening. It has been reported that VOCs in exhaled breath can discriminate lung cancer from benign nodules (16). Gas chromatography–mass spectrometer (GC-MS) has been considered as the gold standard of exhaled breath VOC analysis, but it is not applicable in clinical practice because it requires complex sample pretreatment and time-consuming detecting processes (17). Several researchers have tried to diagnose pulmonary nodules by breath test, and they used different methods (18) (19, 20). High-pressure photon ionization time-of-flight mass spectrometry (HPPI-TOFMS) is a direct mass spectrometry that does not require sample pretreatment and only takes 1 min to analyze one sample; thus, HPPI-TOFMS is a suitable tool for clinical application (13, 17). HPPI-TOFMS has been reported to be effective for detection of lung cancer (13), esophageal cancer (21), and tuberculosis (22), but it is unknown whether HPPI-TOFMS could discriminate lung cancer from benign pulmonary nodules.

In this study, we first performed the breath test by HPPI-TOFMS and trained a model in a retrospective cohort and then tested whether this model could discriminate lung cancer from benign pulmonary nodules and compared its diagnostic accuracy with PET-CT in a prospective cohort.

Materials and methods

Participant's recruitment and study design

This study was reported according to the Standards for Reporting Diagnostic Accuracy, and a checklist was attached (23). The prospective specimen collection, retrospective blinded evaluation design (24) was utilized. This study was approved by the Ethics Committee Board of Peking University People's Hospital (2021PHB349), and informed consent was obtained from all participants.

For the retrospective cohort, patients who received thoracic surgery or endobronchial ultrasound–guided transbronchial needle aspirate were consecutively recruited at the Department of Thoracic Surgery, Peking University People's Hospital from September 2021 to October 2021. The inclusion criteria were 1) age > 18 years, 2) pulmonary lesions <3 cm in CT images, 3) the pathological diagnoses were primary lung cancer or benign lung diseases, and 4) no history of cancer and no anticancer treatment before.

For the prospective cohort, patients with pulmonary lesions planning to receive 18F-FDG PET-CT were prospectively recruited at the Department of Nuclear Medicine, Cancer Hospital of Peking University from November 2021 to January 2022. Patients were consecutively recruited with the following criteria: 1) age > 18 years, 2) with pulmonary lesions and plan to have 18F-FDG PET-CT scanning, and 3) no history of cancer within 5 years. After PET-CT scanning, we followed up pathologic diagnoses of all eligible participants. Patients who met the following criteria were excluded: 1) the lung lesions were metastasized from other organs; 2) lung lesions were larger than 3 cm in CT scans; and 3) no pathological diagnosis. For all participants, the clinical data and demographic data were collected from medical records and questionnaires.

Exhaled breath collection

Exhaled breath samples were collected by trained investigators following the same protocol according to our previous studies (13). Exhaled breath was collected before PET-CT scanning and the morning before surgery. All participants fasted for at least 6 h before sample collection. To reduce potential confounding factors, all participants were asked not to ingest spicy food, alcohol, or coffee the night before exhaled breath collection. Disposable face masks and Tedlar bags were used to collect exhaled breath. A disposable face mask was replaced before each collection to avoid cross-contamination. Briefly, participants first gargled with pure water and then performed a single deep nasal inhalation followed by complete exhalation via their mouth into a Tedlar air bag. At both clinical centers, breath samples were collected in a fixed room and the room air was also collected before and after sample collection of participants. All air bags were delivered to lab and detected within 4 h.

High-pressure photon ionization time-of-flight mass spectrometry detection

Exhaled breath was detected by HPPI-TOFMS as previously described (13). The pressure in the HPPI ion source was set at 500 Pa, and two capillaries were arranged in the ion source. The gas-phase exhaled breath sample was directly introduced into the ionization region through a 250- μm i.d. 0.60- m- long stainless-steel capillary from the Tedlar bag. In order to eliminate condensation of exhaled VOCs and minimize possible surface adsorption, the stainless steel capillary was heated to 100°C and the HPPI ion source was heated to 60°C. The TOF signals were recorded by a 400ps time-to-digital converter rate at 25 kHz, and all the mass spectra were accumulated for 60s. Mass spectrum peaks detected by HPPI-TOFMS with m/z <500 were recorded and 32,500 features were extracted from the HPPI-TOFMS data of each exhaled breath sample.

Positron emission tomography–computed tomography imaging

18F-FDG PET-CT was performed as previously reported (25). The patients were instructed to fast for 6 h before 18F-FDG injection. The 18F was manufactured by HM-20 medical cyclotron (Sumitomo Corporation, Japan), and the 18F-FDG was administrated intravenously according to the patient's body weight (3.0–3.7 MBq/kg). Imaging was performed using a PET/CT scanner (Biograph64, SIEMENS, Erlangen, Germany) operated in 3D Flow Motion (bed entry speed 1 mm/s) from the apex of the skull to the midthigh, with a PET axial field of view of 21.6 cm. The PET images were reconstructed by the TrueX + TOF method offered by the vendor. The injected activity was 3.7 MBq/kg, and the time from injection to scan was approximately 60 ± 10 min. Diagnoses of 18F-FDG PET-CT were made by two authors independently based on SUVmax and CT images. Discernments between two authors were solved by discussion.

Statistical analyses

Machine learning models with mass spectrometry data of exhaled breath as input were constructed with the caret package (https://cran.r-project.org/web/packages/caret/). Then, the diagnostic capacity of the breath model and 16 VOCs model was evaluated with the area under the receiver operating characteristic (ROC) curve (AUC) and accuracy through the R packages pROC (https://cran.r-project.org/web/packages/pROC/index.html) in discriminating lung cancer from benign nodules.

Sensitivity, specificity, accuracy, the positive predict value, and the negative predictive value were calculated to evaluate the diagnostic performance of PET-CT and the breath test. The ROC was performed and the AUC was calculated to evaluate the diagnostic performance of PET-CT and the breath test. Baseline characteristics were analyzed with the independent t test or Fisher's exact test. Two-sided P values less than 0.05 were considered statistically significant. All statistical analyses were performed using SPSS software (version 24.0).

Results

Clinical characteristics of enrolled patients

The overall study design is shown in Figure 1. We first retrospectively selected 49 lung cancer patients and 51 benign pulmonary nodules from Peking University People's Hospital as the discovery cohort (PKUPH cohort). Then, 119 patients with pulmonary nodules and received 18F-FDG PET-CT were prospectively recruited from the Cancer Hospital of Peking University. Following this, 119 patients with pulmonary nodules who underwent 18F-FDG PET-CT were prospectively recruited from the Cancer Hospital of Peking University (CAPKU cohort) (Figure 1A, CAPKU cohort). Out of the 119 patients, 56 were excluded based on the exclusion criteria. A total of 63 patients were included as the validation set (Figure 1A, CAPKU enrollment diagram).

Detailed clinical characteristics of eligible participants are shown in Table 1. As shown, lung cancer patients in the PKUPH cohort were all at stage I, and in the CAPKU cohort 76.9% (30/39, Figure 1B), lung cancer patients were at stage I.

Accuracy of high-pressure photon ionization time-of-flight mass spectrometry

HPPI-TOFMS has been proved to be accurate for lung cancer detection in a proof-of-concept study. In this study, we first trained a model to discriminate lung cancer from benign pulmonary nodules in the PKUPH cohort. A total of 100 patients with pulmonary nodules were included. In the discovery cohort, the random forest model reached the best AUC and accuracy. By constructing a model with all features of mass spectrometry peaks, the random forest model reached 82.1% sensitivity, 92.3% specificity, 84.6% accuracy, 97.0% positive predictive value, 63.2% negative predictive value, and 0.872 AUC (95% CI 0.760–0.983) in the CAPKU cohort (Figure 1C).

With perioperative sampling, 16 VOCs have been identified as potential lung cancer–specific biomarkers (26). Using the 16 VOCs, the model reached 71.8% sensitivity, 76.9% specificity, 73.1% accuracy, 90.3% positive predictive value, 47.6% negative predictive value, and 0.744 AUC (95% CI 0.586–0.901) in the validation cohort (Figure 1C).

Accuracy of 18F-FDG positron emission tomography–computed tomography

We first analyzed the diagnostic accuracy of SUVmax in the CAPKU cohort. As shown in Figure 1C, SUVmax reached the best

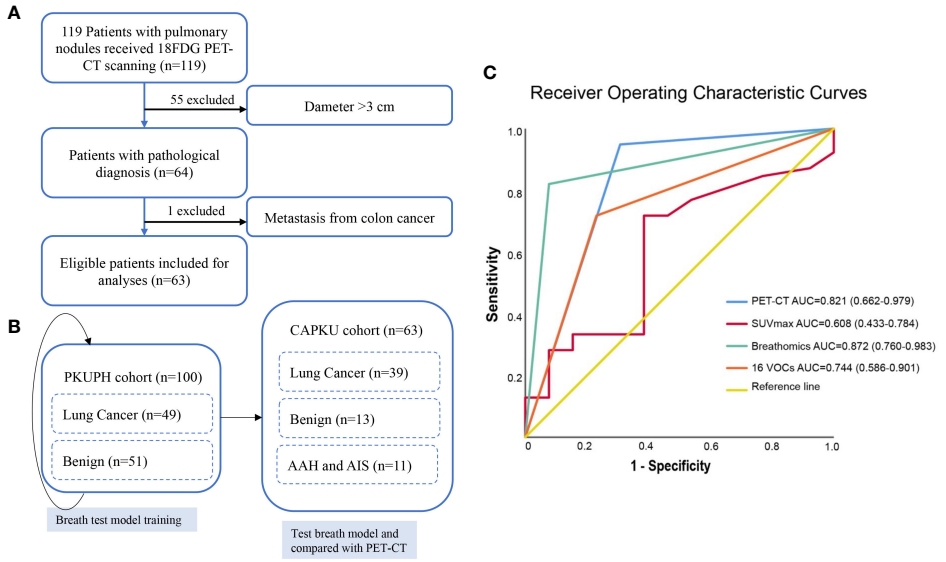


FIGURE 1
Flow diagram of participant recruitment in the Cancer Hospital of Peking University (CAPKU) cohort **(A)** and data analyses process **(B)**. Receiver operating characteristics of the breath test and positron emission tomography–computed tomography in the validation CAPKU cohort **(C)**. AAH, atypical adenomatous hyperplasia; AIS, adenocarcinoma in situ.

AUC of 0.608 (95% CI 0.433–0.784) when the threshold was set as 1.35. SUVmax achieved 71.8% sensitivity, 61.5% specificity, 69.2% accuracy, 84.8% positive predictive value, and 42.1% negative predictive value in the validation cohort.

When combined with CT features, 18F-FDG PET-CT had a better AUC (0.821, 95% CI 0.662–0.979) for discriminating lung cancer from benign nodules (Figure 1C). 18F-FDG PET-CT achieved 94.9% sensitivity, 69.2% specificity, 88.5% accuracy, 90.2% positive predictive value, 81.5% negative predictive value, and 0.744 AUC (95% CI 0.586-0.901) in the validation cohort.

Breath test and 18F-FDG positron emission tomography–computed tomography for detection of atypical adenomatous hyperplasia/adenocarcinoma *in situ*

According to the current WHO classification, AAH and AIS have been classified as benign. However, AAH and AIS are the very early stage of lung adenocarcinoma; they may slowly progress and become invasive adenocarcinoma. Thus, in the validation cohort, we compared diagnostic accuracy of breath test and 18F-FDG PET-CT among pulmonary nodules diagnosed with AAH and AIS. As shown in Table 2, the mean SUVmax of the 11 nodules were 0.95 ranging from 0.4 to 3.2, with only one patient having nodules with SUVmax > 1.35. A representative image was shown in Figure 2. Among 11 patients with AAH/AIS, 4 patients were correctly detected by the breath test and 7 patients were classified as lung cancer according to SUVmax.

Discussion

In this study, we found that SUVmax had the best diagnostic performance when the cutoff value was set as 1.35, which is much lower than the previously reported 2.5 (27, 28). It has been reported that 18F-FDG PET-CT could detect small pulmonary nodules less than 1 cm, lower cutoff may provide accurate diagnosis of benign and malignant pulmonary nodules (29). However, it should be noted that lung cancer risk among patients with PET-negative pulmonary nodules cannot be neglected. Akpoviroro et al. followed up 191 LungRADS-4 patients, and they found that 22.4% (15/67) patients were diagnosed with lung cancer in the PET-negative group (30).

In this study, SUVmax alone achieved an AUC of 0.608; however, the AUC increased to 0.821 when CT image features were added, indicating an improvement in diagnostic accuracy. Among 11 patients with AAH or AIS, 7 were correctly diagnosed as lung cancer by 18F-FDG PET-CT. As the mean SUVmax was 0.95 and only one nodule had SUVmax > 1.35, SUVmax alone is not sufficient for the detection of AIS/AAH. Furthermore, the inclusion of CT images significantly enhanced the accuracy of diagnosis in 18F-FDG PET-CT. These data suggested that SUVmax alone is not enough to discriminate lung cancer from benign pulmonary nodules and CT images should be an indispensable part of pulmonary nodule follow-up.

TABLE 1 Baseline characteristics of included participants.

	PKUPH Cohort			CHPKU Cohort		
	Lung cancer	Non-cancer lesions	*P*	Lung cancer	Non-cancer lesions*	*P*
Number of participants	49	51		39	24	
Gender			0.07			0.16
Male n (%)	20 (38.3%)	24 (48.3%)		20 (47.4%)	9 (67.9%)	
Female n (%)	29 (61.7%)	27 (51.7%)		19 (52.6%)	15 (32.1%)	
Age	54.8 ± 11.3	55.4 ± 11.3	<0.001	56.0 ± 9.8	55.9 ± 10.4	0.67
Smoking			0.963			0.188
Ever-smokers n (%)	8 (21.7%)	9.00%		16 (21.1%)	7 (39.3%)	
Never-smokers n (%)	41 (78.3%)	42 (78.5%)		23 (78.9%)	17 (60.7%)	
Pathology n (%)						
AD	49 (100%)			19 (100%)		
SCC				2		
SCLC				2		
TNM stage n (%)						
I	100 (100%)			30 (89.5%)		
II				2		
III				7 (5.3%)		

*Atypical adenomatous hyperplasia and adenocarcinoma in situ were classified as benign in this table. AD, adenocarcinoma; SCC, squamous cell carcinoma; SCLC, small cell lung cancer.

Compared with PET-CT, the breath test showed better diagnostic accuracy in the CAPKU cohort. The 16 VOCs achieved 71.8% sensitivity and 76.9% specificity, while all exhale breath mass spectrometry features achieved 82.1% sensitivity and 92.3% specificity. Although the diagnostic accuracy decreased slightly, the 16 VOCs still achieved accuracy >70%. These data suggest that the breath test may be used for diagnosis of pulmonary nodules or follow-up, especially for PET-negative pulmonary nodules.

Discriminating lung cancer from benign pulmonary nodules is very challenging, and many methods have been tried, such as circulating cell-free DNA, metabolomics, and exhaled breath. Thus, this study provides new insights into the current lung

TABLE 2 Detailed diagnoses by breath test and PET-CT among 11 participants with atypical adenomatous hyperplasia or adenocarcinoma *in situ*.

ID	Pathology	Prediction based on all Breathomics data	Prediction based on 16 VOCs	Prediction based on PET-CT	SUVmax
Patient 1	AAH	Benign	Benign	Benign	0.7
Patient 2	AIS	Benign	Lung cancer	Lung cancer	1
Patient 3	AIS	Benign	Benign	Benign	0.9
Patient 4	AIS	Benign	Lung cancer	Lung cancer	0.9
Patient 5	AIS	Benign	Lung cancer	Lung cancer	0.8
Patient 6	AIS	Lung cancer	Benign	Lung cancer	0.9
Patient 7	AIS	Benign	Benign	Benign	0.5
Patient 8	AIS	Benign	Benign	Lung cancer	0.5
Patient 9	AIS	Lung cancer	Lung cancer	Benign	0.4
Patient 10	AIS	Benign	Benign	Lung cancer	3.2
Patient 11	AIS	Benign	Benign	Lung cancer	0.7

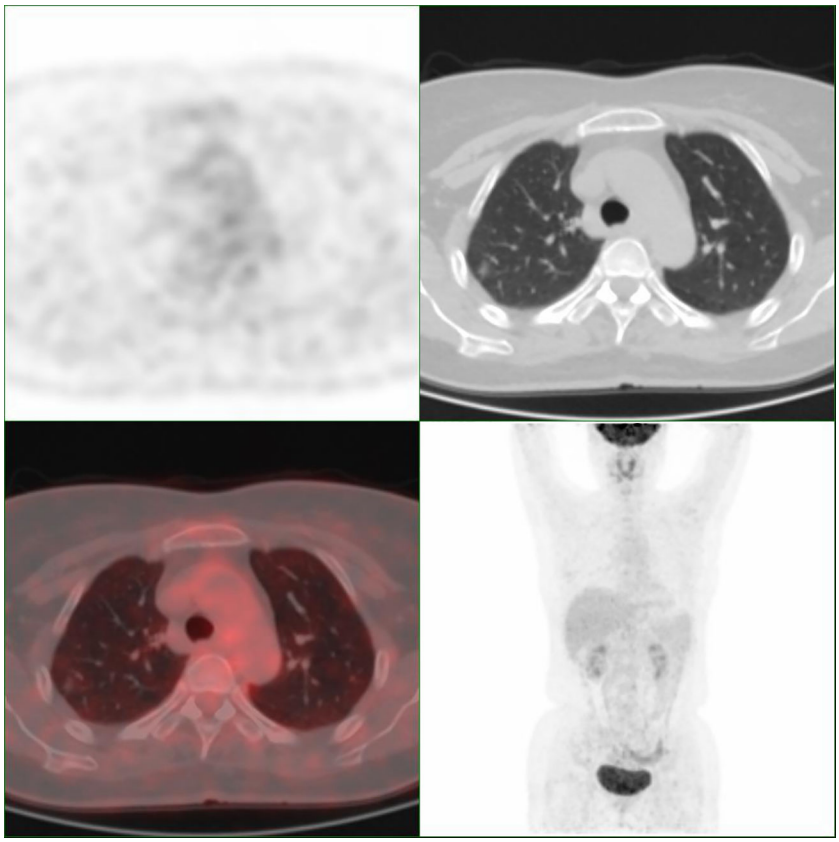


FIGURE 2
Representative images of Patient 4 with adenocarcinoma *in situ*. The nodule has an SUVmax of 0.9.

cancer screening strategy. Exhaled breath is easy and non-invasive to collect, and the breath test by HPPI-TOFMS is fast and feasible, which is very useful to help identify high-risk populations. A well-designed study is warned to investigate how to integrate the breath test into the current LDCT lung cancer screening model.

It is crucial to note the limitations of the study when interpreting the results. While over 90% of incidentally detected pulmonary nodules are generally benign, participants in this study were highly selected before they were recruited, which resulted in a lower proportion of benign nodules. On the other hand, although we have identified 16 lung cancer– specific VOCs in a previous study (26), we did not compare results from HPPI-TOFMS with the current gold- standard GC-MS as reported by Markar et al. (31).

Conclusions

In summary, the study demonstrated the efficacy of a breath test utilizing HPPI-TOFMS for discriminating lung cancer from benign pulmonary nodules, and the accuracy achieved by the exhaled breath test was comparable with 18F-FDG PET-CT.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee Board of Peking University People's Hospital. The patients/participants provided their written informed consent to participate in this study.

Author contributions

Conceptualization, XD, NL, ZY, and MQ; methodology, XD, PW, and GL; software, HC; formal analysis, XD; investigation, XD, HC, NL, and MQ; writing—original draft preparation, XD and MQ; writing—review and editing, XD, NL, ZY, and MQ; visualization, XD and MQ; supervision, NL, ZY, and MQ; project administration,

NL, ZY, and MQ; funding acquisition, MQ. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by National Natural Science Foundation of China, grant number 82173386.

Acknowledgments

Language polishing was partially assisted by ChatGPT3.5 via HIPLOT (https://hiplot.com.cn).

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

1. T. National Lung Screening Trial Research, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* (2011) 365(5):395–409. doi: 10.1056/NEJMoa1102873

2. Callister ME, Baldwin DR, Akram AR, Barnard S, Cane P, Draffan J, et al. British Thoracic society standards of care: British thoracic society guidelines for the investigation and management of pulmonary nodules. *Thorax* (2015) 70(Suppl 2): ii1–ii54. doi: 10.1136/thoraxjnl-2015-207168

3. MacMahon H, Naidich DP, Goo JM, Lee KS, Leung ANC, Mayo JR, et al. Guidelines for management of incidental pulmonary nodules detected on CT images: from the fleischner society 2017. *Radiology* (2017) 284(1):228–43. doi: 10.1148/radiol.2017161659

4. Naidich DP, Bankier AA, MacMahon H, Schaefer-Prokop CM, Pistolesi M, Goo JM, et al. Recommendations for the management of subsolid pulmonary nodules detected at CT: a statement from the fleischner society. *Radiology* (2013) 266(1):304–17. doi: 10.1148/radiol.12120628

5. Nasim F, Ost DE. Management of the solitary pulmonary nodule. *Curr Opin Pulm Med* (2019) 25(4):344–53. doi: 10.1097/MCP.0000000000000586

6. van Klaveren RJ, Oudkerk M, Prokop M, Scholten ET, Nackaerts K, Vernhout R, et al. Management of lung nodules detected by volume CT scanning. *N Engl J Med* (2009) 361(23):2221–9. doi: 10.1056/NEJMoa0906085

7. McWilliams A, Tammemagi MC, Mayo JR, Roberts H, Liu G, Soghrati K, et al. Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med* (2013) 369(10):910–9. doi: 10.1056/NEJMoa1214726

8. Gould MK, Donington J, Lynch WR, Mazzone PJ, Midthun DE, Naidich DP, et al. Evaluation of individuals with pulmonary nodules: when is it lung cancer? diagnosis and management of lung cancer, 3rd ed: American college of chest physicians evidence-based clinical practice guidelines. *Chest* (2013) 143(5 Suppl):e93S–e120S. doi: 10.1378/chest.12-2351

9. Garcia-Velloso MJ, Bastarrika G, de-Torres JP, Lozano MD, Sanchez-Salcedo P, Sancho L, et al. Assessment of indeterminate pulmonary nodules detected in lung cancer screening: diagnostic accuracy of FDG PET/CT. *Lung Cancer* (2016) 97:81–6. doi: 10.1016/j.lungcan.2016.04.025

10. Mazzone PJ, Lam L. Evaluating the patient with a pulmonary nodule: a review. *JAMA* (2022) 327(3):264–73. doi: 10.1001/jama.2021.24287

11. Horvath I, Lazar Z, Gyulai N, Kollai M, Losonczy G. Exhaled biomarkers in lung cancer. *Eur Respir J* (2009) 34(1):261–75. doi: 10.1183/09031936.00142508

12. Marzorati D, Mainardi L, Sedda G, Gasparri R, Spaggiari L, Cerveri P. A review of exhaled breath: a key role in lung cancer diagnosis. *J Breath Res* (2019) 13(3):034001. doi: 10.1088/1752-7163/ab0684

13. Meng S, Li Q, Zhou Z, Li H, Liu X, Pan S, et al. Assessment of an exhaled breath test using high-pressure photon ionization time-of-Flight mass spectrometry to detect lung cancer. *JAMA Netw Open* (2021) 4(3):e213486. doi: 10.1001/jamanetworkopen.2021.3486

14. Ibrahim W, Carr L, Cordell R, Wilde MJ, Salman D, Monks PS, et al. Breathomics for the clinician: the use of volatile organic compounds in respiratory diseases. *Thorax* (2021) 76(5):514–21. doi: 10.1136/thoraxjnl-2020-215667

15. van de Goor R, van Hooren M, Dingemans AM, Kremer B, Kross K. Training and validating a portable electronic nose for lung cancer screening. *J Thorac Oncol* (2018) 13(5):676–81. doi: 10.1016/j.jtho.2018.01.024

16. Chen X, Muhammad KG, Madeeha C, Fu W, Xu L, Hu Y, et al. Calculated indices of volatile organic compounds (VOCs) in exhalation for lung cancer screening and early detection. *Lung Cancer* (2021) 154:197–205. doi: 10.1016/j.lungcan.2021.02.006

17. Wang DH, Lee HS, Yoon D, Berry G, Wheeler TM, Sugarbaker DJ, et al. Progression of EGFR-mutant lung adenocarcinoma is driven by alveolar macrophages. *Clin Cancer Res* (2017) 23(3):778–88. doi: 10.1158/1078-0432.CCR-15-2597

18. Chen K, Nie Y, Park S, Zhang K, Zhang Y, Liu Y, et al. Development and validation of machine learning-based model for the prediction of malignancy in multiple pulmonary nodules: analysis from multicentric cohorts. *Clin Cancer Res* (2021) 27(8):2255–65. doi: 10.1158/1078-0432.Ccr-20-4007

19. Rai SN, Das S, Pan J, Mishra DC, Fu XA. Multigroup prediction in lung cancer patients and comparative controls using signature of volatile organic compounds in breath samples. *PloS One* (2022) 17(11):e0277431. doi: 10.1371/journal.pone.0277431

20. Guirao A, Molins L, Ramon I, Sunyer G, Vinolas N, Marrades R, et al. Trained dogs can identify malignant solitary pulmonary nodules in exhaled gas. *Lung Cancer* (2019) 135:230–3. doi: 10.1016/j.lungcan.2019.06.008

21. Huang Q, Wang S, Li Q, Wang P, Li J, Meng S, et al. Assessment of breathomics testing using high-pressure photon ionization time-of-Flight mass spectrometry to detect esophageal cancer. *JAMA Netw Open* (2021) 4(10):e2127042. doi: 10.1001/jamanetworkopen.2021.27042

22. Fu L, Feng Y, Ren T, Yang M, Yang Q, Lin Y, et al. Detecting latent tuberculosis infection with a breath test using mass spectrometer: a pilot cross-sectional study. *Biosci Trends* (2023) 17(1):73–7. doi: 10.5582/bst.2022.01476

23. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* (2015) 351:h5527. doi: 10.1136/bmj.h5527

24. Ransohoff DF. The process to discover and develop biomarkers for cancer: a work in progress. *J Natl Cancer Inst* (2008) 100(20):1419–20. doi: 10.1093/jnci/djn339

25. Zhou N, Zhang A, Su H, Zhao W, Li N, Yang Z. (18)F-FDG distribution pattern improves the diagnostic accuracy of single pulmonary solid nodule. *Front Oncol* (2022) 12:983833. doi: 10.3389/fonc.2022.983833

26. Wang P, Huang Q, Meng S, Mu T, Liu Z, He M, et al. Identification of lung cancer breath biomarkers based on perioperative breathomics testing: a prospective observational study. *EClinicalMedicine* (2022) 47:101384. doi: 10.1016/j.eclinm.2022.101384

27. Truong MT, Ko JP, Rossi SE, Rossi I, Viswanathan C, Bruzzi JF, et al. Update in the evaluation of the solitary pulmonary nodule. *Radiographics* (2014) 34(6):1658–79. doi: 10.1148/rg.346130092

28. Chen J, Mehta V, Chowdhary V, Chaya N, Lowry JW. Outcome of PET-negative solid pulmonary nodules: a retrospective study. *Acad Radiol* (2021) 28(5):628–33. doi: 10.1016/j.acra.2020.03.016

29. Khalaf M, Abdel-Nabi H, Baker J, Shao Y, Lamonica D, Gona J. Relation between nodule size and 18F-FDG-PET SUV for malignant and benign pulmonary nodules. *J Hematol Oncol* (2008) 1:13. doi: 10.1186/1756-8722-1-13

30. Akpoviroro O, Asokan S, Steiling K, Rebello A, Burks EJ, Litle VR, et al. Lung cancer risk in suspicious lung nodules with negative positron emission tomography. *Ann Thorac Surg* (2022) 113(6):1821–6. doi: 10.1016/j.athoracsur.2021.06.041

31. Markar SR, Wiggins T, Antonowicz S, Chin ST, Romano A, Nikolic K, et al. Assessment of a noninvasive exhaled breath test for the diagnosis of oesophagogastric cancer. *JAMA Oncol* (2018) 4(7):970–6. doi: 10.1001/jamaoncol.2018.0991

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 08 June 2023
DOI 10.3389/fonc.2023.1196883

Check for updates

OPEN ACCESS

EDITED BY
Hongda Liu,
Nanjing Medical University, China

REVIEWED BY
Cui Youbin,
Jilin University, China
Wang Shuai,
Fudan University, China

*CORRESPONDENCE
Hui Tian
tianhuiql@email.sdu.edu.cn

RECEIVED 30 March 2023
ACCEPTED 25 May 2023
PUBLISHED 08 June 2023

CITATION
Qiu J, Li R, Wang Y, Ma X, Qu C, Liu B, Yue W and Tian H (2023) A nomogram combining thoracic CT and tumor markers to predict the malignant grade of pulmonary nodules ≤3 cm in diameter. *Front. Oncol.* 13:1196883. doi: 10.3389/fonc.2023.1196883

COPYRIGHT
© 2023 Qiu, Li, Wang, Ma, Qu, Liu, Yue and Tian. 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 nomogram combining thoracic CT and tumor markers to predict the malignant grade of pulmonary nodules ≤3 cm in diameter

Jianhao Qiu[1], Rongyang Li[1], Yukai Wang[1], Xiuyuan Ma[2], Chenghao Qu[1], Binyan Liu[3], Weiming Yue[1] and Hui Tian[1]*

[1]Department of Thoracic Surgery, Qilu Hospital of Shandong University, Jinan, Shandong, China, [2]Department of Cardiology, Qianfoshan Hospital in the Shandong Province, Jinan, Shandong, China, [3]Department of Breast Surgery, Qilu Hospital of Shandong University, Jinan, Shandong, China

Background: With the popularity of computed tomography (CT) of the thorax, the rate of diagnosis for patients with early-stage lung cancer has increased. However, distinguishing high-risk pulmonary nodules (HRPNs) from low-risk pulmonary nodules (LRPNs) before surgery remains challenging.

Methods: A retrospective analysis was performed on 1064 patients with pulmonary nodules (PNs) admitted to the Qilu Hospital of Shandong University from April to December 2021. Randomization of all eligible patients to either the training or validation cohort was performed in a 3:1 ratio. Eighty-three PNs patients who visited Qianfoshan Hospital in the Shandong Province from January through April of 2022 were included as an external validation. Univariable and multivariable logistic regression (forward stepwise regression) were used to identify independent risk factors, and a predictive model and dynamic web nomogram were constructed by integrating these risk factors.

Results: A total of 895 patients were included, with an incidence of HRPNs of 47.3% (423/895). Logistic regression analysis identified four independent risk factors: the size, consolidation tumor ratio, CT value of PNs, and carcinoembryonic antigen levels in blood. The area under the ROC curves was 0.895, 0.936, and 0.812 for the training, internal validation, and external validation cohorts, respectively. The Hosmer-Lemeshow test demonstrated excellent calibration capability, and the fit of the calibration curve was good. DCA has shown the nomogram to be clinically useful.

Conclusion: The nomogram performed well in predicting the likelihood of HRPNs. In addition, it identified HRPNs in patients with PNs, achieved accurate treatment with HRPNs, and is expected to promote their rapid recovery.

KEYWORDS

pulmonary nodules, consolidation tumor ratio, CT values, carcinoembryonic antigen, nomogram

Introduction

Lung cancer is one of the leading causes of cancer related deaths worldwide with a 5-year overall survival (OS) rate of 19% (1). The popularity and development of low-dose thorax computed tomography (CT) have increased the diagnosis for patients with early-stage lung cancer (2). Pulmonary nodules (PNs) are early imaging manifestations of lung cancer, which are defined as focal, round-like, subsolid, or solid lung shadows with a diameter ≤3 cm (3, 4). PNs can be classified into a variety of pathological types, of which adenocarcinoma is the most common histological type; squamous cell and neuroendocrine carcinoma may also occur. Based on the fifth edition of the World Health Organization classification of thoracic tumors, atypical adenomatoid hyperplasia (AAH) and adenocarcinoma *in situ* (AIS) are classified as glandular precursor lesions. Adenocarcinoma is divided into two subtypes: minimally invasive adenocarcinoma (MIA) and invasive adenocarcinoma (IAC) (5, 6). Existing studies have demonstrated that radical resection can be achieved in both AIS and MIA patients. Among them, the 10-year OS rate was as high as 100% in AIS patients, while the 10-year OS rate in MIA patients was not significantly different from that of AIS (97.8%) (7, 8). For patients with IAC, however, the prognosis of early-stage Ia lung cancer is highly variable due to the different pathological subtypes. For lepidic invasive adenocarcinoma, the 5-year OS can reach 86%-100%, whereas the 5-year OS for solid invasive adenocarcinoma (SPA) and micropapillary invasive adenocarcinoma (MPA) is less than 60% (9–11). Therefore, ensuring that high-risk pulmonary nodules (HRPNs) be promptly resected at an early stage while avoiding overtreatment of low- risk pulmonary nodules (LRPNs) is an urgent need.

Histopathology remains the gold standard for a diagnosis of lung cancer (12, 13); however, given the invasive nature of surgery, the preoperative malignancy of early-stage PNs is not highly valued. The detection of tumor markers, such as carcinoembryonic antigen (CEA), cytokeratin fragment antigen 21-1 (CYFRA21-1), squamous cell carcinoma antigen (SCC), progastrin-releasing peptide (Pro-GRP), carbohydrate antigen 125 (CA125), and neuron-specific enolase (NSE), for lung cancer is widely used in clinical practice (14–17); however, it can easily cause errors in clinical diagnosis owing to its relatively low specificity. Thoracic CT is useful for diagnosing lung cancer. Previous studies have demonstrated that the consolidation tumor ratio (CTR) of PNs can be a critical parameter to predict the degree of malignancy and to determine the surgical method with a specificity of 98.7% (18). However, since early-stage lung cancers with different invasive degrees have overlapping CT morphologies, the CTR of the same PNs measured by different radiologists may be subjective to some extent, affecting the risk assessment of early-stage PNs. Accurately predicting the malignancy of patients' PNs without invasive procedures remains a considerable challenge for clinicians, and there is a need to construct an effective model for preoperative prediction of the risk for PNs with diameter ≤3 cm.

As an emerging visual statistical prediction model, a nomogram is often used to quantify the risk of clinical events, such as cancer (19, 20). No nomogram has incorporated CT imaging features and lung cancer tumor markers to differentiate HRPNs from LRPNs accurately. Therefore, our goal was to develop a nomogram based on important demographic information, clinical parameters, CT imaging characteristics, and tumor markers to assess the risk of PNs ≤3 cm in diameter prior to treatment and assist thoracic surgeons in making clinical decisions.

Abbreviations: AAH, Atypical adenomatoid hyperplasia; AIS, Adenocarcinoma *in situ;* AUC, Area under the ROC curve; BEAS, Bronchus-encapsulated air sign; CA125, Carbohydrate antigen 125; CEA, Carcinoembryonic antigen; CT, Computed tomography; CTR, Consolidation tumor ratio; CYFRA21-1, Cytokeratin fragment antigen 21-1; DCA, Decision curve analysis; HRPNs, High-risk pulmonary nodules; IAC, Invasive adenocarcinoma; LRPNs, Low-risk pulmonary nodules; MIA, Minimally invasive adenocarcinoma; MPA, Micropapillary invasive adenocarcinoma; NSE, Neuron-specific enolase; OS, Overall survival; PN, Pulmonary nodules; Pro-GRP, Progastrin-releasing peptide; ROC, Receiver operating characteristic; SCC, Squamous cell carcinoma; SD, Standard deviation; SPA, Solid invasive adenocarcinoma.

Patients and methods

Protocol and ethics statement

This retrospective study was conducted in accordance with the principles of the Declaration of Helsinki, and the Ethics Committee of Qilu Hospital of Shandong University approved the protocol (registration number: KYLL- 202008-023-1). All patients gave informed consent for the use of their clinical data.

Patients' selection

A retrospective analysis of 1064 patients hospitalized for PNs in Qilu Hospital of Shandong University from April to December 2021 was conducted by searching a prospectively maintained database. The exclusion criteria were as follows: (i) age <18 years, (ii) no thoracic CT within two week before surgery, (iii) preoperative puncture biopsy or fiberoptic bronchoscopy confirmed lung malignancy, (iv) multiple pulmonary nodules, (v) postoperative pathology of pulmonary metastatic tumors, and (vi) incomplete perioperative data. Overall, 812 patients with 812 lesions met the inclusion criteria and were randomly allocated to either the training or internal validation cohort in a 3:1 ratio according to the random split-sample method. In addition, following the inclusion criteria, the thoracic surgery database of Qianfoshan Hospital of Shandong Province was searched, and 83 patients treated at this hospital from January to April 2022 were included as an independent external validation cohort. The training cohort was used to develop a predictive nomogram, while the validation cohorts were used to validate the nomogram's performance.

CT scanning

Before the surgical intervention, all patients underwent a thoracic CT scan (Brilliance iCT 256, Philips Healthcare, USA).

The CT imaging parameters were as follows: tube voltage, 120 kV; tube current, 150 mA; field of view, 350 mm; slice thickness, 1.0 mm; and gantry rotation time, 270 ms.

Image assessment

The CT imaging features were independently assessed by two experienced thoracic surgeons (Jianhao Qiu and Rongyang Li) who were blinded to the patient's clinical and pathological information prior to the assessment. Each lesion was assessed based on nodule diameter, CT values, CTR, lobulation, spiculation, bronchus-encapsulated air sign (BEAS), cavity sign, pleural traction, and vascular convergence. The nodule diameter was defined as the diameter (mm) of the largest lesion on an axial image at a given level (Figure 1). The CT values of the PNs were measured on the thoracic window with the largest nodule diameter using either a circular or oval region of interest, covering at least one-half of the nodule's larger surface area, excluding obvious vessels and bronchi (Figure 1). The CT value measurement was repeated three times for all PNs, and the average CT values obtained six times by two thoracic surgeons were chosen as the representative PNs values. The CTR was defined as the ratio between the largest diameter of the solid component of the PNs and the largest diameter of the ground-glass nodule. For the irregularly shaped nodules, the average of the long and short diameters was used (Figure 1). Lobulation was defined as a portion of the surface of the lesion showing a scallop-like sign. Spiculation was defined as cords extending from the rim of the nodule into the lung parenchyma without reaching the pleural surface. BEAS was defined as small air-like low-density shadows with a smooth inner edge and a diameter <5 mm. Low-density shadows with a diameter ≥5 mm were defined as cavity signs. A linear, tentorial, or stellate shadow between the PNs and pleura was considered pleural traction. Vascular convergence was defined as a packing of the inner portion of the PNs or abnormal angulation to the PNs compared to that of normal lung parenchyma (Supplementary Figures 1, 2).

Histopathologic features analysis

All pathological specimens were fixed in formalin, sectioned, and stained with hematoxylin-eosin, following standard pathological section preparation procedures. Light microscopic histopathological evaluation of hematoxylin-eosin- stained slides was performed by two experienced pathologists blinded to patient data, who independently evaluated each tissue section. Following the fifth edition of the World Health Organization classification of thoracic tumors, we classified PNs diagnosed as benign lesions, AAH, AIS, MIA, and well- differentiated lung cancer into the low-risk group. The pathological types of PNs classified as high-risk include: (i) moderately differentiated lung cancer; (ii) poorly differentiated lung cancer; (iii) well-differentiated lung adenocarcinoma with high-grade components accounting for ≥5%; and (iv) well-differentiated lung adenocarcinoma with high-risk factors. High-grade components include solid, micropapillary, cribriform, and complex glandular structures (fused glands or single cells infiltrating the desmoplastic stroma). High-risk factors include lymph node metastasis, neural invasion, visceral pleural invasion, vascular tumor thrombus, and airspace spread.

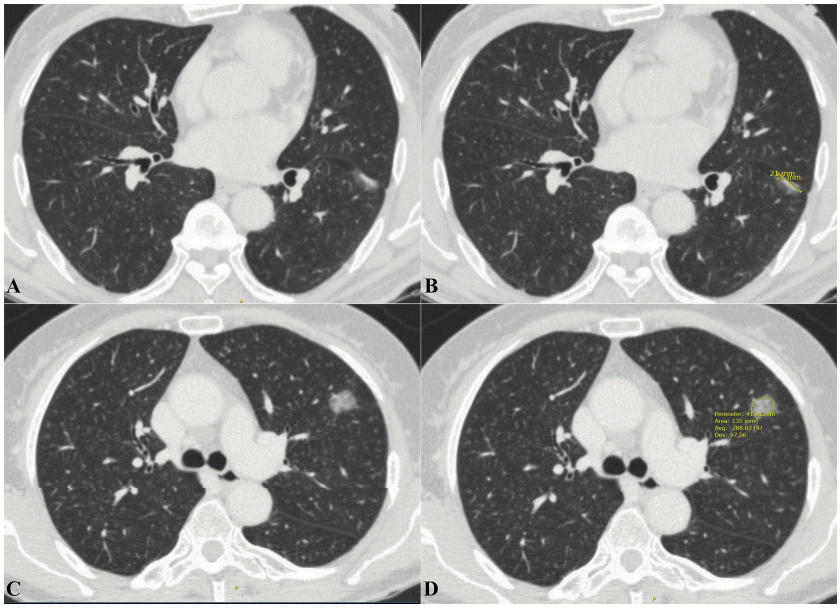


FIGURE 1
Extraction of imaging data related to pulmonary nodules from thoracic CT. **(A)** Axial thoracic CT images showed a mixed density ground-glass nodule in the left lower lobe. Histology confirmed it as IAC. **(B)** The long diameter of the solid component was measured to be 14mm and the long diameter of the whole pulmonary nodule to be 21mm. The CTR value was about 0.677. **(C)** Axial thoracic CT images showed a high-density ground-glass nodule in the left upper lobe. Histology confirmed it as IAC. **(D)** Mean CT values of ground-glass nodule is -288.63 HU. CT, Computed tomography; IAC, Invasive adenocarcinoma; CTR, Consolidation tumor ratio.

Nomogram construction

To identify potential risk factors for HRPNs, univariable logistic regression analysis was conducted. Any factor with a P value of less than 0.2 in the univariable analysis was included in further multivariable analysis. The predictive model was based on the independent risk factors (multivariable logistic regression analysis, P <0.05). The results of the multivariable logistic regression model were then used to construct a nomogram using the packages "rms" and "DynNom" in the R Project software (version 4.2.1; http://www.r-project.org/). A regression model was used to calculate the score for each variable, and the predicted probability of HRPN was derived by summing each variable score.

Nomogram performance

The predictive nomogram's performance was assessed based on discrimination, calibration, and clinical utility. Discrimination is the ability of a model to correctly discriminate events from non-events. Receiver operating characteristic (ROC) curves were used to assess the discriminatory effectiveness of the predicted nomogram (21). The calibration measures how well the predicted probabilities match the true results. The Hosmer-Lemeshow test was used to assess the calibration ability, and a P-value >0.05 indicated satisfactory calibration (22).Calibration was then assessed further by constructing a nomogram calibration plot. Internal and external verifications were conducted using the bootstrapping method with 1000 repetitions (23). A decision curve analysis (DCA) was conducted to assess the clinical utility of the predictive nomograms based on the net benefit at different probability thresholds (24).

Statistical analysis

Continuous normally distributed variables were compared using at a t-test, expressed as mean ± standard deviation (SD). If continuous variables are not normally distributed, the data are expressed as medians (interquartile range [IQR]) and compared between groups using the Mann-Whitney U test. Compared of categorical variables were performed using Pearson's chi-squared or Fisher's exact test. A two-sided P value of <0.05 was used for statistical significance. The SPSS software (v25.0; IBM Corp., Armonk, NY, USA) and the R Project software (v4.2.1; http://www.R-project.org) were used for the data analysis.

Results

Patient characteristics

The identification and selection process for eligible patients is shown in Figure 2. Overall, 895 patients with an HRPN incidence rate of 47.3% were included (423/895). The proportion of patients with IAC was 51.7% (463/895), while the proportion of patients with MIA was 24.5% (219/895) and the proportion of patients with AIS was 17.8% (159/895). Subsequently, 812 patients from Qilu Hospital of Shandong University were randomly assigned in a 3:1 ratio to the training (n = 609) or internal validation cohort (n = 203). Patients from Qianfoshan Hospital of Shandong Province were included in the independent external validation cohort (n = 83). No variables differed significantly between the training and validation cohorts (Table 1). Patients were divided into HRPNs and non-HRPN groups based on the presence or absence of HRPNs. The characteristics of the training and validation cohorts are presented in Table 2.

Identification of HRPN risk factors

Univariable and multivariable logistic regression analyses were performed on the training cohort to identify HRPN risk factors (Table 3). Univariable logistic regression analysis showed that age, gender, body mass index, smoking history, chronic lung disease, abnormal pulmonary function, size, CTR, and CT values of PNs, BEAS or cavity sign, lobulation, spiculation, pleural traction, vascular convergence, and blood tumor marker (CEA, CYFRA21-1, SCC, Pro-GRP, and NSE) levels were potential risk factors (P< 0.2). Further multivariable logistic regression analysis (forward stepwise regression) showed that the PN size (P<0.001), CTR (odds ratio [OR] = 3.8338; 95% confidence interval [CI]: 1.661-8.868; P=0.002), CT value of PNs (P<0.001), and blood CEA levels (OR = 1.701; 95% CI: 1.702–2.701; P =0.024) were independently associated with HRPNs. For some continuous variables in the study factors (CTR, CEA, CYFRA21-1, SCC, Pro- GRP, and NSE), their optimal cutoff values were determined by plotting their ROC curves. The best cutoff value was used as the standard for converting into binary variables and included in the regression analysis (Supplementary Table 1).

Nomogram construction

Four independent HRPN risk factors were included in the logistic regression models. A nomogram of predicted HRPNs based on the coefficients from the multiple logistic regression model was plotted using the "rms" package in the R statistical software (Figure 3). The nomogram shows seven axes, with axes 2-5 representing the four variables in the predictive model. Each variable was scored on a numeric scale ranging from 0 to 100. A total score can be calculated by summing each factor's score. By projecting the total score onto the lower total score scale axis, we can predict the likelihood of HRPNs. Furthermore, in order to facilitate the widespread use of our predictive nomograms among thoracic surgeons, we built an operator interface on a web page using the "Dynnom" package to calculate the predicted probabilities of HRPNs. By selecting the patient's preoperative imaging features and tumor marker levels, the user can obtain the predicted probability that PN is a high-risk type.

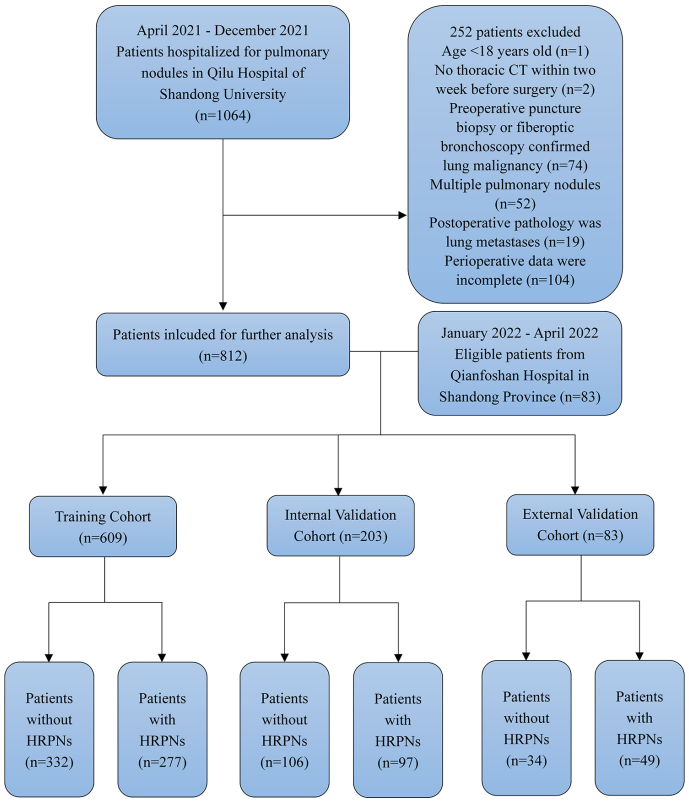


FIGURE 2
Flow diagram of patient selection through the study. HRPNs, High-risk pulmonary nodules.

Predictive performance and nomogram validation

The discriminative ability of the prediction model and nomogram was assessed using an ROC curve (Figure 4). The area under the ROC curve (AUC) was 0.895 (95% CI: 0.870-0.920), 0.936 (95% CI:0.903-0.970), and 0.812 (95% CI:0.717-0.906) for the training, internal validation, and external validation cohorts, respectively. This indicated that the nomogram's prediction accuracy was relatively good. For the predicted probability of HRPNs, the optimal cut-off value was approximately 45.09%, with a sensitivity and specificity of 0.846 and 0.812, respectively (Supplementary Table 1). The Hosmer-Lemeshow test showed excellent calibration ability (P=0.418, 0.916, and 0.975 in the training, internal validation, and external validation cohorts, respectively). The nomograph's calibration curve of HRPN predicted probabilities also showed excellent concordance between the predicted and actual results (Figure 5).

Predictive nomogram's clinical utility

The predictive nomogram's clinical utility was assessed using DCA(Figure 6). The results showed that the nomogram for HRPNs' prediction provided a larger net gain with a broader range of threshold probabilities for predicting HRPNs risk across both the training and validation cohorts. It also proved evidence that the nomogram can be applied clinically and assist surgeons in making better clinical decisions.

Discussion

With the popularity of lung cancer screening using low-dose thoracic thin-slice CT, the PN detection rate has increased (2). For mixed ground-glass nodules, it is difficult to make a definitive preoperative histopathological prediction of PN based on its imaging features (such as lobulation, spiculation, or pleural stretch) (25). Consequently, many low-grade PNs are overtreated, leading to increased hospitalization costs, longer hospital stays, and a higher risk of postoperative complications (26). Therefore, in this study, we combined patients' preoperative imaging information and blood tumor marker levels to develop a clinical prediction model and designed a nomogram with good predictive performance for the degree of PNs risk. Clinicians can estimate the probability that a patient's PN is HRPN before surgery using this predictive nomogram, thereby making sound treatment decisions for PNs with various risks.

This study showed that PN diameter, CT value, CTR, and blood CEA level were independent risk factors for HRPNs. One of the most important imaging features for determining the malignancy of PNs is their diameter. As the diameter of the nodule increases, the

TABLE 1 Baseline characteristics of included patients and comparison between groups.

Characteristics	All	TC	IVC	EVC	P value	
	(n=812)	(n=609)	(n=203)	(n=83)	TC vs IVC	TC vs EVC
Age (years), median (IQR)	58.5 (52-66)	58 (52-66)	59 (51-67)	59 (51-66)	0.854	0.815
Gender, N(%)					0.422	0.015
Female	521 (64.16)	396 (65.02)	125 (61.58)	42 (50.60)		
Male	291 (35.84)	213 (34.98)	78 (38.42)	41 (49.40)		
BMI (kg/m^2), median (IQR)	25.0 (22.9-27.2)	24.9 (22.9-27.1)	25.2 (23.0-27.6)	24.6 (23.1-26.7)	0.318	0.795
Smoking history, N(%)					0.201	0.065
(-)	636 (78.33)	484 (79.47)	152 (74.88)	58 (69.88)		
(+)	176 (21.67)	125 (20.53)	51 (25.12)	25 (30.12)		
CLD, N(%)					0.388	0.259
(-)	651 (80.17)	493 (80.95)	158 (77.83)	72 (86.75)		
(+)	161 (19.83)	116 (19.05)	45 (22.17)	11 (13.25)		
PF, N(%)					0.266	0.437
Normal	477 (58.74)	365 (59.93)	112 (55.17)	54 (65.06)		
Abnormal	335 (41.26)	244 (40.07)	91 (44.83)	29 (34.94)		
Family history of cancer, N(%)					0.315	0.060
(-)	667 (82.14)	495 (81.28)	172 (84.73)	75 (90.36)		
(+)	145 (17.86)	114 (18.72)	31 (15.27)	8 (9.64)		
Nodule diameter (d), N(%)					0.615	0.165
5 mm ≤ d ≤ 10 mm	229 (28.20)	177 (29.07)	52 (25.62)	22 (26.51)		
10 mm < d ≤ 20 mm	404 (49.75)	298 (48.93)	106 (52.22)	35 (42.17)		
20 mm < d ≤ 30 mm	179 (22.05)	134 (22.00)	45 (22.16)	26 (31.32)		
CTR, median (IQR)	0.21 (0.00-0.76)	0.20 (0.00-0.74)	0.24 (0.00-0.82)	0.22 (0.00-0.64)	0.189	0.666
CT value (n)					0.448	0.606
-800HU<n≤-600HU	217 (26.73)	166 (27.26)	51 (25.12)	21 (25.30)		
-600HU<n≤-400HU	245 (30.17)	190 (31.20)	55 (27.10)	28 (33.73)		
-400HU<n≤-200HU	143 (17.61)	105 (17.24)	38 (18.72)	18 (21.69)		
n>-200HU	207 (25.49)	148 (24.30)	59 (29.06)	16 (19.28)		
BEAS or Cavity Sign, N(%)					0.219	0.177
(-)	671 (82.64)	497 (81.61)	174 (85.71)	62 (74.70)		
(+)	141 (17.36)	112 (18.39)	29 (14.29)	21 (25.30)		
Lobulation, N(%)					0.571	0.416
(-)	555 (68.35)	420 (68.97)	135 (66.50)	53 (63.86)		
(+)	257 (31.65)	189 (31.03)	68 (33.50)	30 (36.14)		
Spiculation, N(%)					0.372	0.085
(-)	428 (52.71)	315 (51.72)	113 (55.67)	34 (40.96)		
(+)	384 (47.29)	294 (48.28)	90 (44.33)	49 (59.04)		
PT, N(%)					0.902	0.327
(-)	461 (56.77)	347 (56.98)	114 (56.16)	42 (50.60)		

(Continued)

TABLE 1 Continued

Characteristics	All	TC	IVC	EVC	P value	
	(n=812)	(n=609)	(n=203)	(n=83)	TC vs IVC	TC vs EVC
(+)	351 (43.23)	262 (43.02)	89 (43.84)	41 (49.40)		
VC, N(%)					0.491	0.061
(-)	597 (73.52)	452 (74.22)	145 (71.43)	70 (84.34)		
(+)	215 (26.48)	157 (25.78)	58 (28.57)	13 (15.66)		
CEA (ng/ml), median (IQR)	2.23 (1.45-3.36)	2.22 (1.47-3.37)	2.25 (1.40-3.34)	2.35 (1.56-3.81)	0.961	0.264
CYFRA21-2 (ng/ml), median (IQR)	1.98 (1.44-2.57)	1.97 (1.44-2.53)	2.01 (1.41-2.65)	2.27 (1.60-3.35)	0.967	0.004
SCC (ng/ml), median (IQR)	0.96 (0.72-1.34)	0.94 (0.70-1.33)	1.01 (0.76-1.35)	0.80 (0.60-1.10)	0.248	0.003
Pro-GRP (pg/ml), median (IQR)	42.00 (34.10-51.08)	42.00 (34.05-51.88)	42.56 (34.33-49.70)	34.05 (27.26-41.40)	0.918	<.001
NSE (ng/ml), median (IQR)	19.00 (15.90-23.28)	19.00 (16.00-23.40)	19.10 (15.70-23.10)	14.10 (9.31-18.20)	0.942	<.001
High-risk pulmonary nodule, N(%)					0.626	0.028
(-)	438 (53.94)	332 (54.52)	106 (52.22)	34 (40.96)		
(+)	374 (46.06)	277 (45.48)	97 (47.78)	49 (59.04)		

* P-value for the comparison between training cohort and validation cohort (internal validation cohort and external validation cohort).
IQR, interquartile range; (-), No; (+), Yes; TC, training cohort; IVC, internal validation cohort; EVC, external validation cohort; BMI, body mass index; CLD, chronic lung disease; PF, pulmonary function; CTR, consolidation tumor ratio; BEAS, bronchus encapsulated air sign; PT, pleural traction; VC, vascular convergence; CEA, carcinoembryonic antigen; CYFRA21-1, cytokeratin fragment antigen 21-1; SCC, squamous cell carcinoma antigen; Pro-GRP, progastrin-releasing peptide; NSE, neuron specific enolase.

TABLE 2 Clinical characteristics of patients with or without HRPN in training cohort and validation cohort (internal and external).

Characteristics	Training Cohort			Internal Validation Cohort			External Validation Cohort		
	Non HRPN (n=332)	HRPN (n=277)	P	Non HRPN (n=106)	HRPN (n=97)	P	Non HRPN (n=34)	HRPN (n=49)	P
Age[a] (years)	56 (49-65)	62 (55-67)	<.001	55 (47.75-63)	63 (56-69)	<.001	56.5 (47.75-64)	61 (54.5-67.5)	0.046
Gender[b]			0.058			0.012			0.090
Female	227 (68.37)	169 (61.01)		74 (69.81)	51 (52.58)		21 (61.76)	21 (42.86)	
Male	105 (31.63)	108 (38.99)		32 (30.19)	46 (47.42)		13 (38.24)	28 (57.14)	
BMI[a] (kg/m^2)	24.7 (22.7-27.0)	25.2 (23.2-27.2)	0.077	25.0 (22.4-27.2)	25.3 (23.6-28.0)	0.184	24.7 (23.0-26.5)	24.6 (23.1-26.8)	0.732
Smoking history[b]			0.008			<.001			0.011
(-)	277 (83.43)	207 (74.73)		90 (84.91)	62 (63.92)		29 (85.29)	29 (59.18)	
(+)	55 (16.57)	70 (25.27)		16 (15.09)	35 (36.08)		5 (14.71)	20 (40.82)	
CLD[b]			0.011			0.001			0.739
(-)	281 (84.64)	212 (76.53)		92 (86.79)	66 (68.04)		30 (88.24)	42 (85.71)	
(+)	51 (15.36)	65 (23.47)		14 (13.21)	31 (31.96)		4 (11.76)	7 (14.29)	
PF[b]			0.013			<.001			0.681
Normal	214 (64.46)	151 (54.51)		71 (66.98)	41 (42.27)		23 (67.65)	31 (63.27)	
Abnormal	118 (35.54)	126 (45.49)		35 (33.02)	56 (57.73)		11 (32.35)	18 (36.73)	
Family history of cancer[b]			0.811			0.479			0.085
(-)	271 (81.63)	224 (80.87)		88 (83.02)	84 (86.60)		33 (97.06)	42 (85.71)	
(+)	61 (18.37)	53 (19.13)		18 (16.98)	13 (13.40)		1 (2.94)	7 (14.29)	
Nodule diameter[b] (d)			<.001			<.001			<.001
5 mm ≤ d ≤ 10 mm	152 (45.78)	25 (9.02)		47 (44.34)	5 (5.15)		15 (44.12)	7 (14.28)	

(Continued)

TABLE 2 Continued

Characteristics	Training Cohort			Internal Validation Cohort			External Validation Cohort		
	Non HRPN (n=332)	HRPN (n=277)	P	Non HRPN (n=106)	HRPN (n=97)	P	Non HRPN (n=34)	HRPN (n=49)	P
10 mm < d ≤ 20 mm	150 (45.18)	148 (53.43)		53 (50.00)	53 (54.64)		15 (44.12)	20 (40.82)	
20 mm < d ≤ 30 mm	30 (9.04)	104 (37.55)		6 (5.66)	39 (40.21)		4 (11.76)	22 (44.90)	
CTR[a]	0.00 (0.00-0.14)	0.71 (0.27-1.00)	<.001	0.00 (0.00-0.00)	0.82 (0.41-1.00)	<.001	0.00 (0.00-0.13)	0.51 (0.18-0.77)	<.001
CT value[b] (n)			<.001			<.001			<.001
-800HU<n≤-600HU	158 (47.59)	8 (2.89)		50(47.17)	1(1.03)		15 (44.12)	6 (12.25)	
-600HU<n≤-400HU	129 (38.86)	61 (22.02)		41(38.68)	14(14.43)		14 (41.18)	14 (28.57)	
-400HU<n≤-200HU	28 (8.43)	77 (27.80)		12(11.32)	26(26.81)		4 (11.76)	14 (28.57)	
n>-200HU	17 (5.12)	131 (47.29)		3(2.83)	56(57.73)		1 (2.94)	15 (30.61)	
BEAS or Cavity Sign[b]			0.020			<.001			0.064
(-)	282 (84.94)	215 (77.62)		101(95.28)	73(75.26)		29 (85.29)	33 (67.35)	
(+)	50 (15.06)	62 (22.38)		5(4.72)	24(24.74)		5 (14.71)	16 (32.65)	
Lobulation[b]			<.001			<.001			0.014
(-)	286 (86.14)	134 (48.38)		95(89.62)	40(41.24)		27 (79.41)	26 (53.06)	
(+)	46 (13.86)	143 (51.62)		11(10.38)	57(58.76)		7 (20.59)	23 (46.94)	
Spiculation[b]			<.001			<.001			<.001
(-)	219 (65.96)	96 (34.66)		80(75.47)	33(34.02)		23 (67.65)	11 (22.45)	
(+)	113 (34.04)	181 (65.34)		26(24.53)	64(65.98)		11 (32.35)	38 (77.55)	
PT[b]			<.001			<.001			0.032
(-)	227 (68.37)	120 (43.32)		83(78.30)	31(31.96)		22 (64.71)	20 (40.82)	
(+)	105 (31.63)	157 (56.68)		23(21.70)	66(68.04)		12 (35.29)	29 (59.18)	
VC[b]			0.001			0.689			0.416
(-)	264 (79.52)	188 (67.87)		77(72.64)	68(70.10)		30 (88.24)	40 (81.63)	
(+)	68 (20.48)	89 (32.13)		29(27.36)	29(29.90)		4 (11.76)	9 (18.37)	
CEA[a] (ng/ml)	1.93 (1.31-2.98)	2.49 (1.72-4.05)	<.001	1.76 (1.25-2.74)	2.74 (1.83-4.21)	<.001	2.12 (1.11-3.93)	2.46 (1.66-3.74)	0.418
CYFRA21-2[a] (ng/ml)	1.83 (1.38-2.45)	2.05 (1.55-2.72)	0.001	1.87 (1.34-2.65)	2.13 (1.54-2.66)	0.122	2.23 (1.30-3.32)	2.27 (1.81-3.46)	0.350
SCC[a] (ng/ml)	0.92 (0.72-1.29)	0.98 (0.69-1.36)	0.414	1.07 (0.80-1.37)	0.90 (0.72-1.30)	0.067	0.80 (0.60-1.13)	0.80 (0.60-1.05)	0.503
Pro-GRP[a] (pg/ml)	42.39	41.77	0.901	41.14	43.03	0.064	32.00	34.24	0.846
	(34.83-50.54)	(33.52-53.17)		(32.07-49.16)	(36.49-51.05)		(27.62-41.37)	(26.95-42.45)	
NSE[a] (ng/ml)	19.00	19.00	0.899	19.50	19.00	0.961	13.62	15.20	0.245
	(15.83-23.48)	(16.05-23.15)		(15.35-24.08)	(16.00-22.45)		(8.91-16.20)	(10.18-19.60)	

[a]Variables were compared by the Mann-Whitney U test between groups, and data were expressed as medians (interquartile range [IQR])
[b]Variables were compared using Pearson's chi-square test or Fisher's exact test, and results were expressed as percentages.
HRPN, high-risk pulmonary nodule; (-), No; (+), Yes; BMI, body mass index; CLD, chronic lung disease; PF, pulmonary function; CTR, consolidation tumor ratio; BEAS, bronchus encapsulated air sign; PT, pleural traction; VC, vascular convergence; CEA, carcinoembryonic antigen; CYFRA21-1, cytokeratin fragment antigen 21-1; SCC, squamous cell carcinoma antigen; Pro-GRP, progastrin-releasing peptide; NSE, neuron specific enolase.

depth of invasion and the probability of adverse PN pathological types also increase. Our study showed that PNs with diameters of 10-20 mm and 20-30 mm had a 2.234- fold and 3.873-fold higher risk of HRPNs, respectively compared with PNs of 5–10 mm. This finding is consistent with that of previous studies on the size and malignancy of PNs (25, 27). Previous studies have demonstrated that the average CT value of PNs can discriminate between a variety of invasive lung cancers, and a higher CT value indicates a higher possibility of malignancy (28, 29). In one study, it was found that the average CT value could be used to predict the growth of pure

10.3389/fonc.2023.1196883

TABLE 3 Univariable and multivariable logistic regression analysis of risk factors for HRPN in the training cohort.

Variable	Univariable analysis			Multivariable analysis		
	OR	95%CI	P	OR	95%CI	P
Age (years)	1.049	1.031-1.067	<.001			
Gender			0.058			
Female	Ref.					
Male	1.382	0.989-1.930				
BMI (kg/m^2)	1.039	0.991-1.090	0.115			
Smoking history			0.008			
(-)	Ref.					
(+)	1.703	1.146-2.532				
CLD			0.012			
(-)	Ref.					
(+)	1.689	1.124-2.540				
PF			0.013			
Normal	Ref.					
Abnormal	1.513	1.092-2.097				
Family history of cancer			0.811			
(-)	Ref.					
(+)	1.051	0.699-1.581				
Nodule diameter (d)			<.001			0.001
5 mm ≤ d ≤ 10 mm	Ref.			Ref.		
10 mm < d ≤ 20 mm	5.999	3.712-9.696		2.234	1.242-4.017	
20 mm < d ≤ 30 mm	21.077	11.725-37.888		3.873	1.888-7.944	
CTR			<.001			0.002
≤0.265	Ref			Ref.		
>0.265	21.642	14.154-33.092		3.838	1.661-8.868	
CT value (n)			<.001			<.001
-800HU<n≤-600HU	Ref.			Ref.		
-600HU<n≤-400HU	9.339	4.312-20.227		7.287	3.308-16.048	
-400HU<n≤-200HU	54.312	23.645-124.757		11.846	3.962-35.422	
n>-200HU	152.191	63.655-363.870		23.553	6.909-80.289	
BEAS or Cavity Sign			0.021			
(-)	Ref.					
(+)	1.626	1.077-2.457				
Lobulation			<.001			
(-)	Ref.					
(+)	6.635	4.490-9.804				
Spiculation			<.001			
(-)	Ref.					
(+)	3.654	2.612-5.112				

(Continued)

10.3389/fonc.2023.1196883

TABLE 3 Continued

Variable	Univariable analysis			Multivariable analysis		
	OR	95%CI	P	OR	95%CI	P
PT			<.001			
(-)	Ref.					
(+)	2.828	2.030-3.941				
VC			0.001			
(-)	Ref.					
(+)	1.838	1.273-2.653				
CEA			<.001			0.024
≤1.965 ng/ml	Ref.			Ref.		
>1.965 ng/ml	2.399	1.718-3.350		1.701	1.072-2.701	
CYFRA21-2			<.001			
≤1.785 ng/ml	Ref.					
>1.785 ng/ml	1.863	1.340-2.590				
SCC			0.015			
≤1.655 ng/ml	Ref.					
>1.655 ng/ml	1.850	1.126-3.039				
Pro-GRP			0.078			
≤33.455 pg/ml	Ref.					
>33.455 pg/ml	0.706	0.480-1.039				
NSE			0.180			
≤17.75 ng/ml	Ref.					
>17.75 ng/ml	1.251	0.902-1.735				

HRPN, high-risk pulmonary nodule; (-), No; (+), Yes; BMI, body mass index; CLD, chronic lung disease; PF, pulmonary function; CTR, consolidation tumor ratio; BEAS, bronchus encapsulated air sign; PT, pleural traction; VC, vascular convergence; CEA, carcinoembryonic antigen; CYFRA21-1, cytokeratin fragment antigen 21-1; SCC, squamous cell carcinoma antigen; Pro-GRP, progastrin-releasing peptide; NSE, neuron specific enolase.

ground-glass nodules with an optimum cut-off value of -670HU (30). In another study by Ikeda et al., it was proven that a cut-off value of -584 HU was helpful in distinguishing AAH from AIS and of -472 HU was helpful in distinguishing AIS from IAC (31). Koezuka et al. also confirmed that CT values reflect the cellular structure and density of the lung nodules (32). The highly invasive component was usually present in the site with a high CT value, while the site with a low CT value was diagnosed as having lower invasiveness based on the final pathological result. According to the results of Ikeda, Eguchi and Koezuka et al., the CT value of PNs was

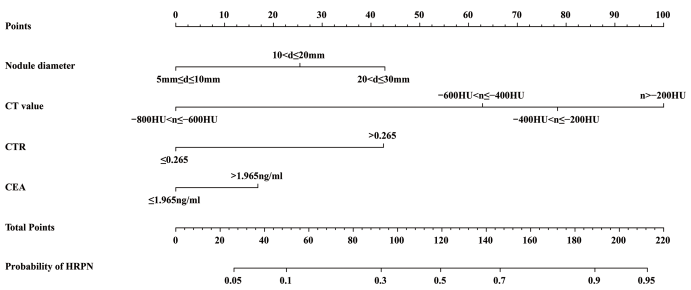


FIGURE 3
A nomogram predicting the probability that PNs with diameter ≤3 cm are HRPNs. Draw a vertical line from each variable's corresponding axis to the point axis to get the points for that variable. Summing the scores for each variable to obtain the total score, the probability of predicting HRPNs can be estimated by projecting the total score onto the lower total score axis. PNs, Pulmonary nodules; HRPNs, High-risk pulmonary nodules.

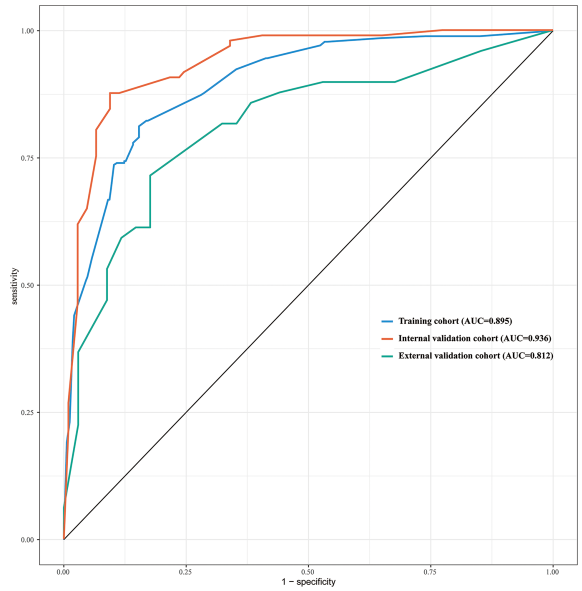


FIGURE 4
ROC curves of the nomogram used to predict the probability of HRPNs in training and validation cohorts. ROC, Receiver operating characteristics; HRPNs, High-risk pulmonary nodules; AUC, Area under the ROC curve.

divided by a gradient of 200HU, which could well reflect the change degree of the solid component of PNs (30–32). Therefore, our prediction model was divided by this gradient. The final results demonstrated that the higher the CT value of the nodule, the higher the risk of HRPNs (OR=7.287, 11.846, and 23.553, respectively) compared with PNs with CT values between -800HU and -600HU. These results are in agreement with those of Ikeda et al. and Eguchi et al (30, 31).

The study of CTR has a long history, and the choice of the optimal cutoff value is still controversial. Suzuki et al. have shown that a CTR ≤0.25 is one of the radiological criteria for non-invasive lung cancer, and the specificity of judging the nodule malignancy before surgery can reach 98.7% (18). According to literature reports, mixed ground-glass nodules with CTR ≤0.5 and diameter ≤1cm have better 5-year OS and recurrence-free survival (33, 34). A study from Japan showed that PNs with a CTR <0.5 and a diameter of less than 2 cm had better recurrence-free survival and OS. And CTR >0.5 often predicted a poor prognosis for patients with these PNs (35). In this study, after multivariable logistic regression analysis, CTR was established as an independent risk factor for HRPNs (OR=3.838, P=0.002). ROC curve analysis was performed on the CTR of the included patients, with an AUC value of 0.869 and an optimal cut-off value of 0.265 (similar to the study of Suzuki K et al.) (18). This indicates that the CTR has a good predictive performance for HRPNs. However, the optimal cutoff value was different from that reported in Sun's study (34), which may be due to the different diameter criteria of the included PNs. In addition, unlike Sun's study, PNs were reclassified in this study using the fifth edition of the World Health Organization classification of thoracic tumors. This will undoubtedly affect the optimal cut-off value for the CTR.

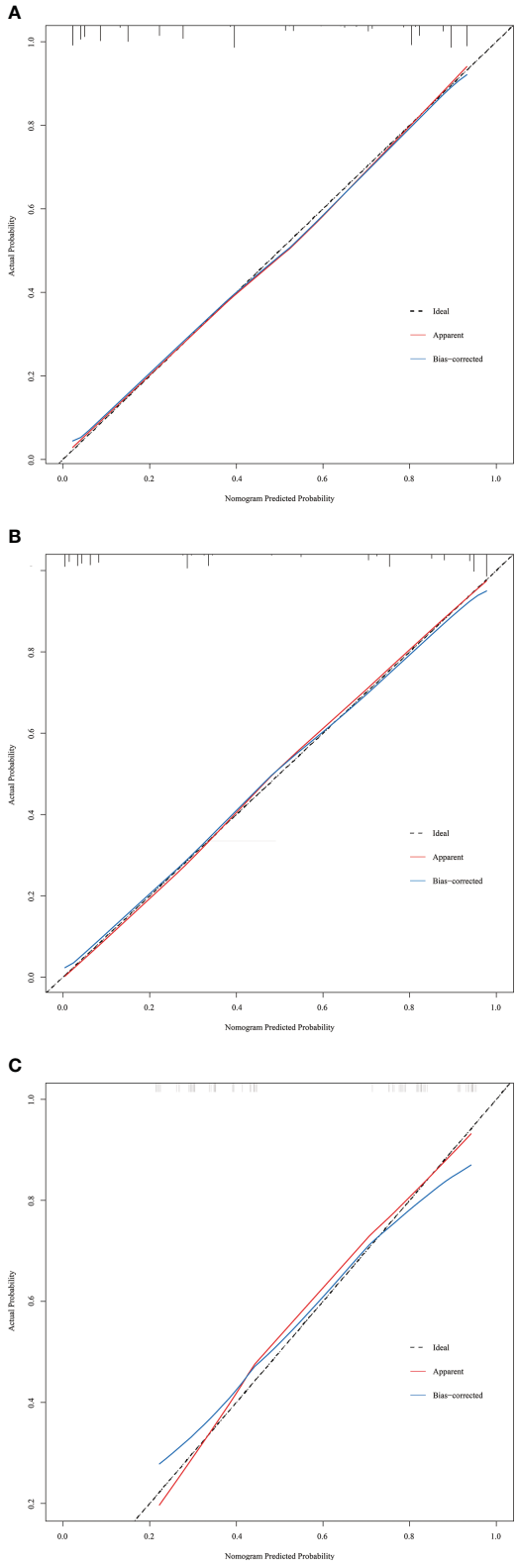


FIGURE 5
Calibration curves of the prediction nomogram in the training cohort **(A)**, internal validation cohort **(B)** and external validation cohort **(C)**. The x-axis represents the nomogram-predicted probability, and the y-axis represents the actual probability of HRPNs. The black pointed line represents the ideal curve, the red solid line represents the apparent curve (non-correction), and the blue solid line represents the bias-correction curve by bootstrapping (B = 1000 repetitions). HRPNs, High-risk pulmonary nodules.

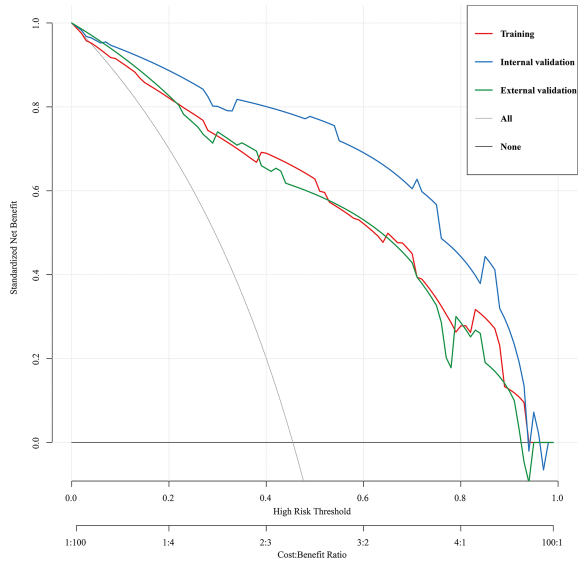


FIGURE 6
Decision curve analysis for the HRPNs nomogram in the training and validation cohorts. The y-axis measures the net benefit, the black line represents the assumption of HRPNs-none-patients, the gray line represents the assumption of HRPNs-all-patients, the red line represents the training cohort, the blue line represents the internal validation cohort, the green lines represents the external validation cohort. HRPNs, High-risk pulmonary nodules.

Regarding tumor markers, we confirmed the diagnostic value of CEA in predicting the risk of HRPNs. Yuan et al. demonstrated that the sensitivities of CEA, CYFRA21-1, and NSE for diagnosing lung cancer were 56.5%, 56.1%, and 19.1%, respectively (16). Molina et al. reported that the diagnostic sensitivity and specificity could reach 88.5% and 82%, respectively, combined with six hematological tumor markers (CA153, CEA, CYFRA21- 1, NSE, Pro-GRP, and SCC) (36). After univariable and multivariable logistic regression analysis of blood tumor markers (CEA, CYFRA21-1, SCC, NSE, and Pro-GRP), only CEA was an independent factor for HRPNs (OR=1.701, 95% CI: 1.072-2.701, P=0.024). This may be due to the high sensitivity of some tumor markers to certain pathological lung cancer types. For example, CYFRA21-1 and SCC are more sensitive to lung squamous cell carcinoma (37), while Pro-GRP and NSE are more sensitive to small cell lung carcinoma (38). Blood CEA levels can be used as a combined index to predict the occurrence of lung cancer, which has good universality. ROC analysis of blood CEA levels in the included patients showed an AUC value of 0.641 with an optimal cutoff value of 1.965 ng/mL (similar to that reported in the study by Zheng et al), indicating that blood CEA levels have better predictive performance for HRPNs than other tumor markers.

It is noteworthy that the multivariable logistic regression analysis showed that the imaging features of lung cancer (lobulation, spiculation, BEAS or cavity sign pleural traction, and vascular convergence) were not independent risk factors for HRPNs, in contrast with previous research findings (39, 40). Based on our analysis, we believe that the imaging features of lung cancer have good sensitivity and specificity for the benign and malignant judgment of PNs; however, it is difficult to use these indicators for evaluating the malignant degree of PNs. This is because, regardless of whether the degree of malignancy is high or low, the above-mentioned imaging features may appear as malignant PNs, which will undoubtedly reduce the accuracy of the judgment of these indicators. According to Liang and Feng's study, pulmonary nodules with final pathology of benign lesions, precancerous lesions, and lung adenocarcinoma can exhibit spiculation on the thoracic CT sign (41, 42). This infers subtly that the imaging characteristics of lung cancer have a low degree of specificity.

Several models for predicting benign and malignant PNs have been reported (40, 43); however, no model has focused on predicting the degree of PN malignancy. The advantages of our method over previously published predictive models are as follows: First, we visualized this predictive model, built a nomogram, and subsequently constructed an operating interface for our nomogram on the web page (http://lungnodules.shinyapps.io/Predict_QiuJH/), which greatly optimized the computational process and enhanced the practicality of using this model in clinical practice. Secondly, we included patients with PNs with a maximum diameter of ≤3 cm on imaging into the model as much as possible and built a preoperative risk prediction model to predict the degree of PN malignancy. Furthermore, this model can be applied to the majority of patients with PNs, greatly increasing its usability. Third, we developed a predictive model that used preoperative imaging features and blood CEA levels to predict the risk of developing HRPNs regardless of the patient's final histopathological outcome. Preoperative non-surgical biopsy (such as CT-guided lung biopsy and bronchoscopy) can be used to obtain PN histopathology, their invasive nature and potential risks limits their clinical usefulness. Therefore, this model can predict the high-risk probability of PNs in patients before surgery to guide the selection of surgical methods to reduce surgical trauma and shorten patients' hospitalization time. Fourth, we used CTR to predict the degree of PN malignancy, which has not been observed in previous studies. Furthermore, we employed internal- external validation in model validation, leveraging patient data from both centers to validate our model. Final validation results also demonstrated the good prediction accuracy of our model. Lastly, we utilized DCA to assess the clinical utility of predictive nomograms. Based on the decision curve, the nomogram model had a clear net benefit (HRPN incidence across all cohorts in this study) in the threshold range of 45-60%, suggesting that the nomogram has a high utility in the clinic.

The clinical predictive nomogram constructed can help thoracic surgeons use preoperative imaging and tumor marker features to assess the probability of HRPNs preoperatively to make better clinical decisions. For patients with HRPNs, standard lobectomy and systematic lymph node dissection can be used to reduce the recurrence rate of lung cancer after surgery, while sublobar resection (lung wedge resection and anatomic segmental lung resection) can be performed in patients with LRPNs to avoid overtreatment and better protect lung function. Related studies from Japan have shown that segmentectomy can be utilized as the standard treatment for PNs with CTR < 0.5 and tumor diameter < 3cm. Additionally, for pulmonary nodules with CTR > 0.5, the recurrence-free survival and disease-free survival of patients were

improved by lobectomy. This outcome matched the predictions made by our prediction model (35, 44). In our hospital, after preoperative evaluation of the patient's PNs, we performed surgical treatment according to the above treatment strategies, which ensured precise patient treatment, reduced the hospitalization time and patient costs, and realized lung cancer-enhanced recovery strategies.

There are several limitations to the present study. First, this retrospective study may limit the generalizability of our predictive nomogram, and uncontrolled confounding factors may also arise. Furthermore, this predictive model has undergone both internal and external validation; hence, selection biases present in the training cohort may be present in the internal validation cohort. Thus, multiple external validations in a more central setting are required to determine if the nomogram is suitable for widespread use in other populations. Finally, we did not include factors that may be related to HRPNs in our study, such as nodule doubling time, blood tissue polypeptide antigen level, and CA125 level. These factors were absent from our database and should be explored in future studies.

Conclusion

Based on PNs preoperative imaging characteristics and blood tumor marker levels, a clinical nomogram to predict the probability of HRPNs was established, and a good prediction effect was achieved. The probability of HRPNs in patients with PNs can be assessed using this nomogram. Different treatment strategies can be applied to HRPNs and LRPNs to achieve precise treatment and accelerated patient recovery.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by The Ethics Committee of Qilu Hospital of Shandong University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

Conception and design: JQ and HT. Administrative support: HT and WY. Provision of study materials or patients: JQ, RL and XM. Collection and assembly of data: JQ, RL, YW, XM and BL. Data analysis and interpretation: JQ, RL and CQ. All authors contributed to the article and approved the submitted version.

Funding

This work was funded by the National Key Research and Development Program (2021YFC2500900, 2021YFC2500904 and 2021YFC2500905), and Natural Science Foundation of Shandong Province (ZR2021LSW006).

Acknowledgments

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.1196883/full#supplementary-material

References

1. Siegel R, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2021. *CA: Cancer J Clin* (2021) 71(1):7–33. doi: 10.3322/caac.21654

2. Duma N, Santana-Davila R, Molina JR. Non-small cell lung cancer: epidemiology, screening, diagnosis, and treatment. *Mayo Clinic Proc* (2019) 94 (8):1623–40. doi: 10.1016/j.mayocp.2019.01.013

3. Hansell D, Bankier AA, MacMahon H, McLoud TC, Müller NL, Remy J. Fleischner society: glossary of terms for thoracic imaging. *Radiology* (2008) 246 (3):697–722. doi: 10.1148/radiol.2462070712

4. Naidich D, Bankier AA, MacMahon H, Schaefer-Prokop CM, Pistolesi M, Goo JM, et al. Recommendations for the management of subsolid pulmonary nodules

detected at CT: a statement from the fleischner society. *Radiology* (2013) 266(1):304–17. doi: 10.1148/radiol.12120628

5. Nicholson AG, Tsao MS, Beasley MB, Borczuk AC, Brambilla E, Cooper WA, et al. The 2021 WHO classification of lung tumors: impact of advances since 2015. *J Thorac Oncol* (2022) 17(3):362–87. doi: 10.1016/j.jtho.2021.11.003

6. Tsao M-S, Nicholson AG, Maleszewski JJ, Marx A, Travis WD. Introduction to 2021 WHO classification of thoracic tumors. *J Thorac Oncol* (2022) 17(1):e1–4. doi: 10.1016/j.jtho.2021.09.017

7. Yotsukura M, Asamura H, Motoi N, Kashima J, Yoshida Y, Nakagawa K, et al. Long-term prognosis of patients with resected adenocarcinoma *in situ* and minimally invasive adenocarcinoma of the lung. *J Thorac Oncol* (2021) 16(8):1312–20. doi: 10.1016/j.jtho.2021.04.007

8. Zhang Y, Deng C, Fu F, Ma Z, Wen Z, Ma X, et al. Excellent prognosis of patients with invasive lung adenocarcinomas during surgery misdiagnosed as atypical adenomatous hyperplasia, adenocarcinoma in situ, or minimally invasive adenocarcinoma by frozen section. *Chest* (2021) 159(3):1265–72. doi: 10.1016/j.chest.2020.10.076

9. Kishi N, Ito M, Miyata Y, Kanai A, Handa Y, Tsutani Y, et al. Intense expression of EGFR L858R characterizes the micropapillary component and L858R is associated with the risk of recurrence in pN0M0 lung adenocarcinoma with the micropapillary component. *Ann Surg Oncol* (2020) 27(3):945–55. doi: 10.1245/s10434-019-07854-2

10. Moreira AL, Ocampo PSS, Xia Y, Zhong H, Russell PA, Minami Y, et al. A grading system for invasive pulmonary adenocarcinoma: a proposal from the international association for the study of lung cancer pathology committee. *J Thorac Oncol* (2020) 15(10):1599–610. doi: 10.1016/j.jtho.2020.06.001

11. Yotsukura M, Asamura H, Suzuki S, Asakura K, Yoshida Y, Nakagawa K, et al. Prognostic impact of cancer-associated active fibroblasts and invasive architectural patterns on early-stage lung adenocarcinoma. *Lung Cancer (Amsterdam Netherlands)* (2020) 145:158–66. doi: 10.1016/j.lungcan.2020.04.023

12. Succony L, Rassl DM, Barker AP, McCaughan FM, Rintoul RC. Adenocarcinoma spectrum lesions of the lung: detection, pathology and treatment strategies. *Cancer Treat Rev* (2021) 99:102237. doi: 10.1016/j.ctrv.2021.102237

13. Travis WD. Lung cancer pathology: current concepts. *Clin Chest Med* (2020) 41 (1):67–85. doi: 10.1016/j.ccm.2019.11.001

14. Cao W, Ma J, Jiang X, Gao G. Afatinib targeted therapy affects the immune function and serum levels of EGFR and gastrin-releasing peptide precursor (pro-GRP) in patients with non-Small-Cell lung cancer (NSCLC). *Dis Markers* (2022) 2022:2869531. doi: 10.1155/2022/2869531

15. Dal Bello MG, Filiberti RA, Alama A, Orengo AM, Mussap M, Coco S, et al. The role of CEA, CYFRA21-1 and NSE in monitoring tumor response to nivolumab in advanced non-small cell lung cancer (NSCLC) patients. *J Transl Med* (2019) 17(1):74. doi: 10.1186/s12967-019-1828-0

16. Yuan J, Sun Y, Wang K, Wang Z, Li D, Fan M, et al. Development and validation of reassigned CEA, CYFRA21-1 and NSE-based models for lung cancer diagnosis and prognosis prediction. *BMC Cancer* (2022) 22(1):686. doi: 10.1186/s12885-022-09728-5

17. Zhou W, Yang Y, Wang Z, Liu Y, Lari Najafi M. Impact of HSP90α, CEA, NSE, SCC, and CYFRA21-1 on lung cancer patients. *J Healthc Eng* (2021) 2021:6929971. doi: 10.1155/2021/6929971

18. Suzuki K, Koike T, Asakawa T, Kusumoto M, Asamura H, Nagai K, et al. A prospective radiological study of thin-section computed tomography to predict pathological noninvasiveness in peripheral clinical IA lung cancer (Japan clinical oncology group 0201). *J Thorac Oncol* (2011) 6(4):751–6. doi: 10.1097/JTO.0b013e31821038ab

19. Jin C, Cao J, Cai Y, Wang L, Liu K, Shen W, et al. A nomogram for predicting the risk of invasive pulmonary adenocarcinoma for patients with solitary peripheral subsolid nodules. *J Thorac Cardiovasc Surg* (2017) 153(2):462–469.e1. doi: 10.1016/j.jtcvs.2016.10.019

20. Wang R, Dai W, Gong J, Huang M, Hu T, Li H, et al. Development of a novel combined nomogram model integrating deep learning-pathomics, radiomics and immunoscore to predict postoperative outcome of colorectal cancer lung metastasis patients. *J Hematol Oncol* (2022) 15(1):11. doi: 10.1186/s13045-022-01225-3

21. Obuchowski NA, Bullen JA. Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. *Phys Med Biol* (2018) 63 (7):07tr01. doi: 10.1088/1361-6560/aab4b1

22. Nattino G, Pennell ML, Lemeshow S. Assessing the goodness of fit of logistic regression models in large samples: a modification of the hosmer-lemeshow test. *Biometrics* (2020) 76(2):549–60. doi: 10.1111/biom.13249

23. Rosenfeld JP, Donchin E. Resampling (bootstrapping) the mean: a definite do. *Psychophysiology* (2015) 52(7):969–72. doi: 10.1111/psyp.12421

24. Vickers AJ, Holland F. Decision curve analysis to evaluate the clinical benefit of prediction models. *Spine J* (2021) 21(10):1643–8. doi: 10.1016/j.spinee.2021.02.024

25. Mazzone PJ, Lam L. Evaluating the patient with a pulmonary nodule: a review. *Jama* (2022) 327(3):264–73. doi: 10.1001/jama.2021.24287

26. Batchelor TJP, Ljungqvist O. A surgical perspective of ERAS guidelines in thoracic surgery. *Curr Opin In Anaesthesiology* (2019) 32(1):17–22. doi: 10.1097/ACO.0000000000000685

27. Heuvelmans MA, Walter JE, Vliegenthart R, van Ooijen PMA, De Bock GH, de Koning HJ, et al. Disagreement of diameter and volume measurements for pulmonary nodule size estimation in CT lung cancer screening. *Thorax* (2018) 73(8):779–81. doi: 10.1136/thoraxjnl-2017-210770

28. Bak SH, Lee HY, Kim JH, Um SW, Kwon OJ, Han J, et al. Quantitative CT scanning analysis of pure ground-glass opacity nodules predicts further CT scanning change. *Chest* (2016) 149(1):180–91. doi: 10.1378/chest.15-0034

29. Tamura M, Shimizu Y, Yamamoto T, Yoshikawa J, Hashizume Y. Predictive value of one-dimensional mean computed tomography value of ground-glass opacity on high-resolution images for the possibility of future change. *J Thorac Oncol* (2014) 9 (4):469–72. doi: 10.1097/JTO.0000000000000117

30. Eguchi T, Kondo R, Kawakami S, Matsushita M, Yoshizawa A, Hara D, et al. Computed tomography attenuation predicts the growth of pure ground-glass nodules. *Lung Cancer (Amsterdam Netherlands)* (2014) 84(3):242–7. doi: 10.1016/j.lungcan.2014.03.009

31. Ikeda K, Awai K, Mori T, Kawanaka K, Yamashita Y, Nomori H. Differential diagnosis of ground-glass opacity nodules: CT number analysis by three-dimensional computerized quantification. *Chest* (2007) 132(3):984–90. doi: 10.1378/chest.07-0793

32. Koezuka S, Sano A, Azuma Y, Sakai T, Matsumoto K, Shiraga N, et al. Combination of mean CT value and maximum CT value as a novel predictor of lepidic predominant lesions in small lung adenocarcinoma presenting as solid nodules. *Sci Rep* (2022) 12(1):5450. doi: 10.1038/s41598-022-09173-1

33. Asamura H, Hishida T, Suzuki K, Koike T, Nakamura K, Kusumoto M, et al. Radiographically determined noninvasive adenocarcinoma of the lung: survival outcomes of Japan clinical oncology group 0201. *J Thorac Cardiovasc Surg* (2013) 146(1):24–30. doi: 10.1016/j.jtcvs.2012.12.047

34. Sun K, You A, Wang B, Song N, Wan Z, Wu F, et al. Clinical T1aN0M0 lung cancer: differences in clinicopathological patterns and oncological outcomes based on the findings on high-resolution computed tomography. *Eur Radiol* (2021) 31 (10):7353–62. doi: 10.1007/s00330-021-07865-2

35. Ito H, Suzuki K, Mizutani T, Aokage K, Wakabayashi M, Fukuda H, et al. Long-term survival outcome after lobectomy in patients with clinical T1 N0 lung cancer. *J Thorac Cardiovasc Surg* (2020) S0022-5223(20)30054-4. doi: 10.1016/j.jtcvs.2019.12.072

36. Molina R, Marrades RM, Augé JM, Escudero JM, Viñolas N, Reguart N, et al. Assessment of a combined panel of six serum tumor markers for lung cancer. *Am J Respir Crit Care Med* (2016) 193(4):427–37. doi: 10.1164/rccm.201404-0603OC

37. Liu L, Liu B, Zhu LL, Zhang W, Li Y. Clinical significance of CYFRA21-1, scc-Ag and telomerase activity in serum and pleural effusion of patients with squamous-cell lung cancer. *Bioanalysis* (2012) 4(19):2367–74. doi: 10.4155/bio.12.203

38. Wang L, Wang D, Zheng G, Yang Y, Du L, Dong Z, et al. Clinical evaluation and therapeutic monitoring value of serum tumor markers in lung cancer. *Int J Biol Markers* (2016) 31(1):e80–7. doi: 10.5301/jbm.5000177

39. Mao L, Chen H, Liang M, Li K, Gao J, Qin P, et al. Quantitative radiomic model for predicting malignancy of small solid pulmonary nodules detected by low-dose CT screening. *Quantitative Imaging In Med Surg* (2019) 9(2):263–72. doi: 10.21037/qims.2019.02.02

40. Wang L, Shen W, Xi Y, Liu S, Zheng D, Jin C. Nomogram for predicting the risk of invasive pulmonary adenocarcinoma for pure ground-glass nodules. *Ann Thorac Surg* (2018) 105(4):1058–64. doi: 10.1016/j.athoracsur.2017.11.012

41. Feng B, Chen X, Chen Y, Yu T, Duan X, Liu K, et al. Identifying solitary granulomatous nodules from solid lung adenocarcinoma: exploring robust image features with cross-domain transfer learning. *Cancers (Basel)* (2023) 15(3):892. doi: 10.3390/cancers15030892

42. Liang ZR, Lv FJ, Fu BJ, Lin RY, Li WJ, Chu ZG. Reticulation sign on thin-section CT: utility for predicting invasiveness of pure ground-glass nodules. *AJR Am J Roentgenol* (2023) 1–10. doi: 10.2214/AJR.22.28892

43. Yu X, Zhang S, Xu J, Huang Y, Luo H, Huang C, et al. Nomogram using CT radiomics features for differentiation of pneumonia-type invasive mucinous adenocarcinoma and pneumonia: multicenter development and external validation study. *AJR Am J Roentgenol* (2023) 220(2):224–34. doi: 10.2214/AJR.22.28139

44. Saji H, Okada M, Tsuboi M, Nakajima R, Suzuki K, Aokage K, et al. Segmentectomy versus lobectomy in small-sized peripheral non-small-cell lung cancer (JCOG0802/WJOG4607L): a multicentre, open-label, phase 3, randomised, controlled, non-inferiority trial. *Lancet* (2022) 399(10335):1607–17. doi: 10.1016/S0140-6736(21)02333-3

frontiers | Frontiers in Oncology

TYPE Case Report
PUBLISHED 28 June 2023
DOI 10.3389/fonc.2023.1182558

Check for updates

OPEN ACCESS

EDITED BY
Hongda Liu,
Nanjing Medical University, China

REVIEWED BY
Alessandro Morabito,
G. Pascale National Cancer Institute Foundation (IRCCS), Italy
Petros Christopoulos,
Heidelberg University Hospital, Germany

*CORRESPONDENCE
Lionel Michaux
lionelmichaux89@gmail.com

RECEIVED 08 March 2023
ACCEPTED 30 May 2023
PUBLISHED 28 June 2023

CITATION
Michaux L, Perrier A, Mehlman C, Alshehhi H, Dubois A, Lacave R, Coulet F, Cadranel J and Fallet V (2023) Therapeutic strategies to overcome EGFR mutations as acquired resistance mechanism in ALK-rearranged non-small-cell lung cancer: Case Reports.
Front. Oncol. 13:1182558.
doi: 10.3389/fonc.2023.1182558

COPYRIGHT
© 2023 Michaux, Perrier, Mehlman, Alshehhi, Dubois, Lacave, Coulet, Cadranel and Fallet. 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.

Therapeutic strategies to overcome EGFR mutations as acquired resistance mechanism in ALK-rearranged non-small-cell lung cancer: Case Reports

Lionel Michaux[1]*, Alexandre Perrier[2], Camille Mehlman[1], Hussa Alshehhi[3], Antonin Dubois[4], Roger Lacave[2], Florence Coulet[2], Jacques Cadranel[1] and Vincent Fallet[1]

[1]Department of Pulmonology and Thoracic Oncology, Assistance Publique Hôpitaux de Paris, Hôpital Tenon and Groupe de Recherche Clinique 4 (GRC 4), Theranoscan, Sorbonne Université, Paris, France, [2]Genetics Department, Assistance Publique Hôpitaux de Paris, Hôpital de la Pitié-Salpêtrière and Sorbonne Université, Paris, France, [3]Pathology Department, Assistance Publique Hôpitaux de Paris, Hôpital Tenon and Sorbonne Université, Paris, France, [4]Department of Pharmacy, Assistance Publique Hôpitaux de Paris, Hôpital Tenon and Sorbonne Université, Paris, France

Introduction: ALK tyrosine kinase inhibitors (ALK TKIs) have improved prognosis in *ALK*-rearranged (*ALK*$^+$) non-small-cell lung cancer (NSCLC). However, drug resistance mechanisms occur inevitably during the course of treatment leading to disease progression. Activation of epidermal growth factor receptor (EGFR) bypass signaling pathway is an uncommon cause of acquired resistance to ALK TKIs.

Method: We present two patients with *EML4-ALK* rearranged NSCLC, developing an acquired *EGFR* resistance mutation after receiving multiple lines of ALK TKIs.

Results: While preclinical models have showed encouraging data, there is a critical need for clinical studies on treatment strategies to overcome this drug resistance. Three real-life therapeutic approaches were used in this report: i) using brigatinib, an inhibitor targeting both ALK and EGFR tyrosine kinases; ii) combining two ALK TKIs together; and iii) delivering doublet platinum chemotherapy. In case 1, time to treatment failure (TTF) was 9.5 months with brigatinib; in case 2, TTF was 10 months with combined TKIs (osimertinib and brigatinib), whereas TTF with chemotherapy was only 2 months. Tolerability profile TKIs combotherapy was acceptable.

Conclusion: These case reports underline the therapeutic complexity of *EGFR*-acquired resistance mutation in *ALK*$^+$ NSCLC and offers some leads to solve this real-life clinical challenge.

KEYWORDS

non-small cell lung cancer, ALK rearrangement, tyrosine kinase inhibitors, resistance mutation, EGFR

Introduction

Giving an ALK tyrosine kinase inhibitor (ALK TKI) is the standard of care for first-line treatment of patients with advanced *ALK* rearrangement driven non-small-cell lung cancer (*ALK*+ NSCLC).

However, nearly all patients develop acquired resistance to ALK TKIs. Resistance can be divided into two categories: in-target and off-target resistance mechanisms. Among these, the activation of alternative bypass signaling pathways have been described. We report two cases of rare resistance mechanisms involving EGFR pathway occurring after sequential treatment by ALK TKIs. We also aimed to describe our therapeutic approaches to overcome these EGFR-driven resistances.

Cases presentation

Case 1

A then 38-year-old, non-smoker woman, presented to the hospital with chronic cough and asthenia in June 2016. Computed tomography (CT) scan showed a middle lobe consolidation with infradiaphragmatic lymphatic nodes. The patient was subsequently diagnosed with a metastatic lung adenocarcinoma. ALK immunohistochemistry (IHC) was positive with a strong intensity score (clone D5F3), and fluorescence *in situ* hybridization (FISH) confirmed an *EML4-ALK* fusion (Vysis LSI ALK Break Apart FISH Probe Kit from Abbott Molecular®). No additional somatic gene alteration, especially *EGFR* mutation, was found (EGFR genotyping was done using RT-PCR on QX200 system, BIO RAD® system and TaqMan® probes from Life Technologies® for exons 18, 20, and 21 and fragment analysis by capillary gel electrophoresis on 3130 ABI system™ from Applied Technologies® for exon 19). As first-line treatment, the patient received crizotinib (250 mg twice daily) for 13 months until brain magnetic resonance imaging (MRI) revealed brain metastases. After 5 months of treatment with alectinib (600 mg twice daily), CT scan showed progression of lung lesions (Figure 1). Next-generation sequencing (NGS) of tumoral lung tissue DNA revealed an *EGFR* L861Q mutation (c.2582T>A, p.(Leu861Gln))—variant allelic frequency (VAF) of 1.7% (Tumor Hotspot MASTR™ Plus, Multiplicom®). While *EML4-ALK* rearrangement was still found, no ALK-dependent resistance mutation was identified. Alectinib was thus discontinued, and brigatinib (90 mg once daily for 1 week, then 180 mg once daily) was given to the patient. The CT scan showed reduction in lung opacities after 8 weeks. Nevertheless, a differential response was observed after 6 months of treatment with maintained lung response but a progression of brain lesions on MRI. Brigatinib was then increased to 240 mg (once daily), and with this dose regimen, the patient had grade 1 myalgia and increased levels of blood creatinine phosphokinases, aspartate aminotransferases, and lipases. After 3 months with brigatinib increased dose, brain metastases progressed, and brigatinib was discontinued. No mutation was shown on plasma circulating tumor DNA, and the patient received a combination of alectinib 600 mg (twice daily) and osimertinib 80 mg (once daily) as fourth-line therapy, which was discontinued after 2 months because of progression. Seven additional lines of treatment were given to the patient, including carboplatin pemetrexed doublet chemotherapy (fifth line) and lorlatinib (sixth line), with 12 and 8 months of response to treatment, respectively. ALK resistance compound mutations (*ALK G1202R* and *ALK G1269A*) were first detected after lorlatinib treatment and were still identified until the patient's death 74 months after the diagnosis.

Case 2

In June 2020, a 42-year-old, non-smoker woman was admitted to the hospital for chronic cough and dyspnea. Imaging tests showed a right upper lobe lung mass with secondary bones and liver lesions. CT-guided liver biopsies revealed a muco-epidermoid lung carcinoma. An immunohistological analysis revealed that TTF1 was negative, while the tumor cells without mucus were positive for p40. Tumor cells were also strongly positive for ALK IHC (clone D5F3). *EML4-ALK* fusion variant 1 was detected by RNA sequencing (Archer® FusionPlex™ Lung panel). No additional somatic gene alteration (such as TP53, KRAS, and KEAP1) was shown in the NGS DNA analysis (SOPHIA Solid Tumor Solution™, SOPHIA Genetics®). Alectinib 600 mg twice daily was started, and the patient experienced a partial response for 12 months. When imaging revealed progression with pelvic bone metastases, lorlatinib was initiated with a 100-mg daily dose regimen. After 2 months, the CT scan revealed new lung opacities that were attributed to a lymphangitic carcinomatosis (Figure 2). Lorlatinib was discontinued, and a carboplatin–paclitaxel doublet chemotherapy was started but was rapidly stopped because of radiological evidences of lung progression. Meanwhile, DNA-based NGS (AmpliSeq™ for Illumina® Focus Panel) in the tissue specimen from bronchus biopsies performed during progression under lorlatinib revealed an exon 20 *EGFR* T790M mutation (VAF, 1.3%) and the RNA sequencing the *EML4-ALK* rearrangement. The patient then started on a combination of brigatinib (90 mg once daily for 1 week, then 180 mg once daily) and osimertinib (80 mg once daily). Partial response in the lung was achieved after 6 weeks. The patient experienced, as treatment emergent adverse events, grade 2 vomiting and grade 1 thrombopenia. The patient remained on these combined therapies for 10 months until the CT scan showed pleural progression. As additional line of treatment, brigatinib and pemetrexed were given to the patient with a response of 6 months. The patient finally died 33 months after the diagnosis.

Discussion

We reported here two cases of well-documented *EGFR* mutations linked to acquired resistance to ALK TKIs in *ALK*+ NSCLC. To our knowledge, neither L861Q nor T790M *EGFR* mutants have been described yet as EGFR bypass ALK resistance mechanism. These *EGFR* mutations were not detected at diagnosis

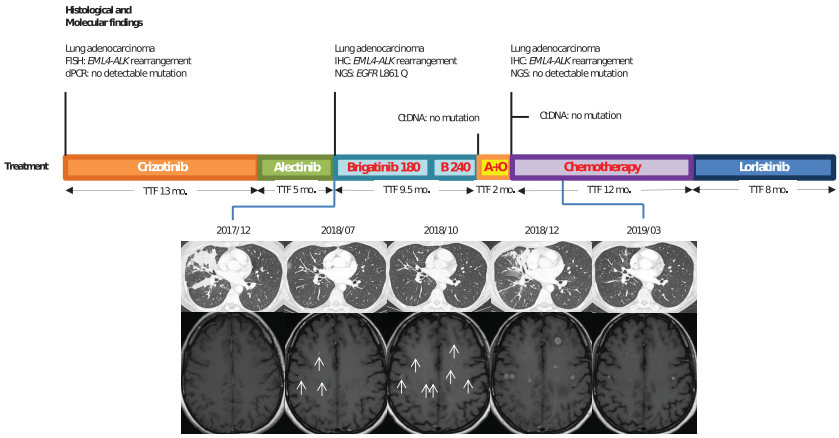


FIGURE 1
Case 1: timeline illustrating the changes in therapeutic regimen in correlation with molecular and radiological findings. FISH, fluorescence *in situ* hybridization; dPCR, digital polymerase chain reaction; TTF, time to treatment failure; Mo., months; ctDNA, circulating tumor DNA; IHC, immunohistochemestry; NGS, next-generation sequencing; Brigatinib 180, brigatinib with a 180-mg daily dose regimen; B240, brigatinib with a 240-mg daily dose regimen; A+O, combined treatment with alectinib and osimertinib. In red, the treatments used when *EGFR* resistance mutation was identified.

even using different highly sensitive detection methods, making the possibility of *de novo* co-alterations unlikely.

Most frequently, *ALK*+ NSCLCs are adenocarcinomas, but in case 2, it was a very rare subtype of NSCLC: mucoepidermoid carcinoma. *EML4-ALK* rearrangements in this rare subtype have been described in small series, including a case report with durable response to alectinib (1, 2).

Initially described in preclinical studies, the activation of EGFR signaling was identified as an infrequent post-ALK TKIs bypass resistance mechanism (3). In the largest clinical cohort that evaluated the mechanisms of resistance post-ALK TKIs, the most frequent alteration reported was the acquisition of an *ALK* resistance mutation in 56% of biopsies post second-generation ALK inhibitors, and no *EGFR* acquired resistance was identified is this cohort (4). In smaller cohorts, *EGFR* mutation as a mechanism of resistance to ALK TKIs has been previously shown (5–7). In the four described patients, activating *EGFR* mutations (L858R or exon 19 deletion) were found in tissue biopsies.

Therapeutic strategy proposed in this context has not been clarified. In the present report, we experienced three clinical therapeutic strategies trying to overcome the activation of EGFR signaling pathway. As a first strategy, we switched alectinib for brigatinib, which was of interest because of brigatinib dual ALK and EGFR activity (8). In cellular assays, brigatinib has been reported to have a substantial activity against *EGFR* exon 19 deletion, with potency only sevenfold reduced compared to *ALK* (8). Moreover, preclinical (9) and clinical (10) data demonstrated that a combined targeted therapy of brigatinib and cetuximab could be beneficial to overcome the triple *EGFR* mutant (*EGFR T790M* and *cis-C797S*) resistance to osimertinib. However, brigatinib exhibits more modest activity against *EGFR L858R* or variants with *T790M* mutation (8), and there is no preclinical data supporting such an activity in case of rare *L861Q EGFR* mutation harbored by our patient. This strategy of using brigatinib to overcome *EGFR*-acquired resistance raises some limit in our first case with EGFR co-alteration. Even if no *ALK* resistance mutation was detected with the *EGFR* mutation after alectinib, the impact of brigatinib on this bypass resistance mutation remains uncertain, since this TKI also overcomes *ALK*-dependent resistance post-alectinib (11, 12). Finally, because of the isolated and asymptomatic brain progression, we chose to escalate

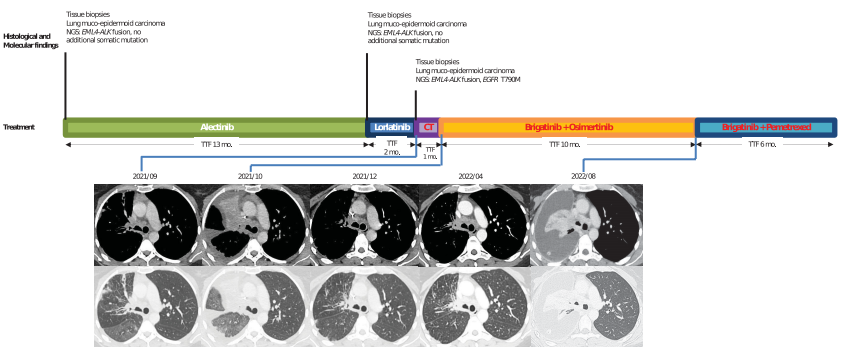


FIGURE 2
Case 2: timeline illustrating the changes in therapeutic regimen in correlation with molecular and radiological findings. NGS, next-generation sequencing; TTF, time to treatment failure; Mo., months; CT, chemotherapy. In red, the treatments used when *EGFR* resistance mutation was identified.

brigatinib dose regimen to 240 mg daily. Despite a good tolerance, this strategy appeared to be ineffective. In the Phase 2 ALTA-2 trial, 13 patients escalated their brigatinib dosage from 180 to 240 mg daily after experiencing disease progression. While this higher dosage demonstrated an acceptable safety profile, the clinical benefit was disappointing, as no confirmed tumoral response was observed, and the progression-free survival was <2 months (13).

As an alternative therapeutic strategy, we experienced a combination treatment with both ALK and EGFR TKIs (alectinib + osimertinib in case 1 and brigatinib + osimertinib in case 2). No grade 3 toxicity was observed in both cases. An encouraging response was reported in case 2 in the fourth line after a rapid progression with both lorlatinib then chemotherapy. Concurrent inhibition of both EGFR and ALK is therapeutically effective in all of the ALK-resistant preclinical models, which have acquired resistance through EGFR pathway activation (3, 14) and has been described to be safe in several case reports in EGFR-resistant lung cancer with acquired ALK fusion co-alteration (15). However, in case 1, patient progressed dramatically after 2 months of combined therapies perhaps because the tumor was no longer EGFR dependent at this time.

Finally, a platinum-based doublet chemotherapy might be beneficial after failure of second- or third-generation ALK TKI in *ALK* rearranged NSCLC (16). Interestingly, our second patient rapidly progressed after 1 month of chemotherapy, while a prolonged (12 months) partial response occurred in the first patient. Nevertheless, this last result must be interpreted with caution given that the *EGFR* mutation was not present after treatment with brigatinib.

In conclusion, we reported two cases of *EGFR* mutant emerging after at least two lines of ALK TKIs in *ALK*-rearranged NSCLC. In order to overcome drug resistance, we adopted different strategies with variable efficacy. These case reports highlight both the complexity of drug resistance mechanisms and the therapeutic challenges in developing strategies to overcome drug resistance in oncogenic-driven NSCLC.

Data availability statement

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

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

LM: Conceptualization, Methodology, Investigation, Writing – Original draft. AP: Writing – Reviewing and Editing. CM: Writing – Reviewing and Editing. HA: Writing – Reviewing and Editing. AD: Writing – Reviewing and Editing. RL: Writing – Reviewing and Editing. FC: Writing – Reviewing and Editing. JC: Writing – Reviewing and Editing, Supervision. VF: Conceptualization, Methodology, Investigation, Writing – Original draft. All authors contributed to the article and approved the submitted version.

Conflict of interest

LM received support for attending meetings and travels from Bristol Myers Squibb and Olympus, all outside of the submitted work. CM received support for attending meeting and travel from ISIS Medical; outside of the submitted work. AD received support for attending meetings and travels from Astellas, Janssen, Bristol Myers Squibb, and Amgen, all outside of the submitted work. FC received payment for her institution for educational events from Astra Zeneca, outside of the submitted work. JC received grants for his institution from Pfizer, AbbVie, Sanofi, and Sophia Genetics; payment for participation to boards of experts from Amgen, Aztra Zeneca, Boehringer Ingelheim, Bristol Myers Squibb, Janssen, Merck Sharp and Dohme, Novartis, Pfizer, and Takeda; all outside of the submitted work. VF received payment for lectures from Pfizer, Takeda, Bristol Myers Squibb, Aztra Zeneca, Roche, and Boehringer Ingelheim; support for attending meetings and travels from Takeda, Janssen, Pfizer, Aztra Zeneca, and Novartis; all outside of the submitted work.

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. Huo Z, Wu H, Li J, Li S, Wu S, Liu Y, et al. Primary pulmonary mucoepidermoid carcinoma: histopathological and molecular genetic studies of 26 cases. *PloS One* (2015) 10(11):e0143169. doi: 10.1371/journal.pone.0143169

2. Sakatani T, Masuda Y, Morikawa T, Usui K. Anaplastic lymphoma kinase-positive lung cancer with mucoepidermoid carcinoma differentiation: a case report. *CRO* (2020) 13(2):1037–41. doi: 10.1159/000510042

3. Sasaki T, Koivunen J, Ogino A, Yanagita M, Nikiforow S, Zheng W, et al. A novel ALK secondary mutation and EGFR signaling cause resistance to ALK kinase inhibitors. *Cancer Res* (2011) 71(18):6051–60. doi: 10.1158/0008-5472.CAN-11-1340

4. Gainor JF, Dardaei L, Yoda S, Friboulet L, Leshchiner I, Katayama R, et al. Molecular mechanisms of resistance to first- and second-generation ALK inhibitors in ALK-rearranged lung cancer. *Cancer Discovery* (2016) 6(10):1118–33. doi: 10.1158/2159-8290.CD-16-0596

5. Doebele RC, Pilling AB, Aisner DL, Kutateladze TG, Le AT, Weickhardt AJ, et al. Mechanisms of resistance to crizotinib in patients with ALK gene rearranged non-small cell lung cancer. *Clin Cancer Res* (2012) 18(5):1472–82. doi: 10.1158/1078-0432.CCR-11-2906

6. Wang W, Jiang X, Song Z, Zhang Y. Patients harboring EGFR mutation after primary resistance to crizotinib and response to EGFR-tyrosine kinase inhibitor. *Onco Targets Ther* (2016) 9:211–5. doi: 10.2147/OTT.S97100

7. McCoach CE, Le AT, Gowan K, Jones K, Schubert L, Doak A, et al. Resistance mechanisms to targeted therapies in *ROS1*$^+$ and *ALK*$^+$ non–small cell lung cancer. *Clin Cancer Res* (2018) 24(14):3334–47. doi: 10.1158/1078-0432.CCR-17-2452

8. Zhang S, Anjum R, Squillace R, Nadworny S, Zhou T, Keats J, et al. The potent ALK inhibitor brigatinib (AP26113) overcomes mechanisms of resistance to first- and second-generation ALK inhibitors in preclinical models. *Clin Cancer Res* (2016) 22(22):5527–38. doi: 10.1158/1078-0432.CCR-16-0569

9. Uchibori K, Inase N, Araki M, Kamada M, Sato S, Okuno Y, et al. Brigatinib combined with anti-EGFR antibody overcomes osimertinib resistance in EGFR-mutated non-small-cell lung cancer. *Nat Commun* (2017) 8:14768. doi: 10.1038/ncomms14768

10. Wang Y, Yang N, Zhang Y, Li L, Han R, Zhu M, et al. Effective treatment of lung adenocarcinoma harboring EGFR-activating mutation, T790M, and cis-C797S triple mutations by brigatinib and cetuximab combination therapy. *J Thorac Oncol* 15 (8):1369–75. doi: 10.1016/j.jtho.2020.04.014

11. Stinchcombe TE, Doebele RC, Wang X, Gerber DE, Horn L, Camidge DR. Preliminary clinical and molecular analysis results from a single-arm phase 2 trial of brigatinib in patients with disease progression after next-generation ALK tyrosine kinase inhibitors in advanced ALK+ NSCLC. *J Thorac Oncol* (2021) 16(1):156–61. doi: 10.1016/j.jtho.2020.09.018

12. Nishio M, Yoshida T, Kumagai T, Hida T, Toyozawa R, Shimokawaji T, et al. Brigatinib in Japanese patients with ALK-positive NSCLC previously treated with alectinib and other tyrosine kinase inhibitors: outcomes of the phase 2 J-ALTA trial. *J Thorac Oncol* (2021) 16(3):452–63. doi: 10.1016/j.jtho.2020.11.004

13. Ou SI, Nishio M, Ahn MJ, Mok T, Barlesi F, Zhou C, et al. Efficacy of brigatinib in patients with advanced ALK-positive NSCLC who progressed on alectinib or ceritinib: ALK in lung cancer trial of brigAtinib-2 (ALTA-2). *J Thorac Oncol* (2022) 17(12):1404–14. doi: 10.1016/j.jtho.2022.08.018

14. Miyawaki M, Yasuda H, Tani T, Hamamoto J, Arai D, Ishioka K, et al. Overcoming EGFR bypass signal-induced?acquired resistance to ALK tyrosine kinase inhibitors in ALK-translocated lung cancer. *Mol Cancer Res* (2017) 15(1):106–14. doi: 10.1158/1541-7786.MCR-16-0211

15. Hou H, Sun D, Zhang C, Liu D, Zhang X. ALK rearrangements as mechanisms of acquired resistance to osimertinib in EGFR mutant non-small cell lung cancer. *Thorac Cancer* (2021) 12(6):962. doi: 10.1111/1759-7714.13817

16. Shaw AT, Solomon BJ, Besse B, Bauer TM, Lin CC, Soo RA, et al. ALK resistance mutations and efficacy of lorlatinib in advanced anaplastic lymphoma kinase-positive non-Small-Cell lung cancer. *J Clin Oncol* (2019) 37(16):1370–9. doi: 10.1200/JCO.18.02236

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 30 June 2023
DOI 10.3389/fonc.2023.1104137

Check for updates

OPEN ACCESS

EDITED BY
Hongda Liu,
Nanjing Medical University, China

REVIEWED BY
Zhiming Li,
Sun Yat-sen University Cancer Center (SYSUCC), China
Eswar Shankar,
The Ohio State University, United States

*CORRESPONDENCE
Xiao Zhu
bioxzhu@yahoo.com
Chuzhong Wei
2843909948@qq.com

†PRESENT ADDRESS
Xiao Zhu,
Zhejiang Provincial People's Hospital Affiliated to Hangzhou Medical College, Hangzhou Medical College, Hangzhou, China

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

RECEIVED 21 November 2022
ACCEPTED 23 May 2023
PUBLISHED 30 June 2023

CITATION
Zhang W, Wei C, Huang F, Huang W, Xu X and Zhu X (2023) A tumor mutational burden-derived immune computational framework selects sensitive immunotherapy/chemotherapy for lung adenocarcinoma populations with different prognoses.
Front. Oncol. 13:1104137.
doi: 10.3389/fonc.2023.1104137

COPYRIGHT
© 2023 Zhang, Wei, Huang, Huang, Xu 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.

A tumor mutational burden-derived immune computational framework selects sensitive immunotherapy/chemotherapy for lung adenocarcinoma populations with different prognoses

Wenlong Zhang[1‡], Chuzhong Wei[1*‡], Fengyu Huang[1], Wencheng Huang[1], Xiaoxin Xu[1] and Xiao Zhu[2*†]

[1]Huizhou First Hospital, Guangdong Medical University, Huizhou, China, [2]Computational Oncology Laboratory, The Marine Biomedical Research Institute, Guangdong Medical University, Zhanjiang, China

Background: Lung adenocarcinoma (LUAD) kills millions of people every year. Recently, FDA and researchers proved the significance of high tumor mutational burden (TMB) in treating solid tumors. But no scholar has constructed a TMB-derived computing framework to select sensitive immunotherapy/chemotherapy for the LUAD population with different prognoses.

Methods: The datasets were collected from TCGA, GTEx, and GEO. We constructed the TMB-derived immune lncRNA prognostic index (TILPI) computing framework based on TMB-related genes identified by weighted gene co-expression network analysis (WGCNA), oncogenes, and immune-related genes. Furthermore, we mapped the immune landscape based on eight algorithms. We explored the immunotherapy sensitivity of different prognostic populations based on immunotherapy response, tumor immune dysfunction and exclusion (TIDE), and tumor inflammation signature (TIS) model. Furthermore, the molecular docking models were constructed for sensitive drugs identified by the pRRophetic package, oncopredict package, and connectivity map (CMap).

Results: The TILPI computing framework was based on the expression of TMB-derived immune lncRNA signature (TILncSig), which consisted of AC091057.1, AC112721.1, AC114763.1, AC129492.1, LINC00592, and TARID. TILPI divided all LUAD patients into two populations with different prognoses. The random grouping verification, survival analysis, 3D PCA, and ROC curve (AUC=0.74) firmly proved the reliability of TILPI. TILPI was associated with clinical characteristics, including smoking and pathological stage. Furthermore, we estimated three types of immune cells threatening the survival of patients based on multiple algorithms. They were macrophage M0, T cell CD4 Th2, and T cell CD4 memory activated. Nevertheless, five immune cells, including B cell,

endothelial cell, eosinophil, mast cell, and T cell CD4 memory resting, prolonged the survival. In addition, the immunotherapy response and TIDE model proved the sensitivity of the low-TILPI population to immunotherapy. We also identified seven intersected drugs for the LUAD population with poor prognosis, which included docetaxel, gemcitabine, paclitaxel, palbociclib, pyrimethamine, thapsigargin, and vinorelbine. Their molecular docking models and best binding energy were also constructed and calculated.

Conclusions: We divided all LUAD patients into two populations with different prognoses. The good prognosis population was sensitive to immunotherapy, while the people with poor prognosis benefitted from 7 drugs.

KEYWORDS

lung adenocarcinoma, tumor mutational burden, prediction of prognosis, immune landscape, immunotherapy, chemotherapy, molecular docking technology

1 Introduction

The incidence of lung cancer is the second highest in the world (1–3). Lung cancer kills millions of people yearly, and its 5-year survival rates vary from 4-17% on the ground of stages and regional differences (4). The most common histological type of lung cancer is lung adenocarcinoma (LUAD) (5, 6). Last decade, many pioneers studied the LUAD gene, and some of these outstanding scientists achieved remarkable results. KRAS, EGFR, and BRAF are most commonly oncogenes with a mutation in LUAD. TP53, STK11, and KEAP have closely related to tumor suppressors (7).

Long non-coding RNA (lncRNA) is the over 200 bp RNA and is disabled to encode proteins. In recent studies, lncRNA was found to be associated with the development of tumors (8). There are infinite lncRNA types. And lncRNA JPX can increase the number of lung cancer cells and accelerate the growth of tumor cells (9). PD-L1 lncRNA splice isoform facilitates LUAD development by directly enhancing c-Myc activity (10). Moreover, novel lncRNA UPLA1 regulates the activity of LUAD. UPLA1 can facilitate migration, invasion, and proliferation of LUAD and is associated with cell cycle arrest (11). Thus, numerous unknown features between lncRNA and LUAD are worthy of research.

Tumor mutational burden (TMB) indicates the number of mutations per million bases. Recently, FDA approved the pembrolizumab (PD-1 antibody) for the treatment of adult and pediatric patients with unresectable or metastatic high TMB (≥10 mutations/megabase) solid tumors (https://www.accessdata.fda.gov/drugsatfda_docs/label/2020/125514s068lbl.pdf). The immune checkpoint inhibitor (ICI)-based immunotherapy has shown a strong vitality, especially ICIs targeting programmed cell death 1 (PD-1) and programmed cell death-ligand 1 (PD-L1). Yang et al. creatively proposed ICI therapy before surgical resection of the tumor, which improved the survival rate of some lung cancer patients (12). Many studies have shown that tumor patients with high TMB values can achieve better immunotherapy effects (13–16). It shows that TMB is the latest and independent signature in evaluating the efficacy of immunotherapy (17, 18). Some scholars also extended the study of TMB to lung cancer. For example, Hellmann et al. found that lung cancer patients with high TMB had a better treatment response to Nivolumab and Ipilimumab combined immunotherapy (19).

Nevertheless, now the development of TMB encountered some problems. The first was the accuracy of TMB measurement. Secondly, how to apply TMB to the prediction model was also a problem (20). The literature review found that no scholar has constructed a computing framework based on TMB to relate to patients' prognosis and immunotherapy sensitivity. Hence, we decided to contribute in this direction (Figure 1).

2 Materials and methods

2.1 The datasets source

In this study, we collected 528 LUAD and 494 lung squamous cell carcinomas (LUSC) samples from The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov/). The non-small cell lung cancer (NSCLC) samples consisted of aforementioned LUAD and LUSC samples. We only remained with the project of the vial with A and deleted the samples with Vial B or C. Because the vial B or C represents that corresponding samples were fixed by formalin and embedded in paraffin, the effects on RNA-sequence had been proved. We also averaged the RNA-sequence results of multiple samples from the same patient. At last, we obtained 513 LUAD samples. Furthermore, we obtained 288 normal lung samples from the Genotype-Tissue Expression (GTEx) (https://gtexportal.org/). At last, we tried to verify the reliability of the computing framework in external Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/). We collected 9 LUAD datasets with OS (1644 samples), which included GSE11969, GSE13213, GSE26939, GSE31210, GSE36471, GSE63459, GSE68465, GSE68571, and GSE72094. After sorting out these datasets, we

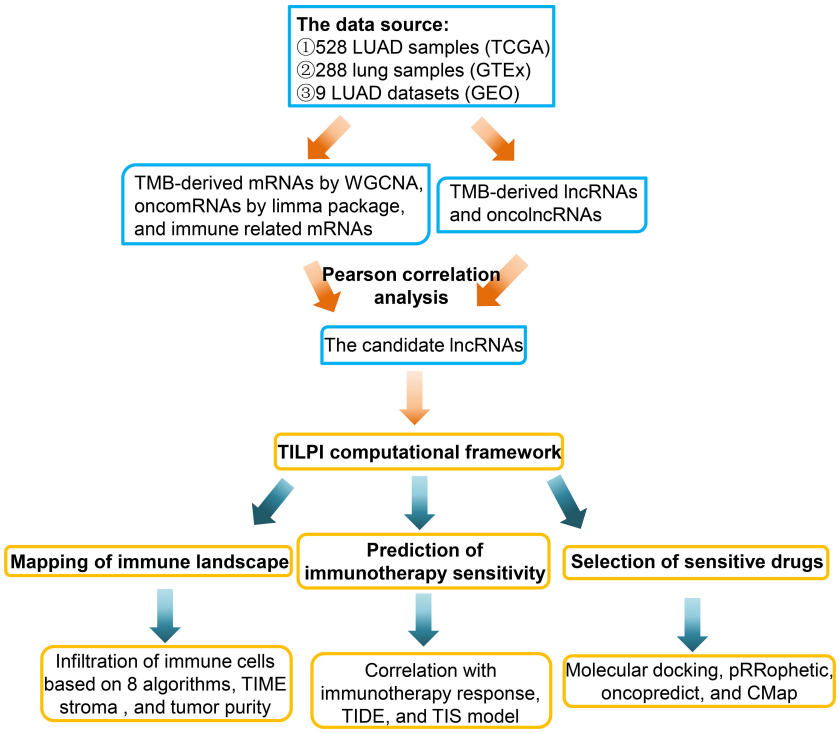


FIGURE 1
This briefly showed the experimental process of this study.

obtained the mRNAs and lncRNAs expression matrix, overall survival (OS) time, survival status, age, gender, smoking, race, pathological stage, and pathological TNM (Table S1). The format of RNA-sequence data we collected was transcripts per kilobase of exon model per million mapped reads (TPM), which facilitated the validation by external datasets. The brief experimental flow was shown in the figure (Figure 1).

2.2 Identify the TMB-derived immune lncRNA set

At the same time, we calculated the TMB value of 513 LUAD samples by the TCGAmutations package of R (21). Nevertheless, we only obtained the TMB value of 509 samples. Then we listed the TMB value in a line. We picked out the 25% minimum as the low TMB group (n = 127) and the 25% maximum as the high TMB group (n = 127). These two groups were used to identify the TMB-derived genes by weighted gene co-expression network analysis (WGCNA) (22). WGCNA holds the idea that the disorder of functional networks leads the tumors and the identification of function-related genes based on the biological network would be more logical. Therefore, it researched the gene functional network analysis in multiple samples rather than simply expressed correlation. We first explored co-expression networks of genes between different TMB groups based on the WGCNA. We constructed the Topological Overlap Matrix (TOM) to decrease the noisy and false relation. Then TOM divides all genes into various module eigengenes (MEs) that consisted of similar functional genes. Then we selected the best soft powers β to build a scale-free network based on the function pickSoftThreshold. In addition, each adjacency matrix was built according to the following formula:

$$\alpha_{ij} = |S_{ij}|^{\beta}$$

(α_i: adjacency matrix between gene i and gene j, S_{ij}: similarity matrix done by Pearson correlation of all gene pairs, β: soft power value). And each adjacency matrix was transformed into a TOM and corresponding dissimilarity (1-TOM). Furthermore, the hierarchical clustering dendrogram based on 1-TOM was constructed, which clustered genes with similar expressions into a co-expression ME. At last, we chose MEs with high correlation coefficients (cor > 0.4, P< 0.05) to conduct further analysis. The TMB-derived lncRNAs and mRNAs of MEs with high correlation coefficients were the candidate genes to construct the computing framework. In addition, the correlation between module membership (MM) and gene significance (GS) of each ME was explored.

Not only that, we built lncRNA and mRNA differentially expression maps between 288 lung samples and 513 LUAD samples based on the limma package of R (23). There were two qualifications to screen the qualified lncRNAs and mRNAs: (1)|log_2FC|≥1, FC refers to the fold change (the expression ratio of lncRNAs or mRNAs between normal samples and LUAD samples). Since the limma package just receives the expression matrix that is log bottomed by 2, so |log_2FC|=1 refers to tumors expressing twice

or half of lncRNAs or mRNAs than normal tissues. (2) False discovery rate (FDR)-adjusted P value< 0.05. This was aimed to obtain oncogenic lncRNAs and mRNAs.

In addition, we collected 2524 immune-related mRNAs from the immunology database and analysis portal (ImmPort) (https://www.immport.org/) and systems biology of the innate immune response (InnateDB) (http://www.innatedb.com/). Furthermore, we intersected TMB-derived mRNAs, oncogenic mRNAs, and immune-related mRNAs to determine candidate mRNAs. We also intersected TMB-derived lncRNAs and oncogenic lncRNAs to determine candidate lncRNAs. Then the Pearson analysis was conducted between candidate lncRNAs and candidate mRNAs. The high correlation coefficient and statistical meaning were qualifications of Pearson analysis (cor > 0.4, P< 0.05). In the end, we successfully identified the TMB-derived immune lncRNA set (TILncSet) and TMB-derived immune mRNA set (TImSet). The TMB-derived immune gene set (TIgeneSet) consisted of TILncSet and TImSet.

2.3 Explore the biological functional pathways of TIgeneSet

Curious about the biological functional pathways of TIgeneSet, we used metascape to explore the potential functions of TIgeneSet (https://metascape.org/). Firstly, we used the Molecular Complex Detection (MCODE) algorithm to construct a protein-protein interaction (PPI) network that showed the functions of TIgeneSet. Secondly, the enrichment analysis in cell type signature also identified cell types close related to TIgeneSet. Thirdly, the enrichment analysis in transcriptional regulatory relationships unraveled by sentence-based text mining (TRRUST) found potential transcription factors related to TIgeneSet. At last, the enrichment analysis in Transcription Factor Targets showed the connected targets of TIgeneSet.

Furthermore, we constructed a node network using the clusterProfiler package of R (24). We mainly used the Gene Ontology (GO) (http://geneontology.org/) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (https://www.kegg.jp/) to conduct the functional enrichment analysis. The filter was P value of analysis less than 0.05.

2.4 Construction and validation of a computing framework for the prediction of prognosis

First, we randomly divided 513 LUAD samples into the training group (n = 257) and the testing group (n = 256). And the TCGA group (n = 513) consisted of the training group and the testing group. The aforementioned TILncSet may lead to the occurrence of LUAD. So TILncSet was more likely to take part in the LUAD progression than other lncRNAs. Therefore, we conducted a statistical analysis based on TILncSet using R-version 4.1.1. And the univariate Cox proportional risk regression analysis, multivariate Cox proportional risk regression analysis, and Kaplan-Meier (KM) method were conducted to identify significant lncRNAs to predict prognosis in the training group. The univariate Cox regression analysis and multivariate Cox regression analysis estimated the TMB-derived immune lncRNA signature (TILncSig). And the formulas of the Cox proportional risk regression analysis were as follows.

$$h\ (t,\ X) = h_0(t) * exp\ (\beta_1 *\ X_1 +\ \beta_2 *\ X_2 + \ldots +\ \beta_n *\ X_n)$$

On the left of the formula, h (t, X) represents the risk function of the individual where X is the predictor or covariate and t is time. The right $h_0(t)$ is the baseline hazard rate of h (t, X) when the X is 0, and it is the quantity to be estimated from the sample data. It's the same for all individuals, so the only difference in risk between individuals is the difference in covariates X. The $exp\ (\beta_1{}^*X_1+\beta_2{}^*X2+\cdots+\beta_n{}^*X_n)$ is called the partial hazard function, and it's different for each individual. $(\beta_1{}^*x_1+\beta_2{}^*x_2+\cdots+\beta_n{}^*x_n)$ is the linear combination of covariate X. $h_0(t)$ is the baseline risk function, which represents the risk when all covariates X are 0. It is the same for all individuals, so the difference in risk among individuals is only the difference in covariates. It's called the partial hazard function, and it's different for each individual. Take the logarithm of both sides of the equation and apply the mathematical transformation:

$$\ln \frac{h\ (t,\ X)}{h_0\ (t)} =\ \beta_1 *\ X_1 +\ \beta_2 *\ X_2 + \ldots +\ \beta_n *\ X_n$$

Then we can figure out the relative risk (RR):

$$RR = \frac{h\ (t,\ X_i)}{h\ (t,\ X_j)} = \frac{h_0\ (t) * \exp\ (\beta^{'} *\ X_i)}{h_0\ (t) * \exp\ (\beta^{'} *\ X_j)}$$

$$= \exp\ [\beta^{'} * (X_i -\ X_j)],\ i,\ j = 1,\ 2,\ \ldots,\ n$$

According to the above theory, we proposed a quantitative computing framework to predict individual prognosis in the training group.

$$\text{Risk Score(TILncSig)} = exp(\ln(h_0(t)) + \sum_{i=1}^{n} \beta(lncRNA_i) * \text{expr}(\text{ lncRNA}_i\ ))$$

Risk Score (TILncSig) is the prognostic index of each LUAD patient. We also called risk score (TILncSig) as TMB-derived immune lncRNA prognostic index (TILPI). The n is the number of lncRNA signatures. $\beta(\text{lncRNA}_i)$ is the coefficient of lncRNA_i obtained by multivariate Cox regression analysis. Expr (lncRNA_i) is the expression level of lncRNA_i.

In addition, we conducted the KM method to estimate the survival probability of individuals. For the nth time point t_n in the study, the survival probability can be calculated as:

$$S(t_n) = S(t_{n-1})\ (1 - \frac{d_n}{r_n})$$

$S\ (t_{n-1})$ is the probability of survival at the time point t_{n-1}. d_n refers to the number of events occurring at the time point t_n. r_n is the alive people number at time point t_n. When $t_0 = 0$, that S (0) = 1.

We used the median TILPI of the training group as a boundary to judge the risk of patients. This boundary was applied to verify the reliability of the computing framework in the testing group and the TCGA group. Furthermore, we confirmed the reliability of the

computing framework. First of all, we used survival analyses and 3D principal component analysis (3D PCA) to verify the difference between the high-TILPI subgroup and the low-TILPI subgroup. Secondly, the performance of TILPI was also evaluated by the time-dependent receiver operating characteristic (ROC) curve. Thirdly, the independent hazard of TILPI and the clinical characteristics were also proved in the study. Furthermore, we also verified the independence of the computing framework by grouping each clinical characteristic. And the Chi-Square test was used to prove clinical characteristics' relationship to the computing framework. In addition, we conducted gene set enrichment analysis (GSEA) to identify the pathways enriching in different risk subgroups (25). The included KEGG pathways of GSEA were 186. And P value< 0.05 was the filter. At last, we constructed a novel model called nomogram to predict individual survival probability, which consisted of TILPI and 8 types of clinical characteristics (age, gender, smoking, race, pathological stage, and pathological TNM) (26).

2.5 Mapping of immune landscape based on computing framework

The tumor immune microenvironment (TIME) is the basis of immunotherapy. Based on 8 quantification algorithms, we described the immune cells' infiltration landscape in detail. The 8 quantification algorithms were Cell type Identification by Estimating Relative Subsets of RNA Transcripts (CIBERSORT) (22 types of immune cells) (27), CIBERSORT-absolute mode (CIBERSORT-ABS) (22 types) (27, 28), Estimating the Proportions of Immune and Cancer cells (EPIC) (8 types) (29), Microenvironment Cell Populations-counter (MCPCOUNTER) (10 types) (30), Quantifying Immune Contexture of Human Tumors (QUANTISEQ) (11 types) (31), Tumor Immune Estimation Resource (TIMER) (6 types) (28, 32), Tumor and Immune System Interaction Database (TISIDB) (28 types) (33), and digitally portraying the tissue cellular heterogeneity landscape (XCELL) (36 types) (34). In addition, we picked out types of immune cells more distributed in different TILPI subgroups based on intersection analyses. There were 2 conditions for intersection analyses. Firstly, the standard-compliant immune cells must be proven more distributed in a risk subgroup with at least 2 algorithms. Secondly, this result can't contradict another algorithm.

TIME has not only all kinds of immune cells but also numerous stromal components. Therefore, we collected the TIME score and stroma score from XCELL. And the cytotoxicity score of MCPCOUNTER was also calculated. Furthermore, we also got the stromal score and tumor purity based on an algorithm called Estimation of STromal and Immune cells in MAlignant Tumour tissues using Expression data (ESTIMATE) (35). The other TIME components were also collected from the tumor immune dysfunction and exclusion (TIDE). They included interferon gamma (IFNG), T-cell-inflamed signature (Merck18 score), CD8, CD274 (PD-L1), cancer-associated fibroblast (CAF), myeloid-derived suppressor cell (MDSC), and tumor-associated macrophage M2 (TAM M2) (36). Furthermore, we planned to obtain the immune subtype of samples based on 6 types of immune subtypes, which consisted of wound healing (immune C1), IFN-γ dominant (immune C2), inflammatory (immune C3), lymphocyte depleted (immune C4), immunologically quiet (immune C5), and TGF-β dominant (immune C6) (37). The difference and correlation analyses above were based on the Wilcoxon test and Pearson correlation coefficient.

2.6 Prediction of immunotherapy sensitivity based on computing framework

Immune checkpoint inhibitor (ICI) was a significant immunotherapy, which decreases the expressions of immune checkpoint proteins. It was reported that the TIDE score was excellent to predict the response to immunotherapy. Therefore, we planned to combine the TIDE score with TILPI to predict immunotherapy sensitivity. We first collected the average expression levels of cytotoxic T lymphocyte (CTL) signatures (CD8A, CD8B, GZM, GZMB, PRF1) to predict the distribution of CTL. According to the average expression levels of CTL, we divided all samples into the hot-tumor subgroups with above-average CTL levels and the cold-tumor subgroups with below-average CTL levels. Every hot tumor has a T cell dysfunction score while every cold tumor has a T cell exclusion score. In hot tumor subgroups, the T cell dysfunction score was derived by systematically identifying genes that were related to CTL infiltration levels to affect patients' OS. T cell dysfunction score of each gene was calculated as follows:

$$\text{Dysfunction} = \frac{d}{StdErr(d)}$$

Then we compared the dysfunction scores of each gene to identify key genes that affected CTL and death hazards. For each hot-tumor sample, the final T-cell dysfunction score was modeled from Cox-PH regression:

$$\text{Hazard} = a{*}\text{CTL} + b{*}\text{P} + d{*}\text{CTL}{*}\text{P}$$

In this model, CTL represents the CTL level. The P represents the expression level of the candidate gene. The coefficient d reflects the influence of interaction between CTL and candidate gene P on death hazard. In the cold-tumor subgroup, the T cell exclusion score is derived by the expression levels of 3 types of cells that restrict T cell infiltration in tumors. They are CAF, MDSC, and TAM. T cell exclusion score of the cold-tumor subgroup was acquired from TIDE (http://tide.dfci.harvard.edu/). In the end, the TIDE score is the combination of the T cell dysfunction score from the hot-tumor subgroup and the T cell exclusion score from the cold-tumor subgroup. Based on the above computation, we analyzed the correlation between the TILPI computing framework and various scores of TIDE.

Moreover, we found a novel immune prediction model called tumor inflammation signature (TIS) (38). It has proved that the TIS model retrospectively predicted the clinical benefit of anti-PD-1 treatment in clinical trials. TIS model also quantifies an adaptive immune response in TIME. TIS model is composed of 18 genes

(CD3D, IDO1, CIITA, CD3E, CCL5, GZMK, CD2, HLA-DRA, CXCL13, IL2RG, NKG7, HLA-E, CXCR6, LAG3, TAGAP, CXCL10, STAT1, GZMB). We also connected TILPI with the TIS model to predict the immunotherapy response. At last, we conducted time-dependent ROC curves for the TIDE model, TIS model, and TILPI computing frameworks in 1, 3, and 5 years' OS.

The 5 published transcriptomics signatures of the immune response were used to validate the possibility that the low TILPI group was suitable for immunotherapy. Tertiary lymphoid structures (TLS) signature is based on differentially expressed genes in tumor tissue with TLS (39). Jerby-Arnon immune resistance are the resistance program combining a gene set related to T cell exclusion, post treatment, and functional resistance (40). Roh immune score is defined by the genes set involved in immune activation associated with tumor rejection (41). Ock anti-CTLA-4 signature is derived from the expression of 105 genes associated with the response to immunotherapy (42). EaSIeR model is based on multi-task machine learning to predict different hallmarks of immune responses (43). All these transcriptomics signatures were calculated following the methodology and code in the original studies. The format of RNA-sequence data we used was TPM.

2.7 Prediction of sensitive drugs and tumor evolutionary status based on computing framework

We wished TILPI computing frameworks perform in predicting the sensitive drugs of individualized chemotherapy. This research was based on the R package called pRRophetic (44). The version of pRRophetic was published in 2016 including 251 types of drugs. The second algorithm to identify sensitive drugs was oncopredict (45). We used semi-inhibitory concentration (IC50) as the boundary to pick out sensitive drugs for different risk subgroups. The drugs with lower IC50 were sensitive for this subgroup. The sensitive drugs for a risk subgroup must meet 2 filters: the P value of the Wilcoxon test< 0.05 and the P value of Spearman correlation analysis< 0.05. Furthermore, we used the connectivity map (CMap) to identify sensitive drugs inhibiting up-regulated TMB-derived oncogenic genes (https://clue.io/). The research was conducted in 28 cell lines, different doses (0.001 uM-90uM), and different processing times (1h-72h). We only selected known compounds and targets. And the absolute normalized CMap score of qualified drugs must be greater than 1.5. At last, we intersected 3 derived drugs to identify candidate drugs.

In addition, we analyze the modes of interaction between the candidate drugs and their targets based on Autodock Vina 1.2.2 (46). The molecular structures of candidate drugs were retrieved from PubChem (https://pubchem.ncbi.nlm.nih.gov/). And the 3D coordinates of their targets were downloaded from the PDB (http://www.rcsb.org/). Then we constructed the molecular docking models by Autodock Vina 1.2.2 (http://autodock.scripps.edu/).

Based on network pharmacology, we next searched for potential targets of candidate drugs targeting LUAD. LUAD targets are from the GeneCards database (www.genecards.org/). The SMILE numbers of the candidate drugs were acquired from the Pubchem database (https://pubchem.ncbi.nlm.nih.gov/) and then sequentially imported into the SwissTargetPrediction database (http://www.swisstargetprediction.ch/) for target prediction. Targets with a probability >0 were selected as potential targets for candidate drugs. The BATMAN-TCM database was also used to obtain the target information for candidate drugs (http://bionet.ncpsb.org.cn/batman-tcm/). The filter was score cutoff >10 and adjusted P_value >0.05. We also used NCI-60 cell line set in CellMiner database to search for genes associated with drug candidates (P<0.05) (47).

The stemness score is known as a significant score for the prediction of tumors' progression (48). The epigenetically regulated-mRNA expression-based stemness score (EREG-mRNAss) was used to assess the tumor evolutionary status because EREG-mRNAss was related to known tumor biological functions, therapy sensitivity, clinical characteristics, and tumor pathology. Furthermore, we obtained another similar stemness score called RNA expression-based stemness score (RNAss). We planned to connect TILPI computing frameworks with stemness scores to estimate the evolutionary status of the tumor. Therefore, we conducted the correlation analysis between TILPI and various stemness scores.

2.8 Statistical analysis

We used R version 4.1.1 to analyze data and create figures and tables (https://www.r-project.org). We also drew diagrams with the help of an online website called bioinformatics (https://www.bioinformatics.com.cn/). The KM method was conducted to verify clinical characteristics' independence of computing frameworks. The log-rank test was used to calculate the P value of survival difference between two subgroups. Furthermore, we used the Chi-Square test to prove clinical characteristics' relationship to the computing framework. The other correlation analyses were based on the Wilcoxon test and Pearson correlation coefficient.

3 Results

3.1 Identify the TMB-derived immune lncRNA set

We collected 509 patients with TMB values and divided them into four equal parts according to the TMB score. Then we took the first 127 and the last 127 patients as the low TMB subgroup and high TMB subgroup. Furthermore, we conducted WGCNA to find TMB-derived mRNA modules, so the mRNA expressed differently between the high and low TMB subgroups was identified. We used Topological Overlap Matrix (TOM) to construct a new neighborhood matrix to reduce error and false correlation. Consequently, we sorted out the mRNA matrix and determined the optimal power value ($\beta = 4$). This value considered both scale independence and mean connectivity (Figure 2A). Moreover, we divided all mRNAs into 19 module eigengenes (MEs) based on the functional correlation (Figures 2B, C). In the end, we obtained 2 MEs

10.3389/fonc.2023.1104137

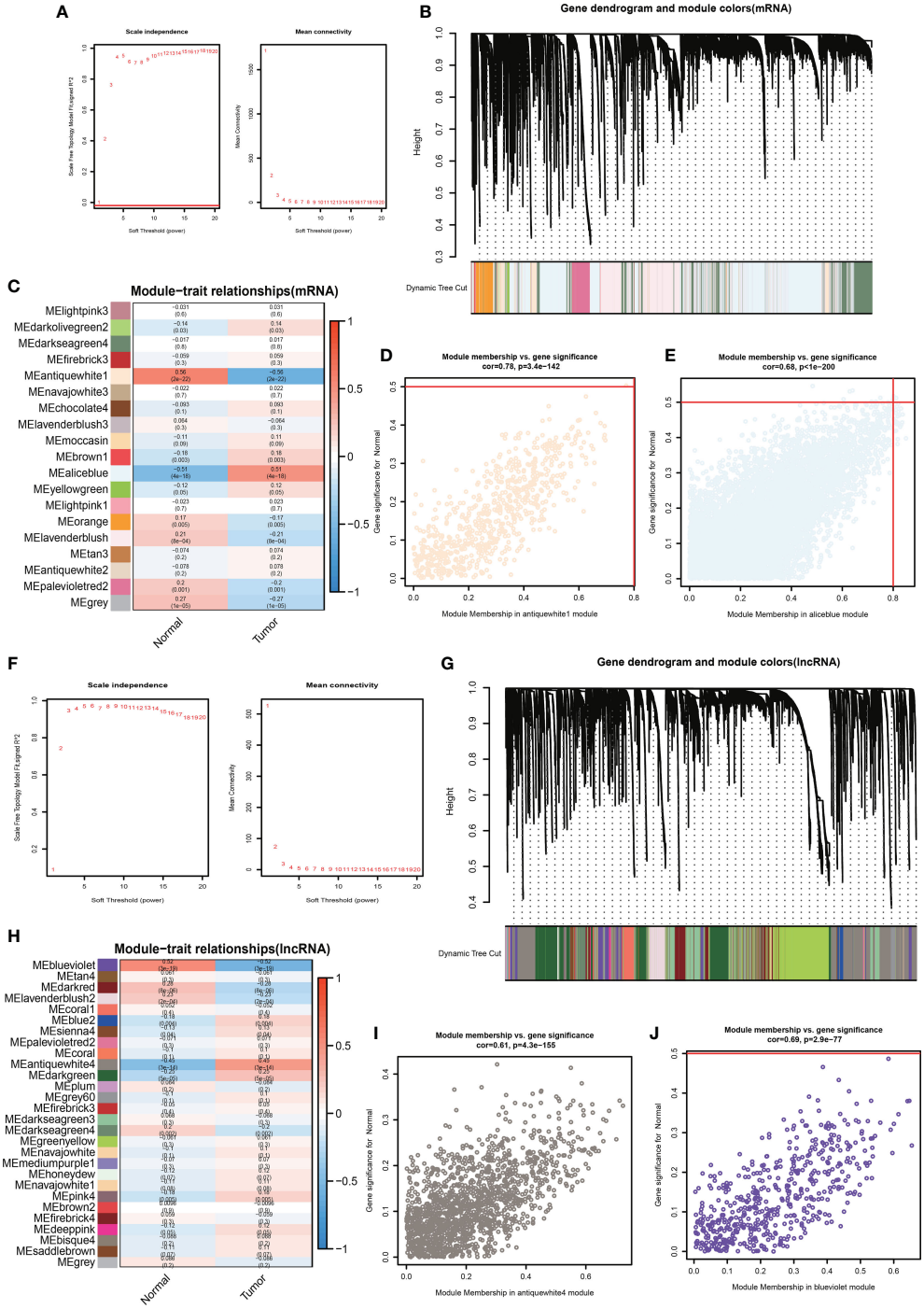


FIGURE 2
Determine the optimal power value (β = 4) taking into account both "scale independence" and "mean connectivity" **(A)**. There were 19 MEs according to lncRNAs' functional relevance **(B, C)**. Verify the correlation between GS and MM. **(F)** The best power value (β) was 3 **(D, E)**. There were 28 MEs **(G, H)**. Verify the correlation between GS and MM **(I-J)**.

expressing differently between the high and low TMB subgroups. They were MEantiquewhite1 (cor = -0.56, P = 2e-22) and MEaliceblue (cor = 0.51, P = 4e-18) (Figure 2C). However, the other 17 MEs were lower associated with the high TMB group (cor<0.4) (Figure 2C). Subsequently, we aimed to confirm the correlation between gene significance (GS) and module membership (MM). We got excellent results for MEantiquewhite1 (cor = 0.68, P< 1e-200) and MEaliceblue (cor = 0.78, P = 3e-142) (Figures 2D, E). But in the rest of the MEs, GS had poor correlations with MM. In conclusion, we obtained 8527 TMB-derived mRNAs from MEantiquewhite1 and MEaliceblue. Moreover, we used WGCNA to find TMB-derived lncRNA modules. And TOM was

10.3389/fonc.2023.1104137

used to construct a new neighborhood matrix to reduce errors and false correlations. Lastly, we made the lncRNA matrix and determined the optimal power value (β = 3). This value considered both scale independence and mean connectivity (Figure 2F). Afterward, according to the functional relevance, we divided all lncRNAs into 28 MEs (Figures 2G, H). Finally, we found that two lncRNA MEs were correlated with TMB. They were MEblueviolet (cor=-0.52, P=3e-19) and MEantiquewhite4 (cor=0.45, P=3e-14) (Figure 2H). And the remaining 26 MEs were irrelevant (cor<0.4). Then, we intended to confirm the correlation between GS and MM. MEblueviolet (cor=0.69, p=2.9e−77) and MEantiquewhite4 (cor=0.61, p=4.3e−155) also did good jobs (Figures 2I, J). However, in the rest 26 MEs, GS was not highly correlated with MM. Finally, 2053 TMB-derived lncRNAs were identified based on MEblueviolet and MEantiquewhite4.

Furthermore, we used the limma package of R to identify differentially expressed mRNAs between 288 normal lung samples from GTEx and 513 LUAD samples from TCGA. Two prerequisites were required before confirming the differential expression of mRNAs: (1) $|log_2FC| \geq 1$ (2) FDR adjusted P value< 0.05. Then, we got a total of 14437 differentially expressed mRNAs. But only 7925 mRNAs met the above two prerequisites (Figures 3A, B). In addition, also for those samples, we found 1187 lncRNAs differentially expressed by limma package. Then we sorted out 871 oncogenic lncRNAs that met the abovementioned premises (Figures 3C, D).

Furthermore, we collected 2524 immune-related mRNAs from the immunology database and analysis portal (ImmPort) and systems biology of the innate immune response (InnateDB). Based on the above, we obtained 8527 TMB-derived mRNAs by WGCNA, 7925 oncogenic mRNAs selected by the limma package, and 2524 immune-related mRNAs. At last, we intersected these three mRNA sets and obtained 267 immune TMB-derived oncogenic mRNAs (Figure 3E). As for lncRNA, we identified 871 oncogenic lncRNAs and 2053 TMB-derived lncRNAs. Then we analyzed the intersection and got 176 eligible TMB-derived oncogenic lncRNAs (Figure 3F). Then we conducted the Pearson correlation analysis to investigate their correlation (Figure 3G). Finally, we obtained 36 immune TMB-derived oncogenic lncRNAs (cor>0.4), which was called TMB-derived immune lncRNA set (TILncSet). And the 43 immune TMB-derived oncogenic mRNAs (cor>0.4) were called TMB-derived immune mRNA set (TImSet). The TMB-derived immune gene set (TIgeneSet) consisted of TILncSet and TImSet.

3.2 Explore the functional biological pathways of TIgeneSet

As we all know, although lncRNA can't encode proteins, it has immeasurable effects on cellular life activities. We intended to examine

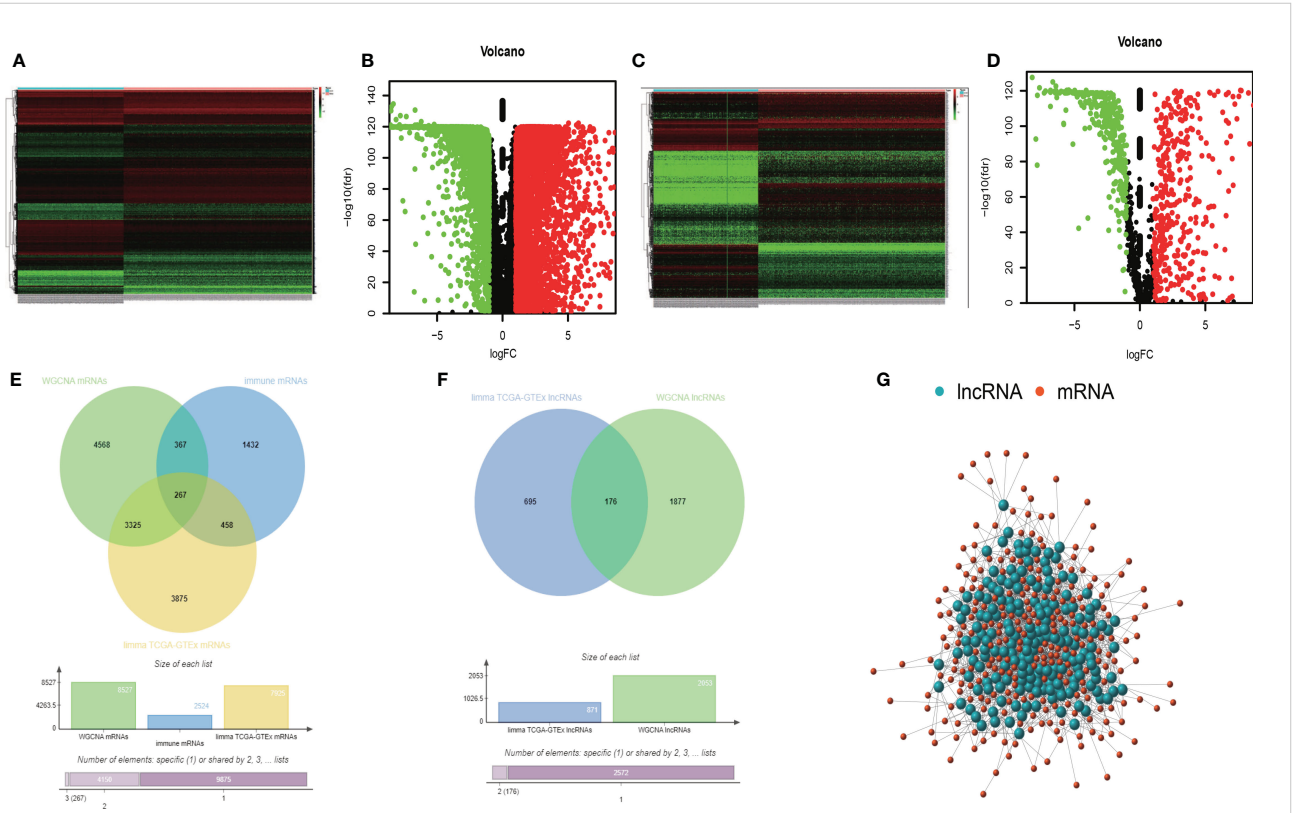


FIGURE 3
The heatmap of differently expressed mRNAs **(A)**. The volcano map of differently expressed mRNAs **(B)**. The heatmap of differently expressed lncRNAs **(C)**. The volcano map of differently expressed lncRNAs **(D)**. The intersection analysis of TMB-derived mRNAs, oncogenic mRNAs, and immune- related mRNAs **(E)**. The intersection analysis of TMB-derived lncRNAs and oncogenic lncRNAs **(F)**. The Pearson correlation network between lncRNAs and mRNAs **(G)**.

10.3389/fonc.2023.1104137

what the 43 mRNAs and 36 lncRNAs from these experiments would play in LUAD. Therefore, we utilized a meta scape to explore the potential functions of these 79 genes. The meaningful enriched pathways must meet the following prerequisites: P value<0.01, a minimum count of 3, and an enrichment factor >1.5. Furthermore, we made a network of enriched terms based on 20 pathway clusters with the smallest P value. In the network, each gene of pathway clusters was represented by a node and was colored by pathway cluster ID (Figure 4A). The size of nodes referred to gene counts of pathway clusters, and the color was correlated with the P value (Figure 4B). The function of 20 pathway clusters was brilliant, which included positive regulation of protein phosphorylation, regulation of MAPK cascade,

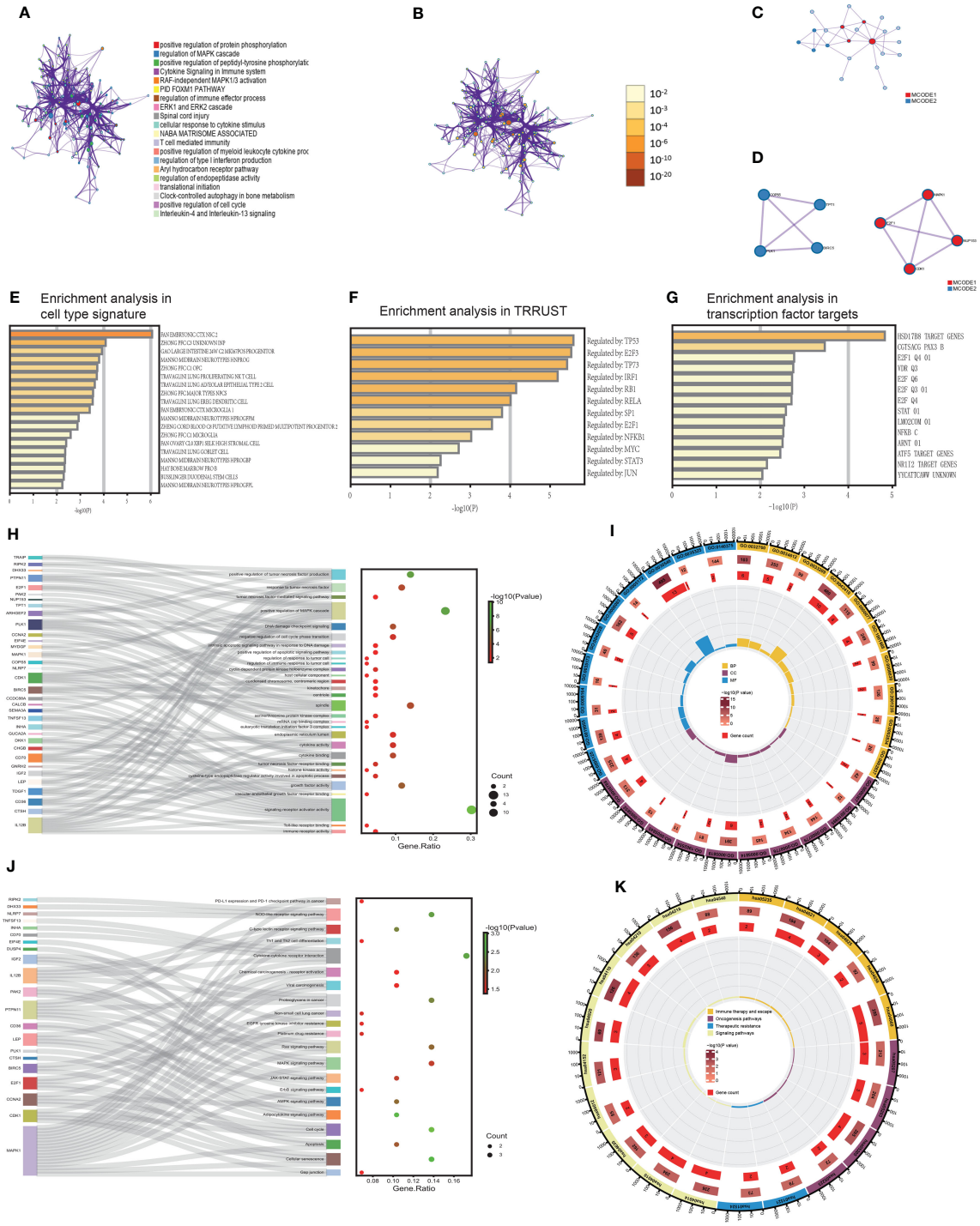


FIGURE 4
The functional biological pathways were clustered **(A, B)**. The PPI network and MCODE algorithm of TIgeneSet **(C, D)**. The enrichment analysis in cell type signature **(E)**. The enrichment analysis of TRRUST. **(G)** The enrichment analysis in transcription factor targets **(F)**. The GO functional enrichment analysis of TIgeneSet **(H, I)**. The KEGG functional enrichment analysis of TIgeneSet **(J, K)**.

cytokine signaling in the immune system, regulation of immune effector process, ERK1 and ERK2 cascade, T cell-mediated immunity, translational initiation, positive regulation of cell cycle.

Wielding the above genes, we also constructed a protein-protein interaction (PPI) network based on three pathways with the smallest P value (Figure 4C). Their functional descriptions were positive protein phosphorylation, regulation of MAPK cascade, and RAF-independent MAPK1/3 activation. If a subset contains between 3 and 500 proteins, we will conduct the Molecular Complex Detection (MCODE) algorithm to show its density in the network. The PPI network showed the relationship between 22 protein subsets of 3 pathways. MCODE1 represented the four red high-density protein subsets, and the other four blue high-density protein subsets were represented by MCODE2 (Figure 4C). MCODE1 included E2F1, MAPK1, CDK1, and NUP153. These four proteins possessed significant interactions between the two (Figure 4D). And MCODE2 included COPS5, TPT1, BIRC5, and PLK1. Except for the fact that there was no interaction between TPT1 and BIRC5, the rest of the proteins had functional interactions between them (Figure 4D). Backing to the macroscopic level, MCODE1 and MOCDE2 also interacted through 4 proteins (Figure 4D).

The enrichment analysis in cell type signature showed the relationship between 76 genes and some cell types, such as lung proliferating NK T cells and lung goblet cells. Lung proliferating NK T cells were associated with the innate form of the immune barrier. And the abnormal proliferation of lung goblet cells refers to LUAD (Figure 4E). The relationship between these genes and LUAD was inseparable. In addition, the enrichment analysis in transcriptional regulatory relationships unraveled by sentence-based text mining (TRRUST) indicated the relationship between 76 genes and transcription factors. These transcription factors or target genes corresponding to transcription factors were TP53, E2F3, TP73, IRF1, RB1, RELA, SP1, E2F1, NFKB1, MYC, STAT3, and JUN (Figure 4F). The above were sorted by P value. TP53 is closely related to the functional activity of LUAD cells. TP73 is highly homologous to TP53, its function involves all aspects of cellular life activities, and its transcriptionally translated protein p73 is a carcinostatic factor. IRF1 negatively regulated the expression of the oncogene kpna2 in LUAD cells under conditions of growth stimulation and hypoxia. SP1 is associated with LUAD transfer. E2F1 and KLF6 form a positive feedback pathway in LUAD, regulating the cell cycle and leading to cisplatin resistance in LUAD. MYC drives the evolution of small-cell lung cancer subtypes. Activating the STAT3 signaling pathway can promote the development of LUAD. Therefore, the relationship between these transcription factors or target genes and LUAD is indivisible.

Furthermore, the enrichment analysis in transcription factor targets showed 14 connected targets (Figure 4G). The E2F family is involved in developing LUAD and affects prognosis and efficacy. E2F1 is correlated with the cell cycle and LUAD resistance. Abnormalities in the STAT pathway are closely related to cell hyperplasia, differentiation, and LUAD development.

We manipulated the GO and KEGG functional enrichment analysis to explore further what roles 36 lncRNAs and 43 mRNAs play in LUAD. P value<0.05 was the filter in GO and KEGG analysis. Then we found 961 meaningful pathways in the GO analysis. There were 827 pathways for biological process (BP), 51 for cell component (CC), and 83 for molecular function (MF). We respectively selected ten excellent pathways in BP, CC, and MF to draw intuitive diagrams. Pathways of BP were associated with the regulation of tumor cell response, apoptosis pathway, cell cycle, tumor-related signaling pathway, and regulation of tumor necrosis factor. Pathways of CC were associated with mitotic processes, protein translation processes, and protein kinase complexes. Pathways of MF were associated with immune-related receptors, cytokines, protein kinases, their related functions, and regulation of growth factors. Then we obtained a diagram that showed the correlation between functional pathways and genes. In the figure, each functional pathway was wired to the corresponding genes. In the adjacent dot plot, larger dots indicated more genes associated with the corresponding pathway, and the dots were colored according to the P value. The gene ratio was the ratio of the number of genes associated with the pathway to the total number of genes obtained (Figure 4H). Obviously, in BP, the positive regulation of the MAPK cascade was associated with the most genes. This pathway was associated with ten genes, which were TDGF1, LEP, IGF2, DHX33, CD36, DKK1, SEMA3A, PTPN11, MYDGF, and RIPK2. And other pathways, like a response to tumor necrosis factor and positive regulation of tumor necrosis factor production, were associated with 5 and 6 genes (Figure 4H). In CC, the spindle was connected with the most genes. They are PLK1, BIRC5, MAPK1, CDK1, ARHGEF2, and TPT1 (Figure 4H). In MF, signaling receptor activator activity was correlated with the most genes. These 13 genes were TDGF1, LEP, IGF2, GNRH2, IL12B, CD70, CHGB, DKK1, GUCA2A, INHA, TNFSF13, SEMA3A, and CALCB. In conclusion, a sector chart was constructed to show the results visually (Figure 4I).

We got 46 functional pathways in the KEGG analysis and selected 21 significant ones. These functional pathways were divided into immune therapy and escape, oncogenesis, therapeutic resistance, and signaling pathways. And there were relationships between functional pathways and genes (Figure 4J). Each pathway was wired to the associated genes (Figure 4J). In immune therapy and escape, cytokine-cytokine receptor interaction was associated with five genes, including LEP, IL12B, CD70, INHA, and TNFSF13. And NOD-like receptor signaling pathway possesses four corresponding genes. There are NLRP7, DHX33, MAPK1, and RIPK2 (Figure 4J). The number of genes associated with this pathway was similar to oncogenesis pathways. The four genes related to proteoglycans in cancer were IGF2, MAPK1, IL12B, and PTPN11. We also found four genes associated with the cell cycle, which included PLK1, CCNA2, CDK1, and E2F1. CCNA2, MAPK1, CDK1, and E2F1 were enriched in cellular senescence (Figure 4J). In therapeutic resistance, we proved that these genes took part in both EGFR tyrosine kinase inhibitor resistance and platinum drug resistance (Figure 4J). In signaling pathways, the RAS signaling pathway and MAPK signaling pathway were also enriched pathways (Figure 4J). Moreover, we drew a sector chart to show the results visually (Figure 4K).

3.3 Construction and validation of the TILPI computing framework

3.3.1 Construct the TILPI computing framework in the training group

To enhance the practicability of TILncSet, we hoped it could determine the prognosis of patients with LUAD. According to the above results, we knew that TILncSet possibly induces LUAD, which is more likely to participate in the pathogenic pathway of LUAD than other lncRNAs. Therefore, we sought prognostic lncRNA signatures of LUAD based on TILncSet. We divided all meaningful TCGA samples (n=513) (the TCGA group) into two groups, namely the training group (n=257) and the testing group (n=256) (Table S1). We counted some clinical characteristics of patients in each group, and it was easy to see that these clinical characteristics were particularly evenly distributed in each group, so the reliability of the grouping was preliminarily verified (Table S1). To screen for prognostic-related lncRNAs, univariate Cox proportional hazard regression analysis was used to analyze the relationship between expression levels of 36 TMB-derived immune lncRNAs and OS in the training group, and 7 TMB-derived immune lncRNAs were found to be significantly associated with the prognosis of LUAD patients (AC091057.1, AC129492.1, AC112721.1, TARID, AC114763.1, LINC00592, AC025166.1) (Figure 5A). In addition, we hoped to screen out lncRNAs with independent prognostic value from these 7 candidate lncRNAs, and conducted multivariate Cox proportional hazards regression analysis for these 7 candidate lncRNAs. Finally, 6 of 7 candidate lncRNAs (AC091057.1, AC129492.1, AC112721.1, TARID, AC114763.1, LINC00592) were identified as independent prognostic lncRNAs (Figure 5B). The six all lncRNAs of TILncSig were risk factors for LUAD patients because their coefficients based on multi-Cox analysis were all positive. Then a TMB-derived immune lncRNA signature (TILncSig) was constructed. Next, we built a computing framework to evaluate the risk score generated by TILncSig's expressions in individuals. The computing framework was as follows:

$$\mathrm{TILPI} = \exp(\ln(h(t_0)) + \sum_{i=1}^{n} \mathrm{coef}(\mathrm{lncRNA}_i) * \mathrm{expr}(\mathrm{lncRNA}_i))$$

$$\ln(h(t_0)) = -0.7850$$

TILPI is a prognostic risk score for the LUAD patients. For each individual, his TILPI = exp (-0.7850 + 0.0908 * expression (AC091057.1) + 0.3175 * expression (AC129492.1) + 0.1009 * expression (AC112721.1) + 0.1987 * expression (TARID) + 0.2754 * expression (AC114763.1) + 0.1442 * expression (LINC00592)). The median score of the LUAD patients in the training group (median =0.8510) was used as a risk cutoff to classify patients into the low-risk group with low TILPI (TILPI ≤0.8510) or high-risk group with high TILPI (TILPI >0.8510).

To further verify the effectiveness of grouping, we utilized 3D principal component analysis (3D PCA) to verify the reliability in the training group, and the results proved that our grouping was reliable (Figure 5C). Kaplan–Meier analysis showed that the survival time of patients in the low-risk group are significantly better than patients in the high-risk group (P<0.001) (Figure 5D). In the training group, the 5-year survival rate of the high-risk subgroup was 7.8%, worse than 11.6% in the low-risk subgroup. Based on the above data, we demonstrated that TILPI did a good job of relating to the OS of patients. To prove the credibility of TLIPI, we used the time-dependent ROC curves to observe the training group. Finally, the area under the curve (AUC) value of the TILPI was 0.730 in the training group. It meant that TILPI had excellent credibility in judging OS in the training group, but other clinical characteristics didn't (age (ACU=0.542), gender (ACU=0.625), race (AUC=0.511), smoking (AUC=0.506), pathological stage (AUC=0.648), pathological T (AUC=0.635), pathological N (AUC=0.597), pathological M (AUC=0.478)) (Figure 5E).

We also used univariate Cox proportional hazards regression analysis to verify the independence of clinical characteristics (age, gender, smoking, pathological stage, and pathological TNM) and TILPI in the training groups. The results showed that age, pathological stage, pathological T, pathological N, and TILPI were independent risk factors (P<0.05), while gender, smoking, and pathological M wasn't (P>0.05) (Figure 5F) (Table S2). Next, we used multivariate Cox proportional hazards regression analysis on age, pathological stage, pathological T, pathological N, and TILPI. In the end, only pathological N and TILPI were eligible independent risk factors in the training group (P<0.05) (Figure 5G) (Table S2). We considered that the expressions of TILncSig were factors that affected individual TILPI. Firstly, all of TILncSig were expressed less in the low-risk subgroup than in the high-risk subgroup (P<0.001) (Figure 5H). Therefore, AC091057.1, AC129492.1, AC112721.1, TARID, AC114763.1, and LINC00592 were likely disadvantageous to LUAD patients in the training group.

3.3.2 Verify the reliability of the TILPI computing framework in the TCGA group and the testing group

In the previous experiment in the training group, the potential of the computing framework to relate to OS was demonstrated. However, it was still necessary to further verify its reliability in the testing group. When the same TILncSig and risk cutoff as those derived from the training group was applied to the testing group, 256 patients of the testing group was classified into the low-risk group (n = 121) and high-risk group (n = 135) with significantly different overall survival. As Figure 5J showed that the overall survival of 135 patients in the high-risk group was much poorer than 121 patients in the low-risk group (P=0.014). Of course, we also used 3D PCA to verify the reliability of grouping in the testing group. Meanwhile, the result proved that our grouping was reliable (Figure 5I). The time-dependent ROC curves showed that TILPI (AUC=0.750), pathological stage (AUC=0.750), and pathological N (AUC=0.657) had credibility in judging OS in the testing group, but other clinical characteristics didn't [age (ACU=0.532), gender (ACU=0.546), race (AUC=0.516), smoking (AUC=0.542), pathological T (AUC=0.638), pathological M (AUC=0.530)] (Figure 5K). Furthermore, we verified the independence of clinical characteristics and TILPI in the testing groups. And the result showed that only the pathological stage and TILPI were statistically meaningful (P<0.05) (Figures 5L, M) (Table S2). Moreover, the expressions of TILncSig influencing TILPI in the testing group were similar to the training group. AC091057.1,

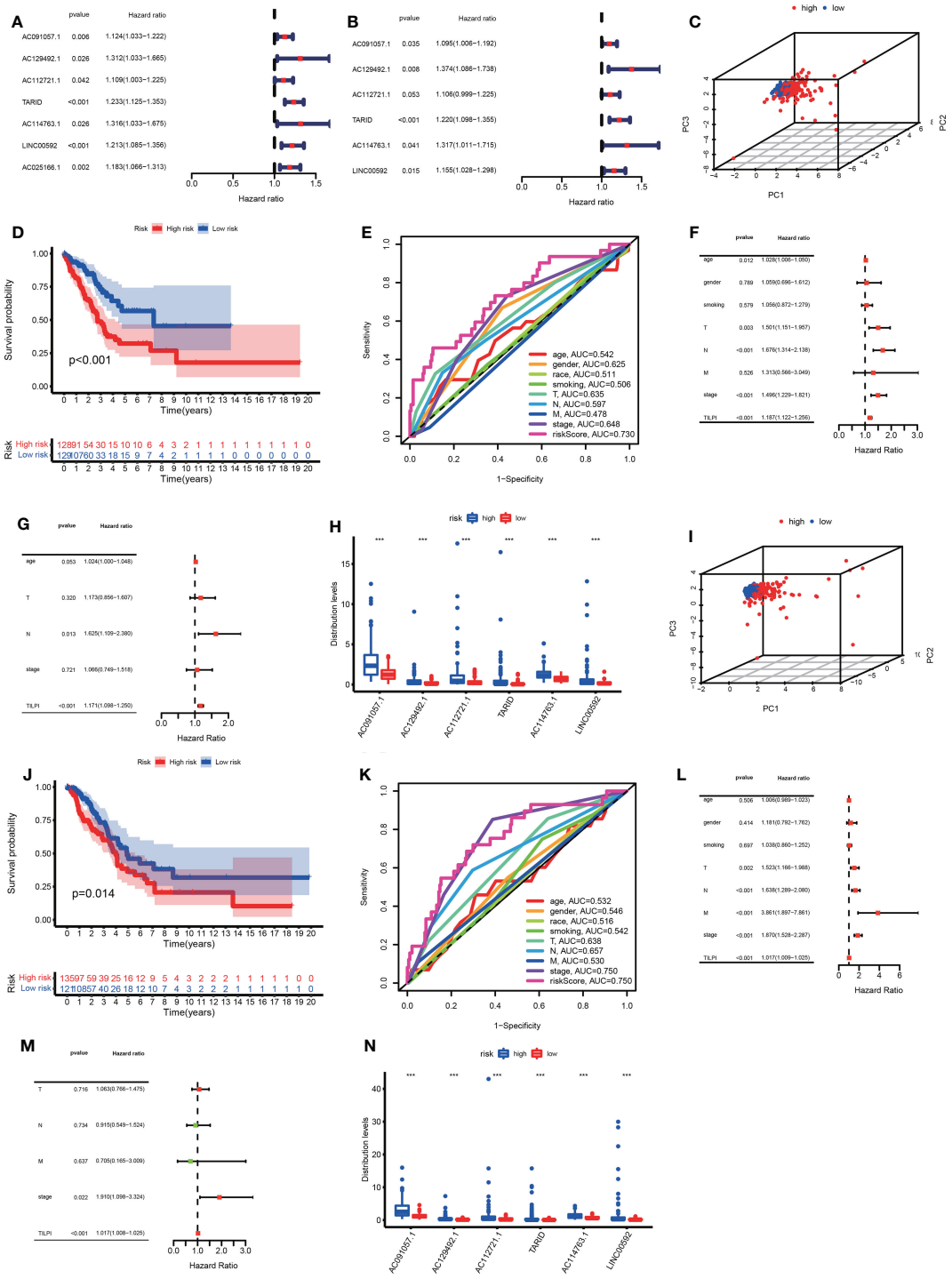


FIGURE 5
We successfully identified TILncSig based on univariate and multivariate Cox proportional hazards regression analyses in the training group **(A, B)**. In the training group, the 3D PCA, survival analysis, and ROC curve verify the reliability of the TILPI computational framework **(C-E)**. The univariate and multivariate Cox analyses prove the independent prognostic hazard of TILPI **(F, G)**. The expression of TILncSig was different in different risk subgroups **(H)**. Verify the reliability of the TILPI computational framework in the testing group **(I-N)**. ***P < 0.001.

AC129492.1, AC112721.1, TARID, AC114763.1, and LINC00592 were all life-threatening in the testing group, too (P<0.001) (Figure 5N).

The prognostic performance of the TILncSig in the TCGA group was similar to the above results. We similarly used the median TILPI of the training group (0.8510) as a cutoff to divide the TCGA group (n=513) into the low-risk subgroup (n=250) and high-risk subgroup (n=263). As we expected, 3D PCA also showed that grouping based on TILPI was reliable (Figure 6A). The overall survival of 263 patients in

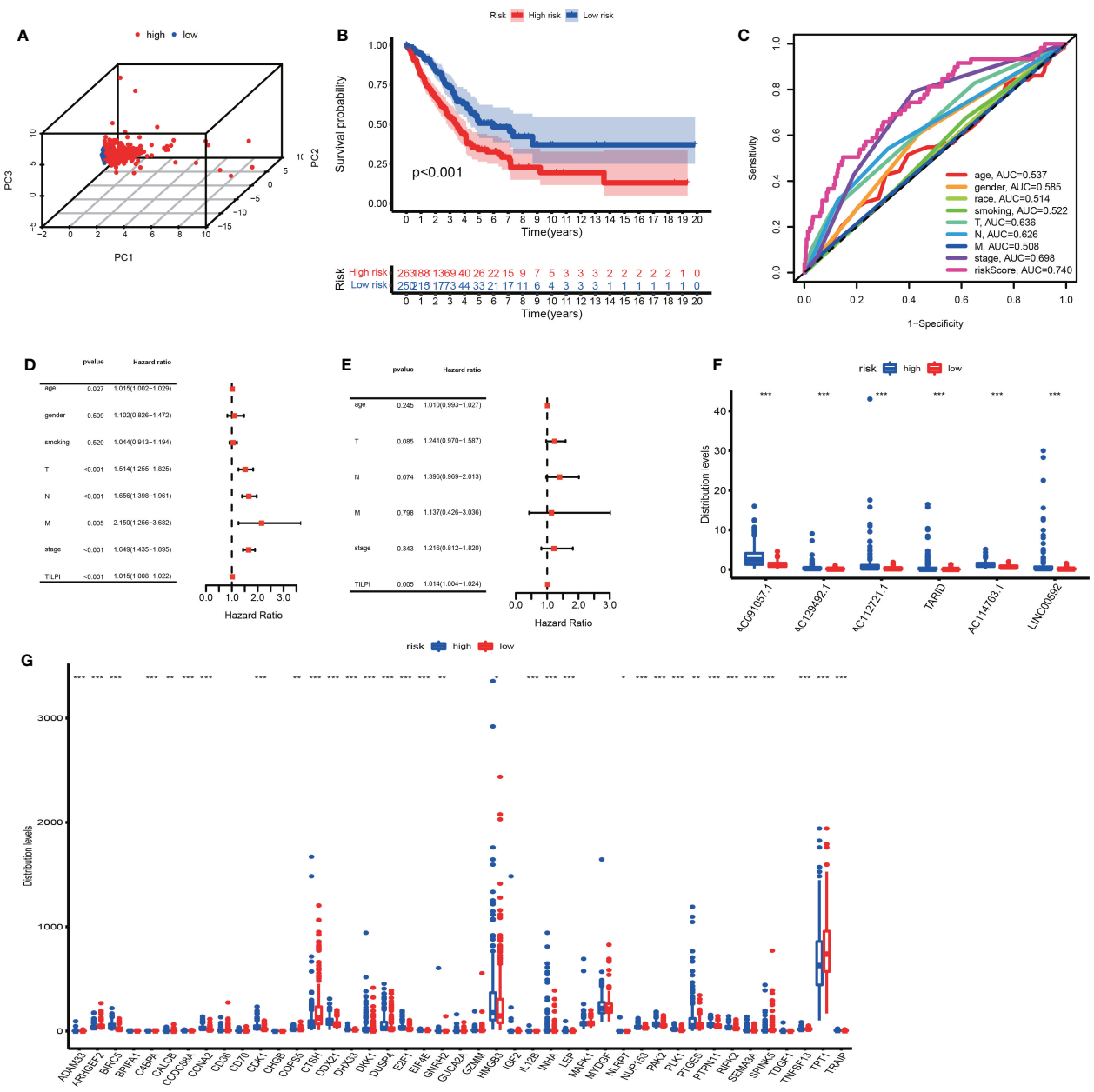


FIGURE 6
Verify once again the reliability of the TILPI computational framework in the TCGA group **(A-G)**. *P < 0.05; **P < 0.01; ***P < 0.001.

the high-risk group was much poorer than 250 patients in the low-risk group (P<0.001) (Figure 6B). And Similar to the training group, the survival rate was 9.9% in the high-risk group at 5 years lower than 13.2% in the low-risk group (Figure 6B). The results of the time-dependent ROC curves in the TCGA group showed that TILPI (AUC=0.740) and pathological stage (AUC=0.698) had credibility in judging OS in the TCGA group while other clinical characteristics didn't (age (ACU=0.537), gender (ACU=0.585), race (AUC=0.514), smoking (AUC=0.522), pathological T (AUC=0.636), pathological N (AUC=0.626), pathological M (AUC=0.508)) (Figure 6C). Furthermore, according to the independence analyses based on univariate and multivariate Cox proportional hazards regression of clinical characteristics and TILPI in the TCGA group, the only meaningful result was TILPI (P=0.005) (Figures 6D, E) (Table S2). In addition, we also demonstrated that AC091057.1, AC129492.1, AC112721.1, TARID, AC114763.1, and LINC00592 were all life-threatening in the TCGA group too (P<0.001) (Figure 6F).

Unlike the other two groups, we investigated whether the expressions of 43 above mRNAs affected the individual TILPI in the TCGA group. We found that 30 of the 43 mRNAs were associated with TILPI. The 21 mRNAs were expressed more in the high-risk subgroup (BIRC5, CCDC88A. CCNA2, CDK1, CALCB, COPS5, DDX21,

DHX33, DKK1, DUSP4, E2F1, EIF4E, HMGB3, INHA, LEP, NLRP7, NUP153, PAK2, PLK1, PTGES, PTPN11, RIPK2, SEMA3A, TRAIP). And nine mRNAs were expressed more in the low-risk subgroup (ADAM33, ARHGEF2, C4BPA, CTSH, GNRH2, IL12B, SPINK5, TNFSF13, TPT1) (P<0.05) (Figure 6G). These results suggested that the expressions of 9 mRNAs would improve the prognosis of LUAD patients, while 21 other mRNAs were adverse.

Subsequent survival analyses examined the likelihood of a single gene predicting survival. AC114763.1, LINC00592, TARID, and AC091057.1 independently predicted patient survival (P<0.05) (Figures S1A-D). AC129492.1 and AC112721.1, on the other hand, did not perform well (Figures S1E, F). TILPI's potential to predict survival was also extended to 1077 patients with NSCLC. Taking the median TILPI (0.8510) of the above training group as cutoff, survival probability of patients in the high TILPI group was significantly lower than those in the low TILPI group (P=0.01) (Figure S1G).

At last, we tried to verify the reliability of TILPI in GEO datasets and collected 9 LUAD datasets with OS (1644 samples). They were GSE11969, GSE13213, GSE26939, GSE31210, GSE36471, GSE63459, GSE68465, GSE68571, and GSE72094. However, the platforms of 9 all external datasets didn't cover TILncSig.

3.3.3 Clinical characteristics independence analysis of TILPI

Curious about whether the prognostic value of the TILPI was independent of common clinical characteristics, multivariate Cox regression analyses were performed on age, gender, the degree of smoking, race, pathological TNM and pathologic stage. (Table S3). Firstly, we divided samples based on age into the old subgroup (age > 65, n = 262) and the young subgroup (age ≤ 65, n = 260). As the figure showed that the low-risk subgroup and the high-risk subgroup exhibited obvious survival differences in the old subgroup (P< 0.001) (Figure 7A), while these differences weren't statistical meaning in the young subgroup (P=0.057) (Figure 7B). For the gender, there were survival differences between the high-risk subgroup and low-risk subgroup in the male samples (n=242, P=0.004) (Figure 7C), while the female samples (n=280) were as well (P<0.001) (Figure 7D). But we found that TILPI wasn't independent of smoking (P>0.05) (Figures 7E-H) or race (Figures 7I-K). Regarding the pathological T, the computing framework was unable to relate to the survival probability of patients in the T1 subgroup (n=172, P=0.386), but it was able to relate to the survival probability of patients in the T2 subgroup (n=281, P=0.007), T3 subgroup (n=47, P=0.001), and T4 subgroup (n=19, P=0.031) (Figures 7L-O). In addition, the computing framework wasn't independent in the N1 subgroup (n=99, P=0.309) (Figure 7P), but it was independent in the N0 subgroup (n=335, P=0.003) and N2 subgroup (n=75, P=0.026) (Figures 7Q, R). What's more, the computing framework was able to divide patients into high-risk subgroups and low-risk subgroups in the M0 subgroup (n=335) (P<0.001) (Figure 7S), while wasn't in the M1 subgroup (n=26, P=0.634) (Figure 7T). Moreover, the computing framework was valid in the stage II subgroup (P=0.046) and stage III subgroup (P=0.007) (Figures 7U, V), but it was invalid in the stage I subgroup (P=0.106) or stage IV subgroup (P=0.886) (Figures 7W, X). In conclusion, there wasn't independence between TILPI and seven clinical characteristics (age, smoking, race, pathological stage, pathological T, pathological N, and pathological M).

3.3.4 Clinical characteristics correlation analysis of TILPI

We were also curious about whether the prognostic value of the TILPI was associated with common clinical characteristics, Chi-square tests were performed on age, gender, the degree of smoking, race, pathological TNM and pathologic stage (Figure 8A) (Table S3). Firstly, it was obvious that the age of patients wasn't associated with TILPI (P = 0.402) (Figure 8B) or race (P=0.208) (Figure 8C). There were more female patients than male patients, but higher TILPI was more likely to occur in male patients than in female patients (P = 0.001) (Figure 8D). Furthermore, we also found that TILPI was related to smoking (P = 0.001) (Figure 8E). What's more, pathological T and pathological N were statistically meaningful (P = 0.001, P = 0.003) (Figures 8F, G), but pathological M was not (P=0.344) (Figure 8H). The pathological stage was closely related to TILPI. In conclusion, five clinical characteristics were associated with TILPI, which included gender, smoking, pathological stage, pathological T, and pathological N.

According to independence and correlation analysis, we knew that TILPI was related to various clinical characteristics (smoking, pathological stage, pathological T, and pathological N) (Table S3). Thus, the reliability of TILPI in predicting prognosis may depend on these clinical features.

3.3.5 GSEA pathway correlation analysis of TILPI

Through GSEA, we found 36 significant enriched pathways in different risk subgroups (P<0.05). Thirty pathways were enriched in the high-risk group. And ten pathways possessed excellent biological functions concerning LUAD (pathways in cancer, cell cycle, p53 signaling pathway, mismatch repair, DNA replication, starch and sucrose metabolism, glycolysis gluconeogenesis, galactose metabolism, pentose and glucuronate interconversions, maturity-onset diabetes of the young) (Figure 9A). In the high-risk group, five pathways were related to cell division and DNA mutations, while the last five pathways were related to energy metabolism, meaning we could kill LUAD cells by affecting their division and energy metabolism. Furthermore, six significant pathways enriched in the low-risk group (allograft rejection, asthma, hematopoietic cell lineage, intestinal immune network for IgA production, renin-angi system, viral myocarditis) (Figure 9B). In the low-risk group, three pathways were all immune response-activated pathways, suggesting that immune activation may be responsible for protecting the low-risk group. Therefore, individuals in the high-risk group may transform into the low-risk group by activating these immune pathways, thus prolonging OS.

3.3.6 Construction of the prognosis nomogram based on TILPI and clinical features

To use TILPI more accurately to stratify the risk of LUAD patients in the clinic, we used TILPI and several clinical

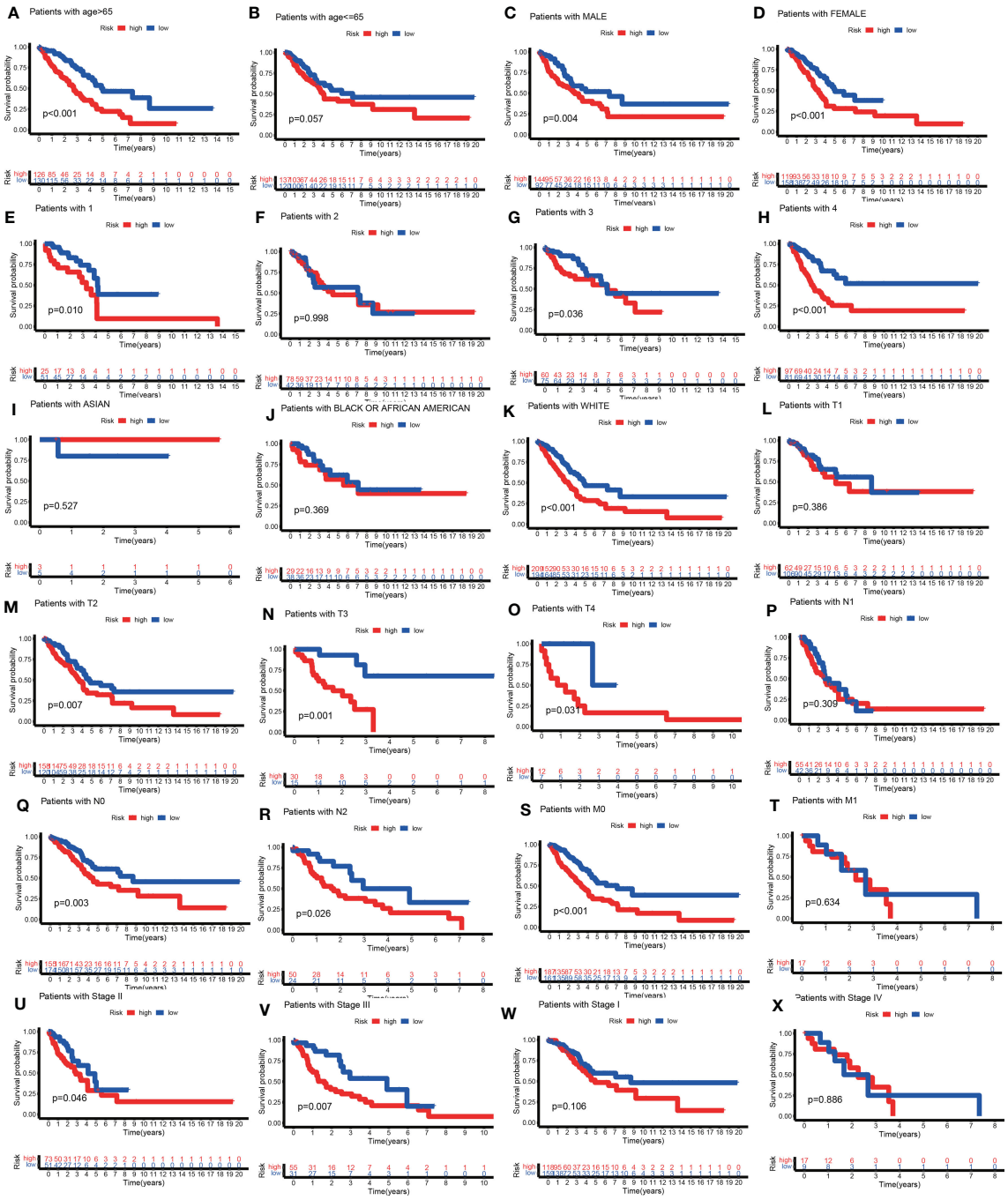


FIGURE 7
Explore the clinical independence of the TILPI computational framework by grouping. The low-risk and the high-risk subgroups exhibited obvious survival differences in the old subgroup **(A)**, while these differences weren't statistical meaning in the young subgroup **(B)**. For the gender, there were survival differences between the high-risk and low-risk subgroups in the male samples **(C)**, while the female samples were as well **(D)**. But TILPI wasn't independent of smoking **(E-H)** and race **(I-K)**. TILPI was unable to relate to the survival probability of patients in the T1 subgroup, but it was able to relate to the survival probability of patients in the T2, T3, and T4 subgroups **(L-O)**. In addition, TILPI was not independent in the N1 subgroup **(P)**, but it was independent in the N0 and N2 subgroups **(Q, R)**. What is more, TILPI was able to divide patients into high-risk and low-risk subgroups in the M0 subgroup **(S)**, while was not in the M1 subgroup **(T)**. Moreover, TILPI was valid in the stage II and stage III subgroups **(U, V)**, but it was invalid in the stage I and stage IV subgroup **(W, X)**.

characteristics (age, gender, race, smoking, pathological stage, pathological TNM) to construct a new model that could calculate the OS probability (Figure 9C). To verify the reliability of this model, we fitted the calculated survival time with the actual survival time. The results showed that the 1-year survival rate prediction, 2-year survival rate prediction, and 3-year survival rate prediction all had a good fit with the actual survival time (Figures 9D-F). It also verified the reliability of TILPI. We confirmed that this new prognostic calculation model enabled to help clinicians to calculate the survival time of patients more easily.

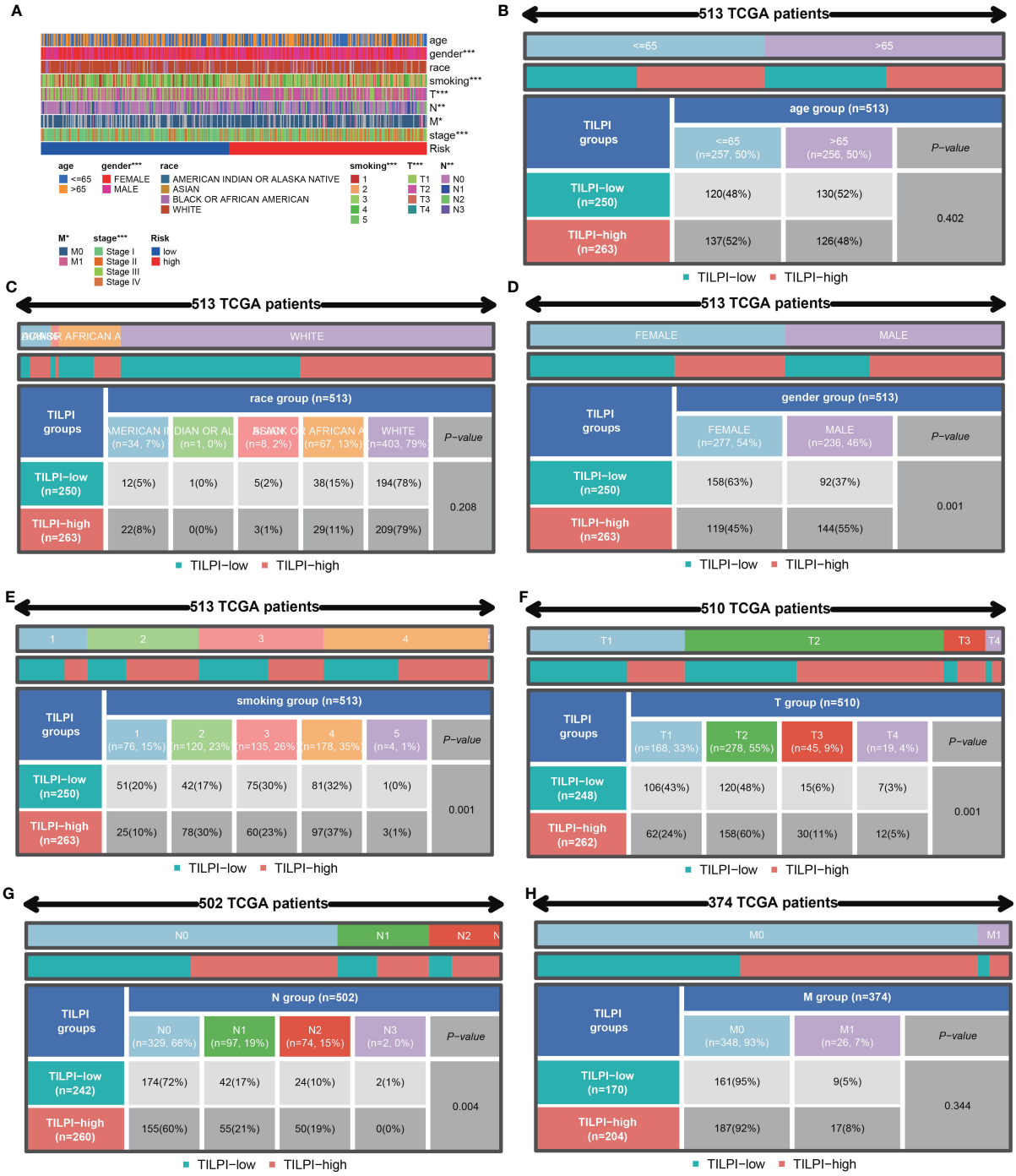


FIGURE 8
The correlation analyses between TILPI computational framework and clinical characteristics. TILPI was correlated with gender, smoking, pathological TNM, and pathological stage **(A)**. TILPI wasn't associated with **(B)**, as was race **(C)**. The higher TILPI was more likely to occur in male patients **(D)**. TILPI was related to smoking **(E)**. What's more, the correlation between TILPI and pathological T was statistically meaningful **(F)**, as was pathological **(G)**. But pathological M was not **(H)**. TILPI, TMB-derived immune lncRNA prognostic index. *P < 0.05; **P < 0.01; ***P < 0.001.

3.4 Mapping of immune landscape based on TILPI

The tumor immune microenvironment (TIME) is the soil of immunotherapy. The ratio of immune cells will show more accurate treatment in individualized immunotherapy if the relationship between immune cells and TIME is close-knit. Therefore, we applied eight algorithms to describe the immune infiltration landscape in detail. These algorithms were cell type identification by CIBERSORT, CIBERSORT-ABS, EPIC, MCPCOUNTER, QUANTISEQ, TIMER, TISIDB, and XCELL. Firstly, we found 22 types of immune cells by CIBERSORT. And we drew four pictures

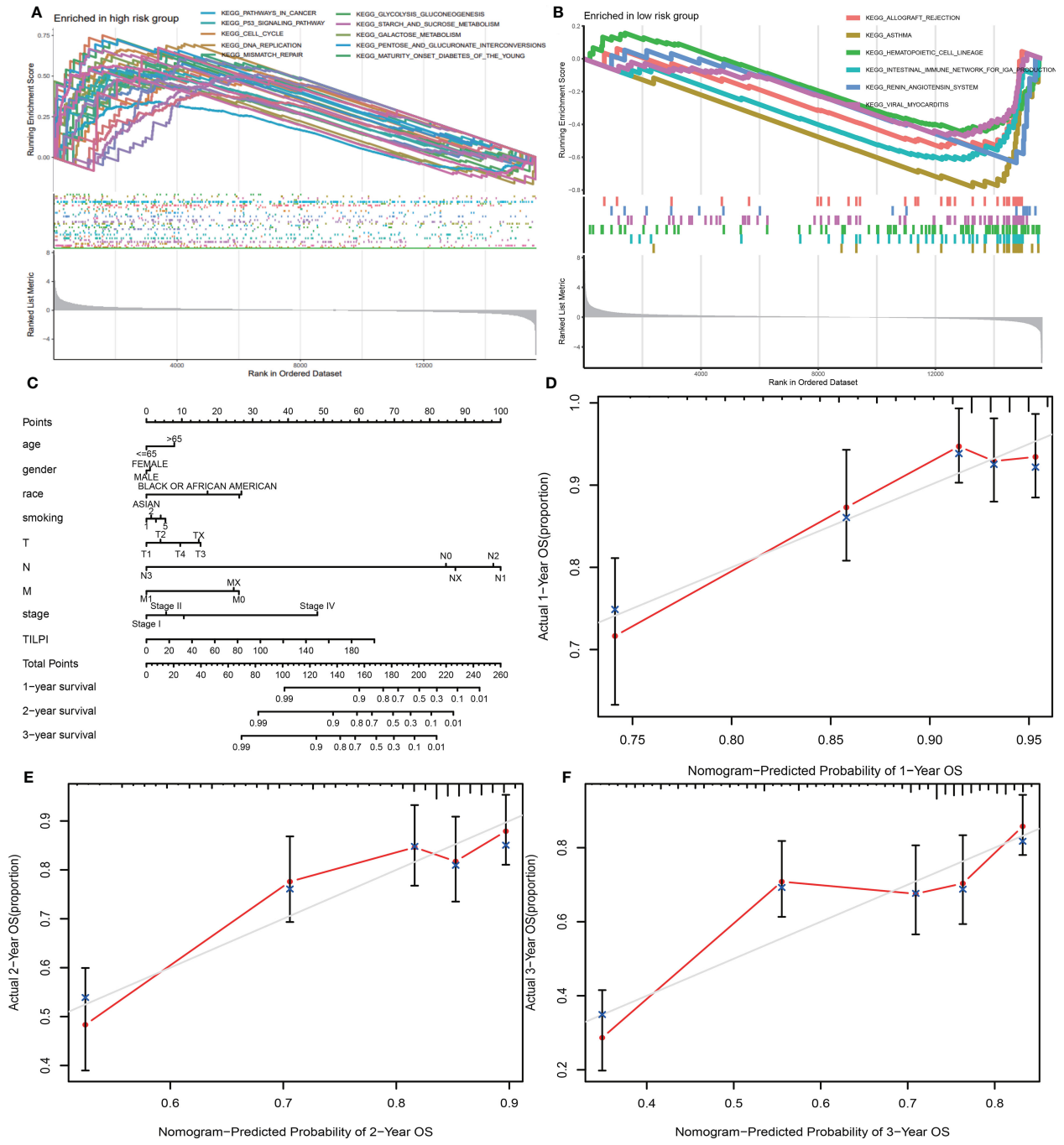


FIGURE 9
30 pathways were enriched in the high-risk group, and ten possessed excellent biological functions concerning LUAD **(A)**. There were six significant pathways enriched in the low-risk subgroup **(B)**. A new model was constructed consisting of TILPI and several clinical characteristics (age, gender, race, smoking, pathological stage, and pathological TNM) **(C)**. The new model's 1, 2, and 3-year survival rate predictions fit well with the actual survival time **(D-F)**.

that described the composition of immune cells. These plots were based on the Wilcoxon test and illuminated the difference between 22 immune cells in high-risk and low-risk subgroups.

Furthermore, we found that the distribution of 11 immune cells in the subgroups was statistically significant (P<0.05) (Figure 10A). There were four types of immune cells more distributed in the high-risk subgroup. They were macrophage M0, macrophage M1, plasma cell, and T cell CD4 memory activated. And the other seven types of immune cells were more distributed in the low-risk subgroup, which included B cell memory, dendritic cells resting, monocyte, mast cell resting, eosinophil, T cell CD4 memory resting, T cell regulatory (Tregs). However, the remaining 11 types of immune cells were not significantly statistical, which included B cell naive, dendritic cell activated, macrophage M2, mast cell activated, neutrophil, NK cell activated, NK cell resting, T cell CD4 naïve, T cell CD8, T cell follicular helper, T cell gamma delta. And the

10.3389/fonc.2023.1104137

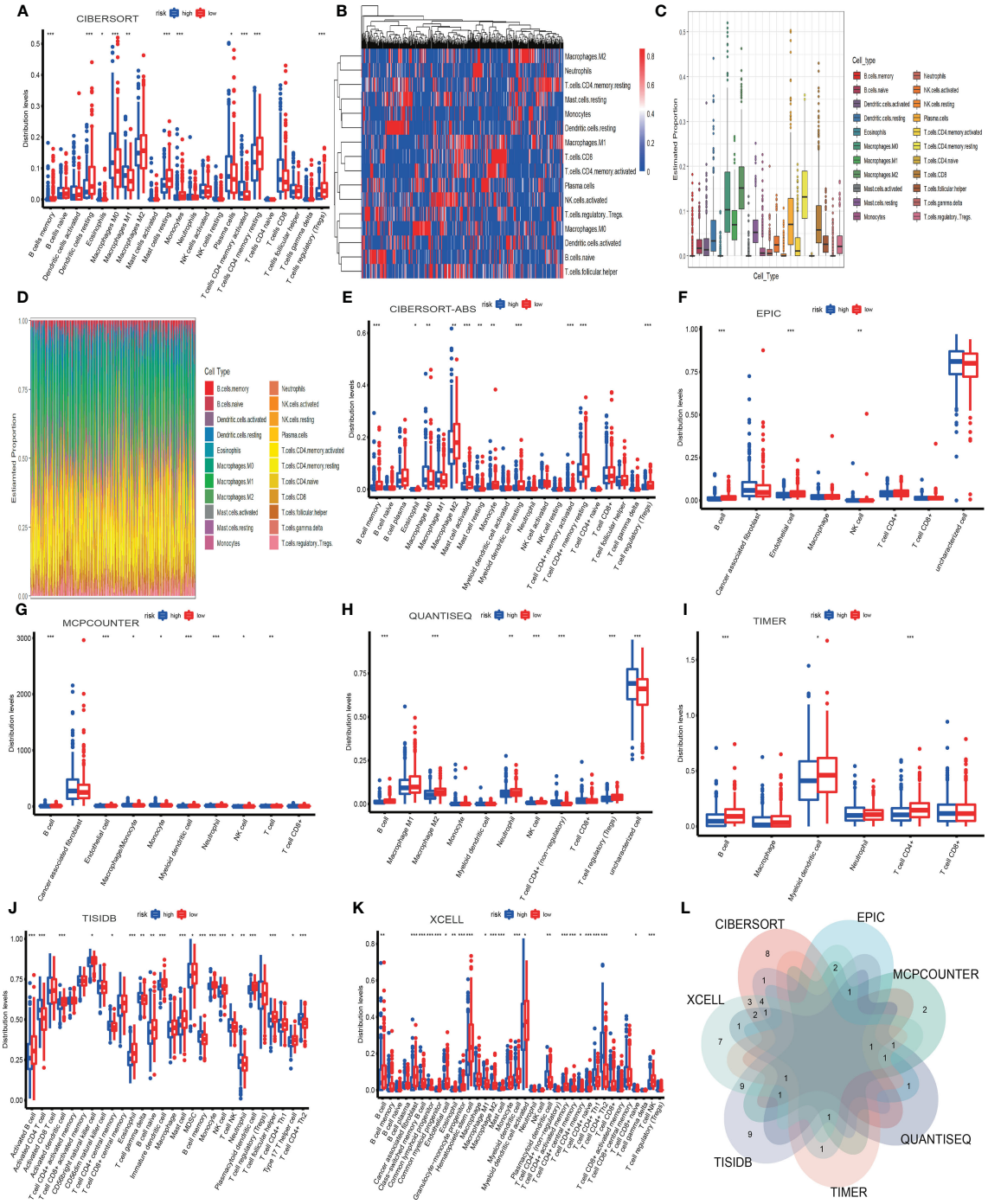


FIGURE 10
The infiltration landscape of 22 immune cells based on CIBERSORT **(A-D)**. The other seven algorithms showed the infiltration of immune cells, which included CIBERSORT-ABS, EPIC, MCPCOUNTER, QUANTISEQ, TIMER, TISIDB, and XCELL **(E-K)**. The intersection of immune cells of different algorithms **(L)**. *P < 0.05; **P < 0.01; ***P < 0.001.

heatmap showed 16 types of immune cell expression. And the more immune cells are expressed, the more the color in the graph tends to be red.

Conversely, the fewer immune cells are expressed, the more the color in the heatmap inclines to be blue (Figure 10B). The plot showed Twenty-two types of immune cells and their expression. Macrophage M2 was the most expressed (Figure 10C). In the landscape diagram, the immune infiltration landscape of each sample was shown. Twenty-two colors represented twenty-two types of cells, the abscissa represented the samples, and the percentage of color length on the ordinate represented the ratio of immune cells (Figure 10D). Secondly, we wielded CIBERSORT-

ABS to obtain a picture based on the Wilcoxon test, which indicated the distribution of 22 types of immune cells in the high-risk and low-risk subgroups (Figure 10E).

Furthermore, we found that the graph's distribution of 11 immune cells was statistically significant (P<0.05). They were three types more distributed in the high-risk subgroup. They were macrophage M0, mast cell resting, and T cell CD4 memory activated. The other eight types were more distributed in the low-risk subgroup, which involved B cell memory, eosinophil, macrophage M2, mast cell activated, monocyte, myeloid dendritic cell resting, T cell CD4 memory resting, T cell regulatory (Tregs). But the remaining ten types were not statistically significant, which included B cell naive, B cell plasma, macrophage M1, myeloid dendritic cell activated, neutrophil, NK cell activated, NK cell resting, T cell CD4 naive, T cell CD8, T cell follicular helper.

Thirdly, we applied EPIC to obtain a graph based on the Wilcoxon test, which implied the distribution of 8 types of immune cells in high-risk and low-risk subgroups. (Figure 10F) Furthermore, we found that graph's distribution of three immune cells was statistically significant (P<0.05). And there was 1 type more distributed in the high-risk subgroup. It was the NK cell. The other two types were more distributed in the low-risk subgroup, B cell and endothelial cell. But the remaining five types were not statistically significant, including CAF, macrophage, T cell CD4, T cell CD8, and uncharacterized cells. Fourthly, we used MCPCOUNTER to draw a picture based on the Wilcoxon test, which showed the distribution of 10 types of immune cells in high-risk and low-risk subgroups (Figure 10G).

Furthermore, we found that CAF and T cell CD8 were not statistically significant, and the distribution of the remaining immune cells in the graph was statistically significant (P<0.05). They were three types more distributed in the high-risk subgroup. They were macrophage/monocyte, monocyte, and NK cell. And the other five types were more distributed in the low-risk subgroup, which included B cell, endothelial cell, myeloid dendritic cell, neutrophil, and T cell. Fifthly, we wielded QUANTISEQ to draw a diagram based on the Wilcoxon test, which showed the distribution of 11 types of immune cells in the high-risk and low-risk subgroups (Figure 10H).

Moreover, we found that the distribution of 7 immune cells in the diagram was statistically significant (P<0.05). And there were two types, including T cell CD4 (non-regulatory) and uncharacterized cells more distributed in the high-risk subgroup. The other five types were more distributed in the low-risk subgroup. They involved B cells, macrophage M2, neutrophils, NK cells, and T cell regulatory (Tregs). Nevertheless, the remaining four types were not statistically significant. They were macrophage M1, monocyte, myeloid dendritic cell, and T cell CD8. Sixthly, we manipulated TIMER to obtain a picture based on the Wilcoxon test, which indicated the distribution of 6 types of immune cells in high-risk and low-risk subgroups (Figure 10I).

Furthermore, we found that the distribution of 3 immune cells in the picture was statistically significant (P<0.05). And all the types were more distributed in the low-risk subgroup. They were B cell, myeloid dendritic cell, and T cell CD4. But the rest of the cells were not statistically significant, which included macrophage, neutrophil, and T cell CD8.

Seventhly, we operated TISIDB to draw a diagram based on the Wilcoxon test, which showed the distribution of 28 types of immune cells in the high-risk and low-risk subgroups (Figure 10J). In addition, we found that the distribution of 20 immune cells in the graph was statistically significant (P<0.05). And there were seven types more distributed in the high-risk subgroup. They were T cell CD4 central memory, activated CD4 T cell, T cell gamma delta, B cell memory, T cell NK, neutrophil, and T cell CD4 Th2. The other 13 types more distributed in the low-risk subgroup, which included activated B cell, activated dendritic cell, CD56bright natural killer cell, eosinophil, B cell naïve, immature dendritic cell, mast cell, MDSC, monocyte, NK cell, plasmacytoid dendritic cell, T cell follicular helper, type 17 T helper cell. However, the remaining eight types were not statistically significant. There were T cell CD4 activated memory, activated CD8 T cell, T cell CD8 activated memory, CD56dim natural killer cell, T cell CD8 central memory, macrophage, T cell regulatory (Tregs), T cell CD4 Th1. Eighthly, we used XCELL to obtain a picture based on the Wilcoxon test, which indicated the distribution of 36 types of immune cells in the high-risk and low-risk subgroups (Figure 10K).

Furthermore, we found that the distribution of 23 immune cells in the picture was statistically significant (P<0.05). And there were seven types more distributed in the high-risk subgroup. They were common lymphoid progenitor, macrophage M1, plasmacytoid dendritic cell, T cell CD4 memory, T cell CD4 Th1, T cell CD4 Th2, T cell CD8 naïve. The other 16 types were more distributed in the low-risk subgroup. They were B cell, CAF, class−switched memory B cell, common myeloid progenitor, endothelial cell, eosinophil, granulocyte−monocyte progenitor, hematopoietic stem cell, macrophage M2, mast cell, myeloid dendritic cell, myeloid dendritic cell activated, T cell CD4 activated memory, T cell CD4 central memory, T cell CD4 naïve, T cell NK. But the remaining 13 types were not statistically significant. They were B cell memory, B cell naïve, B cell plasma, macrophage, monocyte, neutrophil, NK cell, T cell CD4 (non−regulatory), T cell CD8, T cell CD8 central memory, T cell gamma delta, T cell regulatory (Tregs).

Based on Venn diagram, we found that 8 algorithms have multiple overlapping immune cell types (Figure 10L). We estimated three types of immune cells threatening the survival of patients in the high-TILPI subgroup based on multiple algorithms (Table S4). They were macrophage M0, T cell CD4 Th2, and T cell CD4 memory activated. On the contrary, five immune cells, including B cell, endothelial cell, eosinophil, mast cell, and T cell CD4 memory resting, prolonged the survival (Table S4).

TIME has not only immune cells but also numerous stromal components. We obtained the stroma score, immune score, estimate score, and tumor purity from ESTIMATE. Stromal and immune scores were calculated to relate to the levels of stromal invasion and immune cells and thus to infer the tumor tissue's tumor purity which meant the proportion of tumor cells in TIME. In the same way, stroma score (R=-0.14, P=0.0022), immune score (R=-0.16, P=0.00018), and estimate score (R=-0.16, P=0.00034) were negatively correlated with TILPI

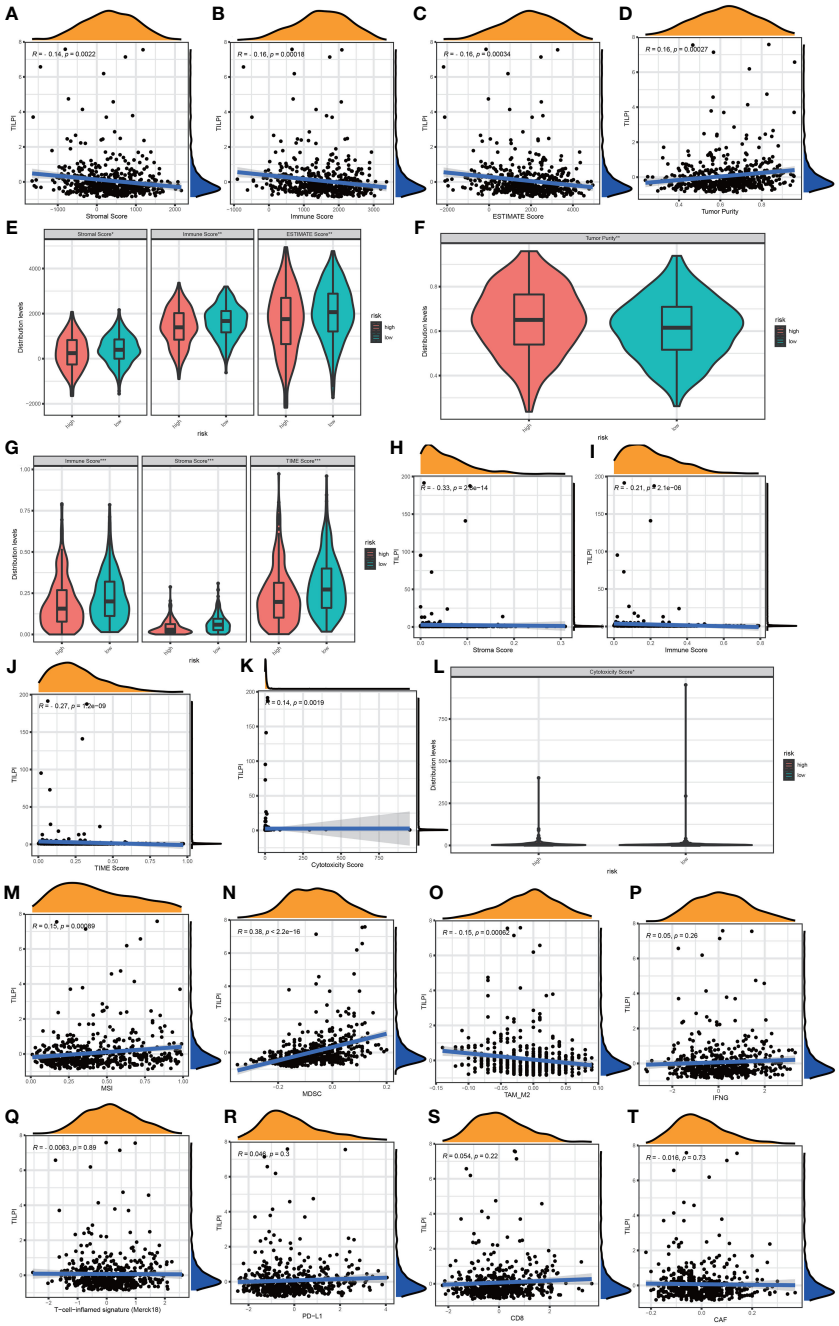


FIGURE 11
TILPI was negatively correlated with stroma score, immune score, and estimate score and positively related to tumor purity in the ESTIMATE algorithm **(A-F)**. TILPI also was negatively correlated with stroma score, immune score, and TIME score in the XCELL algorithm **(G-J)**. TILPI was positively correlated with the cytotoxicity score of MCPcounter **(K, L)**. MSI and MDSC were positively correlated with TILPI, but the TAM-M2 was the contrary. The other TIME components from TIDE weren't related to TILPI **(M-T)**. *P < 0.05; **P < 0.01; ***P < 0.001.

(Figures 11A-C). But the tumor purity was the opposite. It was positively correlated with TILPI (R=0.16, P=0.00027) (Figure 11D). And stroma score, immune score, and estimate score were also higher in the low-TILPI group, while tumor purity was higher in the high-TILPI group (P<0.01) (Figures 11E, F). Furthermore, we obtained three kinds of scores (stroma score, immune score, and TIME score) based on XCELL. Stromal score and immune score related to the level of infiltrating stromal and immune cells in TIME. We found that three kinds of scores were all higher in the low-TILPI group (P<0.001). It also verified that three types of scores were all negatively correlated with TILPI (Figure 11G). We also analyzed their correlation with TILPI by Pearson correlation analysis, and stroma score (R=-0.33, P<0.001), immune score (R=-0.21, P<0.001), and TIME score (R=-0.27, P<0.001) were all negatively correlated with TILPI (Figures 11H-J). Moreover,

10.3389/fonc.2023.1104137

MCPcounter is a model based on the gene expression matrix, and absolute abundance scores of eight immune cells and two stromal cells were generated for each sample. We first analyzed the Pearson correlation between cytotoxicity score based on MCPcounter and TILPI. According to the results, cytotoxicity score and TILPI were positively correlated (R=0.14, P=0.0019) (Figure 11K). To verify this result, we also performed the Wilcoxon test on the TCGA group to observe whether there was a cytotoxicity score difference between the high-TILPI and low-TILPI groups. And the result showed that the cytotoxicity score was higher in the high-TILPI group (P<0.05) (Figure 11L), which further verified cytotoxicity score and TILPI were positively correlated. In addition, we got eight types of scores

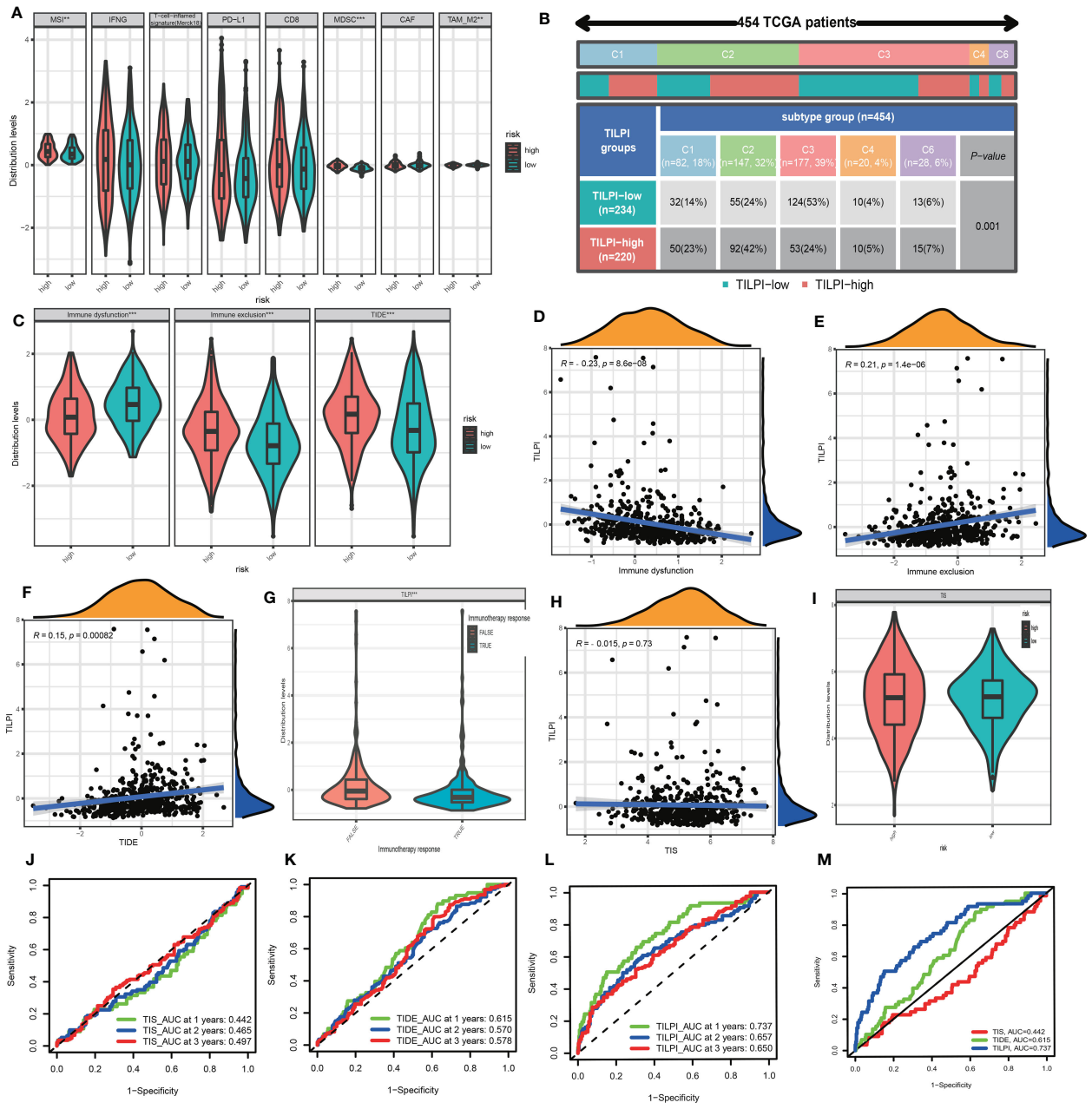


FIGURE 12
MSI and MDSC were positively correlated with TILPI, but the TAM-M2 was the contrary. The other TIME components from TIDE weren't related to TILPI **(A)**. TILPI was also correlated with the immune subtype of distribution. The low-TILPI subgroup was more distributed in immune C3 (n=124), and most of the samples in the high-TILPI subgroup were distributed in immune C2 (n=92) **(B)**. The TIDE model proved that the low-TILPI subgroup was more sensitive to immunotherapy **(C-F)**. The correlation between TILPI and immunotherapy once again demonstrated the sensitivity of the low-TILPI subgroup to immunotherapy **(G)**. TILPI wasn't correlated with the TIS model **(H-I)**. The ROC curves showed TILPI computational framework was better related to prognosis than the TIDE model and TIS model **(J-M)**. **P < 0.01; ***P < 0.001.

by TIDE: MSI, IFNG, Merck18, PD-L1, CD-8, MDSC, CAF, and TAM-M2. We found that MSI (R=0.15, P=0.00089) and MDSC (R=0.38, P<0.001) were positively correlated with TILPI by Pearson correlation analysis (Figures 11M, N). And TAM-M2 was negatively correlated with TILPI (R=-0.15, P=0.00062) (Figure 11O). Unfortunately, IFNG, Merck18, PD-L1, CD-8, and CAF were independent of TILPI (P>0.05) (Figures 11P-T). In the Wilcoxon test, MSI and MDSC were higher in the high-TILPI group (P<0.01, P<0.001), and TAM-M2 was higher in the low-TILPI group (P<0.01), while IFNG, Merck18, PD-L1, CD-8, and CAF were meaningless (P>0.05) (Figure 12A). At last, TILPI was also correlated with the immune subtype of distribution (P = 0.001) (Figure 12B). Only 454 samples were corresponding immune subtypes. They were divided into five subtypes, which were immune C1 (n=82), immune C2 (n=147), immune C3 (n=177), immune C4 (n=20), and immune C6 (n=28). The low-TILPI subgroup was more distributed in immune C3 (n=124), and most of the samples in the high-TILPI subgroup were distributed in immune C2 (n=92) (Figure 12B).

3.5 Association between TILPI and immunotherapy sensitivity

Tumor immune dysfunction and exclusion (TIDE) was a computing framework for evaluating the likelihood of tumor immune escape in gene expression profiles of tumor samples. We calculated the immune dysfunction score, immune exclusion score, and TIDE score. The Wilcoxon test showed immune dysfunction was higher in the low-TILPI group, immune exclusion was higher in the high-TILPI group, and TIDE was higher in the high-TILPI group, which was the most important (Figure 12C). The Pearson correlation analysis showed immune dysfunction was negatively correlated with TILPI (R=-0.23, P<0.001) (Figure 12D), while immune exclusion (R=0.21, P<0.001) and TIDE (R=0.15, P=0.00082) were positively correlated with TILPI (Figures 12E, F). As we know, high TIDE scores indicated severe immune evasion, and it was clear that the high-TILPI group was more prone to immune evasion than the low-TILPI group. Thus, there was no doubt that the low-TILPI group was more suitable for immunotherapy. What's more, the true immunotherapy response had lower TILPI than the false immunotherapy response (P< 0.001), which further verified that the low-TILPI group was more suitable for immunotherapy in the LUAD patients (Figure 12G).

The TIS was a marker of the immune microenvironment gene expression profile. It is based on eighteen genes to relate to the clinical benefit of PD-1-directed therapy. We first analyzed the correlation between TIS and TILPI, but TIS was independent of TILPI (P>0.05) (Figure 12H). And the Wilcoxon test showed TIS was no difference between high-TILPI and low-TILPI (Figure 12I). Furthermore, we calculated the reliability of three models (TIS, TIDE, TILPI) in LUAD patients' 1 year, 2 years, and 3 years OS. The results showed that TIS did not do a good job in relating to the OS of LUAD patients (AUC=0.442, AUC=0.465, AUC=0.497) (Figure 12J), and TIDE didn't either (AUC=0.615, AUC=0.570, AUC=0.578) (Figure 12K). Fortunately, TILPI had excellent credibility in relating to the OS of LUAD patients (AUC=0.737, AUC=0.657, AUC=0.650) (Figure 12L). In summary, TILPI has an advantage over TIS and TIDE, and the time-dependent ROC curves between TILPI (AUC=0.737), TIS (AUC=0.442), and TIDE (AUC=0.615) also verified it (Figure 12M).

Immunotherapy associations based on TIDE predictions also apply to LUSC and NSCLC populations. In the LUSC population, TILPI was significantly lower in the immunotherapy-responding group than in the non-responding group (P=0.0015) (Figure S1H). This was also the case in NSCLC patients, with the immunotherapy response group having a lower TILPI (P<0.001) (Figure S1I). The following five published transcriptomics signatures of immune responses confirmed that the low TILPI group may be more suitable for immunotherapy. As you can see, the low TILPI group had a higher TLS score, implying higher immune activity (Figures S1J, K). The low TILPI group induced weaker immune resistance (Figures S1L, M) and stronger ability to suppress immune resistance (Figures S1N, O), which also indicated that the low TILPI group may be more suitable for immunotherapy. The low TILPI group also had a higher Roh immune score, which was associated with higher immune activation (Figures S1P, Q). The Ock anti-CTLA-4 signature expression level of the low TILPI group was higher, which was also associated with better immunotherapy efficacy (Figures S1R, S). There was no difference in EaSIeR score between different TILPI groups, but there was a tendency for lower EaSIeR score in the lower TILPI group (Figures S1T, U).

3.6 Relationship between TILPI and drug sensitivity

Furthermore, we wished TILPI computing frameworks also relate to sensitive drugs to a specific population. Firstly, we conducted analyses of drug sensitivity based on the pRRophetic package updated in 2016. The judgmental standard of drug sensitivity was IC50. The patients with lower IC50 were sensitive to this drug. The filter was P value of the Wilcoxon test less than 0.05. Then we picked out 12 types of drugs more sensitive in the low-TILPI subgroup, which included AS605240, AZ628, Crizotinib, Erlotinib, KIN001-135, Phenformin, Salubrinal, TAK-715, TL-2-105, WZ3105, YM155, and Z-LLNle-CHO. And 82 types of drugs were also determined for patients with poor prognosis in the high-TILPI subgroup. Secondly, we conducted analyses of drug sensitivity based on the oncopredcit package. The judgmental standard of drug sensitivity was the same as the pRRophetic possessing package. Then 11 types of drugs were identified for the low-TILPI subgroup. They were ABT737, Axitinib, AZD6482, BMS.754807, Doramapimod, GSK269962A, PF.4708671, PRT062607, Ribociclib, SB505124, and ZM447439. And the number of sensitive drugs in the high-TILPI subgroup was 92. Thirdly, we used 95 up-regulated mRNAs from 267 immune TMB-derived oncogenic mRNAs to estimate drugs inhibiting these up-regulated genes base on the Connectivity Map (CMap). Then we got the results of 423,422 lines under different cell lines, dose, and time. We only selected known compounds and targets. And the absolute normalized CMap score of qualified drugs must be greater than 1.5. Therefore, we obtained 285 qualified drugs for LUAD patients (Figure 13A).

At last, we intersected the drugs based on pRRophetic, oncopredcit, and CMap for the population with poor prognosis.

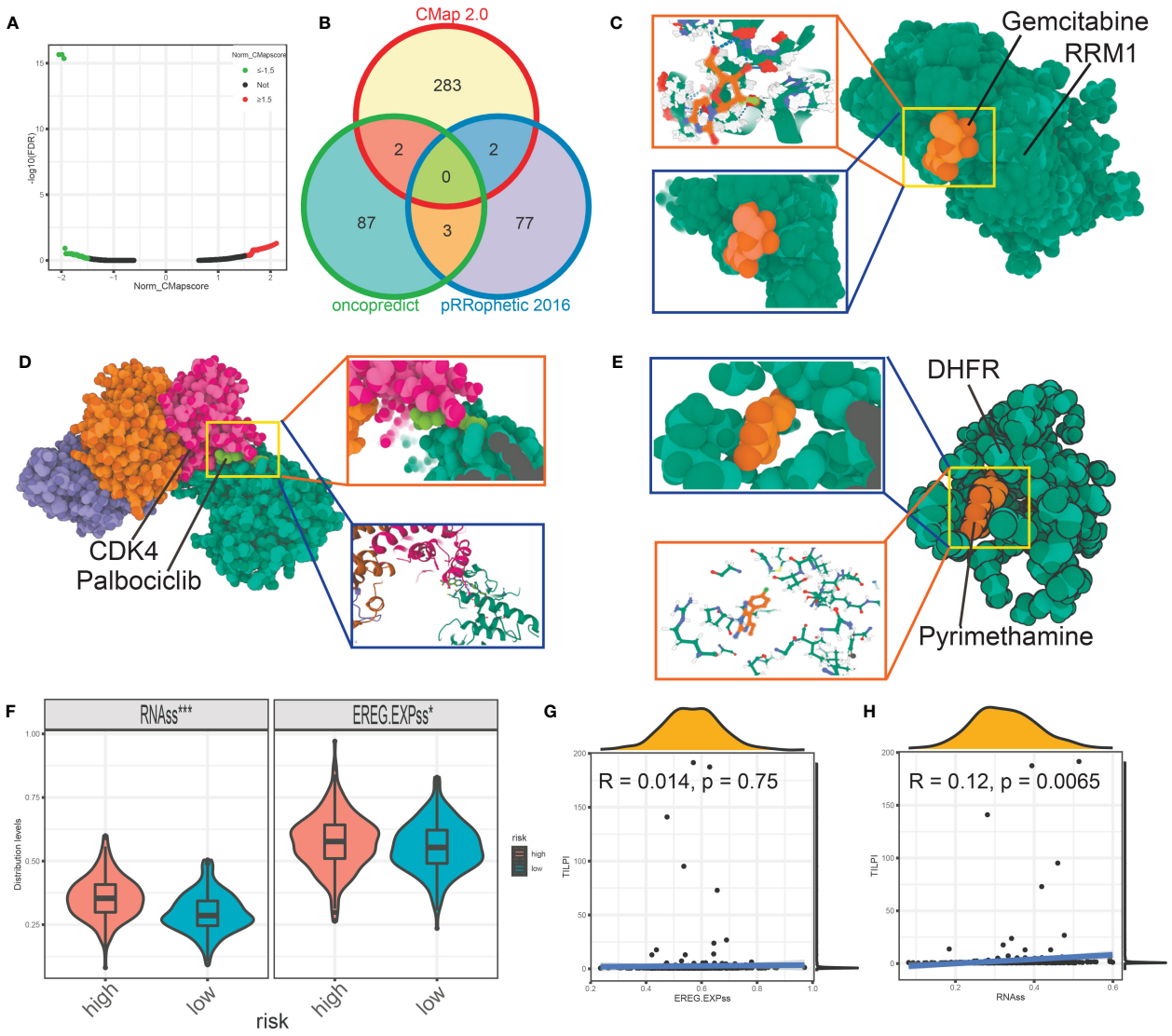


FIGURE 13
The volcano map showed the qualified drugs based on CMap **(A)**. The intersection analysis estimated seven more reliable drugs based on pRRophetic, oncopredict, and CMap **(B)**. The molecular model of gemcitabine docking RRM1 **(C)**. The molecular model showed the details that palbociclib docking CDK4 **(D)**. Pyrimethamine firmly docked its target protein DHFR **(E)**. The stemness score was different between different TILPI subgroups **(F, G)**. *P < 0.05; ***P < 0.001.

That is, in the high-TILPI subgroup. In conclusion, we identified seven sensitive drugs: Docetaxel, Gemcitabine, Paclitaxel, Palbociclib, Pyrimethamine, Thapsigargin, and Vinorelbine for poor prognostic population (Figure 13B). Docetaxel, Paclitaxel, and Vinorelbine are tubulin inhibitors that target ABCB1, BCL2, CYP2C8, MAP2, MAP4, MAPT, NR1I2, TLR4, and numerous subtypes of TUB proteins. And Gemcitabine is the ribonucleoside reductase inhibitor targeting RRM1, CMPK1, RRM2, and TYMS. Palbociclib is the CDK inhibitor that targets CDK4, CDK6, and CCND3. Pyrimethamine is the dihydrofolate reductase inhibitor and targets DHFRP1, HEXA, STAT3, DHFR, and SLC47A1. Thapsigargin is the ATPase inhibitor that targets ATP2A1.

Furthermore, we constructed molecular docking models to evaluate the affinity of 7 candidate drugs to their targets. Firstly, the best binding energy of Gemcitabine for RRM1 was -66.514 kcal/mol (Figure 13C). It indicates that the affinity between Gemcitabine and RRM1 was low, as seen from the figure (Figure 13C). Nevertheless, the best binding energy of the docking model between palbociclib and CDK4 was -7.573 kcal/mol (Figure 13D). The best binding energy for pyrimethamine to its target DHFR was -4.056 kcal/mol (Figure 13E). It was without doubt that palbociclib and pyrimethamine had low binding energy for their targets, indicating high stable binding and potential. Unfortunately, we failed to construct the molecular docking models of docetaxel,

thapsigargin, paclitaxel, and vinorelbine. Seven types of sensitive drugs for the high-TILPI subgroup, especially Palbociclib and pyrimethamine, were worth further exploration.

Based on network pharmacology, we next searched for potential targets of candidate drugs targeting LUAD. A total of 8907 LUAD target genes were collected from genecard database. The numbers of drug targets from SwissTargetPrediction, Batmant-TCM, and Pubchem databases were as follows: docetaxel (341 targets), gemcitabine (116 targets), paclitaxel (244 targets), palbociclib (315 targets), pyrimethamine (65 targets), thapsigargin (1337 targets), and vinorelbine (115 targets). There may be 8 targets of docetaxel acting on LUAD, including ABCA3, TP53, STK11, BIRC5, EGFR, ERBB2, KRAS, and RB1 (Figure S2A). The potential targets of gemcitabine for LUAD are TP53, KRAS, HYAL2, EGFR, and ERBB2 (Figure S2B). The paclitaxel may act on TP53, BIRC5, EGFR, ERBB2, TXNRD1, and KRAS to control the progression of LUAD (Figure S2C). For palbociclib, its potential targets for LUAD treatment may be TP53, BIRC5, ERBB2, SMARCA4, MYC, KRAS, BRAF, and RB1(Figure S2D). However, the effective target of pyrimethamine for LUAD seems to be only TP53 (Figure S2E). The thapsigargin may kill LUAD cells by targeting BIRC5, IRS1, MYC, MVP, HMOX1, ERBB2, CADM1, and TP53 (Figure S2F). The vinorelbine may act on SMARCA4, EPCAM, and ERBB2 to treat LUAD (Figure S2G). In total, there are 18 possible targets of these 7 drugs for LUAD, including ABCA3, BIRC5, BRAF, CADM1, EGFR, EPCAM, ERBB2, HMOX1, HYAL2, IRS1, KRAS, MVP, MYC, RB1, SMARCA4, STK11, TP53, and TXNRD1 (Figure S2H). Of these, the targets BIRC5, ERBB2, KRAS, and TP53 played a role in more than half of the drugs. The R package limma was used to analyze the differences between the high and low TILPI groups, and 57 up-regulated genes were found in the high TILPI group (Figure S2I) (Table S5). There are 19 up-regulated genes that are potential targets of LUAD, among which ABCC2, F2, GAL, INHBE, and UGT2B7 may be potential targets of docetaxel, paclitaxel, pyrimethamine, and thapsigargin in the treatment of high TILPI group (Figure S2J).

In the NCI-60 cell lines of CellMiner database, thirteen of the 57 up-regulated genes of high TILPI group were associated with the therapeutic sensitivity of 7 candidate drugs ($P<0.05$) (Table S6). They are ABCC2, BEST3, CREB3L3, CYP24A1, DSG4, GAL, GIP, IGF2BP1, MUC13, RAB3B, TFF1, TRIM15 and UGT2B7 (Figure S2K).

3.7 Prediction of tumor evolutionary status based on computing framework

The stemness score reliably evaluates the similarity of tumor cells to stem cells. The higher stemness score was correlated with therapy resistance, tumor biological functions, and clinical characteristics. We found that the high-TILPI subgroup possessed higher EREG-mRNAss and RNA expression-based stemness scores (RNAss) (Figure 13F). It could be one of the reasons that the high-TILPI subgroup had a poor prognosis. And TILPI was also positively correlated with EREG-mRNAss but statistically meaningless ($R=-0.014$, $P=0.75$) (Figure 13G). In addition, TILPI was also positively correlated with RNAss ($R=0.12$, $P=0.0065$) (Figure 13H).

4 Discussion

Lung cancer remained second in the global cancer rankings in 2021 (1–3). And LUAD is the most common histological type of lung cancer (5, 6). So far, traditional imagology and histopathology are still the gold standards for diagnosing and prognosis of LUAD. However, we aimed to construct a prognostic model based on lncRNA expression. Until a few years ago, lncRNA was regarded as a superfluous substance transcribed by genes (49). However, some in-depth studies have proved that lncRNA was involved in the biological activities of genes (50–52). LncRNA is closely related to tumor function. We first identified 267 TMB-derived oncogenic mRNAs and 176 TMB-derived oncogenic lncRNAs, and we obtained 43 immune TMB-derived oncogenic mRNAs and 36 TMB-derived oncogenic lncRNAs based on Pearson analysis. Next, we explored the potential functions of 79 mRNAs and lncRNAs by meta scape. And then, we further found the pathogenesis pathways of LUAD. We ascertained 30 GO pathways and 21 KEGG pathways in functional enrichment analysis.

Based on KM method analyses, univariate Cox proportional risk regression, and multivariate Cox proportional risk regression, we further study found that six lncRNAs in TILncSet expression level were closely related to the patient's OS, and all of them were a negative correlation, in other words, they were risk factors (AC091057.1, AC129492.1, AC112721.1, TARID, AC114763.1, LINC00592). Some scholars also have pointed out that AC091057.1 is a risk factor for LUAD patients (53). Other researchers have found that AC129492.1 impacts the prognosis of patients with hepatocellular carcinoma, colon cancer, and osteosarcoma (54–56). It has been reported that AC112721.1 is abnormally expressed in patients with breast cancer and bladder cancer (57, 58). Confusingly, TARID has been shown to activate the expression of the tumor suppressor gene TCF21 by inducing promoter demethylation (59, 60), which is contrary to our results, and the specific reasons remain to be further studied. LINC00592 is ferroptosis-related lncRNA, which has been identified as an independent prognostic predictor of LUAD and may be involved in the immune response to LUAD (61). It can also be used as a prognostic marker for disease-free survival in patients with gastric cancer (62) and is differentially expressed in cervical cancer (63). Unfortunately, no studies on AC114763.1 have yet to be researched. It is expected that these lncRNAs will be further studied.

Based on TMB, we built a novel computing framework called TILPI. It was an innovative step forward. TILPI calculated a risk score for each patient based on the expression of TILncSet. After testing, TILPI distinguished the prognosis of different risk score subgroups well in all groups (training group, test group, and TCGA group), and the higher TILPI, the worse prognosis. The reliability of

TILPI was verified, as the AUC value of the TILPI based on time-dependent ROC curves in all groups was higher than or equal to 0.73.

Furthermore, the independence and correlation analysis of some clinical factors were carried out, and the results showed that TILPI was closely related to the degree of smoking, pathological T, pathological N, and pathological stage. Studies have found that, compared with the smoker LUADs, never-smoker LUADs have a higher prevalence of clinically actionable driver alterations (78%-92% v 49.5%; P<.0001) (64). It also suggests that never-smoking patients have a better prognosis. And TILPI wasn't associated with and independent of race and age. However, studies have found that East Asian LUADs have more stable genomes and better prediction accuracy than European LUADs (65). It may be due to the improper grouping method and insufficient sample size, which made for our failure to find the relationship between race and TILPI, and further improvement is needed. The specific relationship and mechanism between TILPI and other clinical factors remain to be further studied. Fortunately, the Norman plot based on TILPI could still relate to the prognosis of LUAD patients, and we look forward to its clinical performance in the future. Yan Li et al. presented a framework called bioRFR to quantify wellness-to-disease transition in cancer patients by gene expression. They considered that cancer does not progress linearly, making it difficult or impossible to recover once it passes a tipping point. BioRFR was able to identify if a patient has passed this tipping point and provide personalized treatment. We must consider it in the future (66).

We also found ten pathways pathogenic to LUAD and six protectives to LUAD patients through gene enrichment. The prognosis of LUAD patients could be improved by inhibiting these pathogenic pathways or activating these protective pathways. It was a significant finding. Hopefully, it will be validated in the clinic. In summary, we have constructed a new and effective prognostic model for LUAD, accurately distinguishing between low-risk and high-risk LUAD patients. Compared with expensive molecular tests, TILPI is cheaper and more convenient. However, our research is still limited to calculation and analysis, and biological studies are required in the future.

Then, we used eight algorithms to describe the immune infiltration landscape, and we screened out meaningful immune cells by the Wilcoxon test, which were selected from eight immune infiltration landscape sets. Then we took the intersection. Then we ended up with eight immune cells. They were B cell, endothelial cell, eosinophil, mast cell, T cell CD4 memory resting, T cell CD4 Th2, macrophage M0, and T cell CD4 memory activated. B cell, endothelial cell, eosinophil, mast cell, and T cell CD4 memory resting was more in the low-risk subgroup. T cell CD4 Th2, macrophage M0, and T cell CD4 memory activated were more in the high-risk subgroup. We also used ESTIMATE and XCELL algorithms to calculate the stroma score of each sample and combined them with the immune score to reflect tumor purity. The two algorithms' stroma scores were negatively correlated with TILPI. The higher the stroma score was, the lower TILPI was. As predicted, tumor purity was higher in the high TILPI group.

The predicted immunotherapy response based on TIDE suggested that LUAD patients with low TILPI may be better candidates for immunotherapy, and this possibility applied to LUSC and NSCLC patients as well. The five published transcriptomics signatures of immune responses confirmed that the low TILPI group may be more suitable for immunotherapy. They were the TLS signature, Jerby-Arnon immune resistance program, Roh immune score, Ock anti-CTLA-4 signature, and EaSIeR model. In conclusion, TILPI is a good predictor of TME status and is superior to TIDE and TIS. In the study of Xu et al., TIDE was also used to relate to the effect of immunotherapy in patients. The difference is that their results showed that patients in the high-risk group responded better to immunotherapy, while we concluded that the low-risk group responded better to immunotherapy (67). Further research is required on this issue.

We also reviewed the latest developments of seven sensitive drugs. Network pharmacological analysis suggested that there were 18 potential therapeutic targets for LUAD, including ABCA3, BIRC5, BRAF, CADM1, EGFR, EPCAM, ERBB2, HMOX1, HYAL2, IRS1, KRAS, MVP, MYC, RB1, SMARCA4, STK11, TP53, and TXNRD1. Differential analysis based on R package limma found 57 genes up-regulated in the high TILPI group. Among them, ABCC2, F2, GAL, INHBE, and UGT2B7 may be potential targets of docetaxel, paclitaxel, pyrimethamine, and thapsigargin in the treatment of high TILPI group. In the NCI-60 cell lines of CellMiner database, thirteen of the 57 up-regulated genes of high TILPI group were associated with the therapeutic sensitivity of 7 candidate drugs. Both docetaxel and paclitaxel belonged to taxanes, which, combined with platinum drugs, was the first-line treatment option for LUAD. It was reported that docetaxel prolonged OS versus ICI in NSCLC patients (68). Other researchers have also combined paclitaxel with other drugs, such as nanoparticle albumin and ICI (69, 70). Furthermore, gemcitabine plus platinum was the standard chemotherapy for squamous NSCLC. And some novel research proved that the combination of gemcitabine and other drugs exhibited synergistic antitumor efficacy, which included albumin-bound paclitaxel and ICI (71, 72). In addition, vinorelbine was not a new chemotherapy drug, but it still has powerful effects (73, 74). As for CDK inhibitors, palbociclib was more commonly used in breast cancer patients with RB mutations but has recently been studied in NSCLC (75, 76). Pyrimethamine is an antimalarial drug and has also been proven to have antitumor activity (77). But the clinical research between pyrimethamine and chemotherapy for lung cancer was lacking. This clinical research absence also existed in thapsigargin. In conclusion, the dominant position of docetaxel, paclitaxel, gemcitabine, and vinorelbine in the chemotherapy of LUAD must be emphasized. Nevertheless, the potential significance of palbociclib, pyrimethamine, and thapsigargin in the chemotherapy of LUAD waited for further research.

The computing framework combined with prognosis may be novel research related to immunotherapy and chemotherapy. This computing framework also played some roles but also possessed limitations. Firstly, the details of TILncSig affecting OS were waiting for deeper exploration by biologists. Secondly, the details of the immune cells differently distributed in different TILPI subgroups that affected OS need further biological research. Thirdly, we believed that the low-TILPI subgroup was more sensitive to immunotherapy, but this conclusion needs to be verified in the

clinic. Fourthly, the effects of 7 sensitive drugs in the high-TILPI subgroup must also be researched in the clinic. Fifthly, they need more abundant samples. The external datasets should be employed for validation rather than our limited TCGA datasets. Sixthly, prospective analyses of the TILPI computing framework are required because all datasets in the study are retrospective. In any case, we will try our best to solve these problems step by step in further research.

5 Conclusion

We identified TILncSig based on TMB-related genes by WGCNA, oncogenes, and immune genes in LUAD. Then we construct the TILPI computing framework to relate to individual prognosis. TILPI could also map the immune infiltration landscape of immune cells, tumor cells, and stromal components based on creatively combined analyses of multiple algorithms. Furthermore, the TILPI computing framework successfully identified different prognostic LUAD populations and selected sensitive immunotherapy/chemotherapy for them. We believed that the low-TILPI subgroup was more sensitive to ICI, and the high-TILPI subgroup had a better effect on seven drugs.

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 human participants were reviewed and approved by Guangdong Medical University ethics committee. The patients/participants provided their written informed consent to participate in this study.

Author contributions

XZ conceived the work. CW and WZ wrote and drafted the manuscript. CW, WZ, FH, WH, and XX discussed and edited the manuscript. XZ checked the accuracy of statistics and bioinformatics as an expert in statistics and bioinformatics. All authors contributed to the article and approved the submitted version.

Acknowledgments

We are grateful to all those who donated human tissue and built a public database. We thank Dr. Jin Wang (Soochow University) and Dr. Jianming Zeng (University of Macau), for generously sharing their experience and codes. CW thanks XX's beauty as flower and jade for providing him with research motivation.

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.2023.1104137/full#supplementary-material

SUPPLEMENTARY FIGURE 1
AC114763.1, LINC00592, TARID, and AC091057.1 independently predicted patient survival **(A-D)**. AC129492.1 and AC112721.1, on the other hand, did not perform well **(E, F)**. The survival probability of NSCLC patients in the high TILPI group was significantly lower than those in the low TILPI group **(G)**. In the LUSC population, TILPI was significantly lower in the immunotherapy-responding group than in the non-responding group **(H)**. Patients with NSCLC who responded to immunotherapy had a lower TILPI **(I)**. The low TILPI group had a higher TLS score **(J, K)**. The low TILPI group induced weaker immune resistance **(L, M)** and stronger ability to suppress immune resistance **(N, O)**. The low TILPI group also had a higher Roh immune score **(P, Q)**. The Ock anti-CTLA-4 signature expression level of the low TILPI group was higher **(R, S)**. There was no difference in EaSIeR score between different TILPI groups, **(T, U)**.

SUPPLEMENTARY FIGURE 2
There may be 8 targets of docetaxel acting on LUAD **(A)**. The potential targets of gemcitabine for LUAD are TP53, KRAS, HYAL2, EGFR, and ERBB2 **(B)**. The paclitaxel may act on TP53, BIRC5, EGFR, ERBB2, TXNRD1, and KRAS to control the progression of LUAD **(C)**. For palbociclib, there may be as many as eight targets for LUAD **(D)**. The effective target of pyrimethamine for LUAD seems to be only TP53 **(E)**. The thapsigargin may kill LUAD cells by eight targets **(F)**. The vinorelbine may act on SMARCA4, EPCAM, and ERBB2 to treat LUAD **(G)**. In total, there are 18 possible targets of these 7 drugs for LUAD **(H)**. Based on differential expression analysis, 57 up-regulated genes were found in the high TILPI group **(I)**. There are 19 up-regulated genes that are potential targets of LUAD **(J)**. In the NCI-60 cell lines of CellMiner database, thirteen of the 57 up-regulated genes of high TILPI group were associated with the therapeutic sensitivity of 7 candidate drugs **(K)**.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistic. *CA Cancer J Clin* (2019) 69:7–34. doi: 10.3322/caac.21551

2. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistic. *CA Cancer J Clin* (2021) 71:7–33. doi: 10.3322/caac.21654

3. 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

4. Hirsch FR, Scagliotti GV, Mulshine JL, Kwon R, Curran WJJr., Wu YL, et al. Lung cancer: current therapies and new targeted treatments. *Lancet* (2017) 389:299–311. doi: 10.1016/S0140-6736(16)30958-8

5. Herbst RS, Heymach JV, Lippman SM. Lung cancer. *N Engl J Med* (2008) 359:1367–80. doi: 10.1056/NEJMra0802714

6. Rami-Porta R, Call S, Dooms C, Obiols C, Sanchez M, Travis WD, et al. Lung cancer staging: a concise update. *Eur Respir J* (2018) 51:1800190. doi: 10.1183/13993003.00190-2018

7. Swanton C, Govindan R. Clinical implications of genomic discoveries in lung cancer. *N Engl J Med* (2016) 374:1864–73. doi: 10.1056/NEJMra1504688

8. Goodall GJ, Wickramasinghe VO. Rna in cancer. *Nat Rev Cancer* (2021) 21:22–36. doi: 10.1038/s41568-020-00306-0

9. Pan J, Fang S, Tian H, Zhou C, Zhao X, Tian H, et al. lncRNA JPX/miR-33a-5p/Twist1 axis regulates tumorigenesis and metastasis of lung cancer by activating wnt/beta-catenin signaling. *Mol Cancer* (2020) 19:9. doi: 10.1186/s12943-020-1133-9

10. Qu S, Jiao Z, Lu G, Yao B, Wang T, Rong W, et al. PD-L1 lncRNA splice isoform promotes lung adenocarcinoma progression *via* enhancing c-myc activity. *Genome Biol* (2021) 22:104. doi: 10.1186/s13059-021-02331-0

11. Han X, Jiang H, Qi J, Li J, Yang J, Tian Y, et al. Novel lncRNA UPLA1 mediates tumorigenesis and prognosis in lung adenocarcinoma. *Cell Death Dis* (2020) 11:999. doi: 10.1038/s41419-020-03198-y

12. 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) 378:1976–86. doi: 10.1056/NEJMoa1716078

13. Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med* (2014) 371:2189–99. doi: 10.1056/NEJMoa1406498

14. Chan TA, Yarchoan M, Jaffee E, Swanton C, Quezada SA, Stenzinger A, et al. Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. *Ann Oncol* (2019) 30:44–56. doi: 10.1093/annonc/mdy495

15. Samstein RM, Lee CH, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet* (2019) 51:202–6. doi: 10.1038/s41588-018-0312-8

16. Fusco MJ, West HJ, Walko CM. Tumor mutation burden and cancer treatment. *JAMA Oncol* (2021) 7:316. doi: 10.1001/jamaoncol.2020.6371

17. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 blockade in tumors with mismatch-repair deficiency. *N Engl J Med* (2015) 372:2509–20. doi: 10.1056/NEJMoa1500596

18. Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* (2017) 357:409–13. doi: 10.1126/science.aan6733

19. Hellmann MD, Nathanson T, Rizvi H, Creelan BC, Sanchez-Vega F, Ahuja A, et al. Genomic features of response to combination immunotherapy in patients with advanced non-Small-Cell lung cancer. *Cancer Cell* (2018) 33:843–852 e4. doi: 10.1016/j.ccell.2018.03.018

20. Anagnostou V, Bardelli A, Chan TA, Turajlic S. The status of tumor mutational burden and immunotherapy. *Nat Cancer* (2022) 3:652–6. doi: 10.1038/s43018-022-00382-1

21. Ellrott K, Bailey MH, Saksena G, Covington KR, Kandoth C, Stewart C, et al. Scalable open science approach for mutation calling of tumor exomes using multiple genomic pipelines. *Cell Syst* (2018) 6:271–81.e7. doi: 10.1016/j.cels.2018.03.002

22. Langfelder P, Horvath S. WGCNA: an r package for weighted correlation network analysis. *BMC Bioinf* (2008) 9:559. doi: 10.1186/1471-2105-9-559

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

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

25. Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, et al. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, cytoscape and EnrichmentMap. *Nat Protoc* (2019) 14:482–517. doi: 10.1038/s41596-018-0103-9

26. Balachandran VP, Gonen M, Smith JJ, Dematteo RP. Nomograms in oncology: more than meets the eye. *Lancet Oncol* (2015) 16:e173–80. doi: 10.1016/S1470-2045(14)71116-7

27. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* (2015) 12:453–7. doi: 10.1038/nmeth.3337

28. Sturm G, Finotello F, Petitprez F, Zhang JD, Baumbach J, Fridman WH, et al. Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. *Bioinformatics* (2019) 35:i436–45. doi: 10.1093/bioinformatics/btz363

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

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

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

32. Li B, Severson E, Pignon JC, Zhao H, Li T, Novak J, et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol* (2016) 17:174. doi: 10.1186/s13059-016-1028-7

33. Ru B, Wong CN, Tong Y, Zhong JY, Zhong SSW, Wu WC, et al. TISIDB: an integrated repository portal for tumor-immune system interactions. *Bioinformatics* (2019) 35:4200–2. doi: 10.1093/bioinformatics/btz210

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

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

36. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med* (2018) 24:1550–8. doi: 10.1038/s41591-018-0136-1

37. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The immune landscape of cancer. *Immunity* (2018) 48:812–830 e14. doi: 10.1016/j.immuni.2018.03.023

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

39. Cabrita R, Lauss M, Sanna A, Donia M, Skaarup Larsen M, Mitra S, et al. Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature* (2020) 577:561–5. doi: 10.1038/s41586-019-1914-8

40. Jerby-Arnon L, Shah P, Cuoco MS, Rodman C, Su MJ, Melms JC, et al. A cancer cell program promotes T cell exclusion and resistance to checkpoint blockade. *Cell* (2018) 175:984–997 e24. doi: 10.1016/j.cell.2018.09.006

41. Roh W, Chen PL, Reuben A, Spencer CN, Prieto PA, Miller JP, et al. Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. *Sci Transl Med* (2017) 9. doi: 10.1126/scitranslmed.aah3560

42. Ock CY, Hwang JE, Keam B, Kim SB, Shim JJ, Jang HJ, et al. Genomic landscape associated with potential response to anti-CTLA-4 treatment in cancers. *Nat Commun* (2017) 8:1050. doi: 10.1038/s41467-017-01018-0

43. Lapuente-Santana O, Van Genderen M, Hilbers PAJ, Finotello F, Eduati F. Interpretable systems biomarkers predict response to immune-checkpoint inhibitors. *Patterns (N Y)* (2021) 2:100293. doi: 10.1016/j.patter.2021.100293

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

45. Maeser D, Gruener RF, Huang RS. oncoPredict: an r package for predicting *in vivo* or cancer patient drug response and biomarkers from cell line screening data. *Brief Bioinform* (2021) 22. doi: 10.1093/bib/bbab260

46. Morris GM, Huey R, Olson AJ. Using AutoDock for ligand-receptor docking. *Curr Protoc Bioinf* (2008) 24:8.14.1–.40. doi: 10.1002/0471250953.bi0814s24

47. Reinhold WC, Sunshine M, Liu H, Varma S, Kohn KW, Morris J, et al. 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* (2012) 72:3499–511. doi: 10.1158/0008-5472.CAN-12-1370

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

49. Rion N, Ruegg MA. LncRNA-encoded peptides: more than translational noise? *Cell Res* (2017) 27:604–5. doi: 10.1038/cr.2017.35

50. Lee S, Kopp F, Chang TC, Sataluri A, Chen B, Sivakumar S, et al. Noncoding RNA NORAD regulates genomic stability by sequestering PUMILIO proteins. *Cell* (2016) 164:69–80. doi: 10.1016/j.cell.2015.12.017

51. Munschauer M, Nguyen CT, Sirokman K, Hartigan CR, Hogstrom L, Engreitz JM, et al. The NORAD lncRNA assembles a topoisomerase complex critical for genome stability. *Nature* (2018) 561:132–6. doi: 10.1038/s41586-018-0453-z

52. Wei C, Li M, Li X, Lyu J, Zhu X. Phase separation: "The master key" to deciphering the physiological and pathological functions of cells. *Adv Biol (Weinh)* (2022) 6:e2200006. doi: 10.1002/adbi.202200006

53. Jin D, Song Y, Chen Y, Zhang P. Identification of a seven-lncRNA immune risk signature and construction of a predictive nomogram for lung adenocarcinoma. *BioMed Res Int* (2020) 2020:7929132. doi: 10.1155/2020/7929132

54. Wu F, Wei H, Liu G, Zhang Y. Bioinformatics profiling of five immune-related lncRNAs for a prognostic model of hepatocellular carcinoma. *Front Oncol* (2021) 11:667904. doi: 10.3389/fonc.2021.667904

55. Yin T, Zhao D, Yao S. Identification of a genome instability-associated LncRNA signature for prognosis prediction in colon cancer. *Front Genet* (2021) 12:679150. doi: 10.3389/fgene.2021.679150

56. Zhang J, Huang C, Zhu G, He G, Xu W, Li J, et al. Selection of lncRNAs that influence the prognosis of osteosarcoma based on copy number variation data. *J Oncol* (2022) 2022:8024979. doi: 10.1155/2022/8024979

57. Wang J, Zhang C, Wu Y, He W, Gou X. Identification and analysis of long non-coding RNA related miRNA sponge regulatory network in bladder urothelial carcinoma. *Cancer Cell Int* (2019) 19:327. doi: 10.1186/s12935-019-1052-2

58. Vishnubalaji R, Alajez NM. Epigenetic regulation of triple negative breast cancer (TNBC) by TGF-beta signaling. *Sci Rep* (2021) 11:15410. doi: 10.1038/s41598-021-94514-9

59. Arab K, Park YJ, Lindroth AM, Schafer A, Oakes C, Weichenhan D, et al. Long noncoding RNA TARID directs demethylation and activation of the tumor suppressor TCF21 *via* GADD45A. *Mol Cell* (2014) 55:604–14. doi: 10.1016/j.molcel.2014.06.031

60. Arab K, Karaulanov E, Musheev M, Trnka P, Schafer A, Grummt I, et al. GADD45A binds r-loops and recruits TET1 to CpG island promoters. *Nat Genet* (2019) 51:217–23. doi: 10.1038/s41588-018-0306-6

61. Wang Y, Lu G, Xue X, Xie M, Wang Z, Ma Z, et al. Characterization and validation of a ferroptosis-related LncRNA signature as a novel prognostic model for lung adenocarcinoma in tumor microenvironment. *Front Immunol* (2022) 13:903758. doi: 10.3389/fimmu.2022.903758

62. Cheng C, Wang Q, Zhu M, Liu K, Zhang Z. Integrated analysis reveals potential long non-coding RNA biomarkers and their potential biological functions for disease free survival in gastric cancer patients. *Cancer Cell Int* (2019) 19:123. doi: 10.1186/s12935-019-0846-6

63. Yuan LY, Qin X, Li L, Zhou J, Zhou M, Li X, et al. The transcriptome profiles and methylation status revealed the potential cancer-related lncRNAs in patients with cervical cancer. *J Cell Physiol* (2019) 234:9756–63. doi: 10.1002/jcp.27661

64. Devarakonda S, Li Y, Martins Rodrigues F, Sankararaman S, Kadara H, Goparaju C, et al. Genomic profiling of lung adenocarcinoma in never-smokers. *J Clin Oncol* (2021) 39:3747–58. doi: 10.1200/JCO.21.01691

65. Chen J, Yang H, Teo ASM, Amer LB, Sherbaf FG, Tan CQ, et al. Genomic landscape of lung adenocarcinoma in East asians. *Nat Genet* (2020) 52:177–86. doi: 10.1038/s41588-019-0569-6

66. Li Y, Zhang SW. Resilience function uncovers the critical transitions in cancer initiation. *Brief Bioinform* (2021) 22. doi: 10.1093/bib/bbab175

67. Xu F, Huang X, Li Y, Chen Y, Lin L. m(6)A-related lncRNAs are potential biomarkers for predicting prognoses and immune responses in patients with LUAD. *Mol Ther Nucleic Acids* (2021) 24:780–91. doi: 10.1016/j.omtn.2021.04.003

68. Park K, Ozguroglu M, Vansteenkiste J, Spigel D, Yang JCH, Ishii H, et al. Avelumab versus docetaxel in patients with platinum-treated advanced NSCLC: 2-year follow-up from the JAVELIN lung 200 phase 3 trial. *J Thorac Oncol* (2021) 16:1369–78. doi: 10.1016/j.jtho.2021.03.009

69. Yoneshima Y, Morita S, Ando M, Nakamura A, Iwasawa S, Yoshioka H, et al. Phase 3 trial comparing nanoparticle albumin-bound paclitaxel with docetaxel for previously treated advanced NSCLC. *J Thorac Oncol* (2021) 16:1523–32. doi: 10.1016/j.jtho.2021.03.027

70. Taniguchi Y, Shimokawa T, Takiguchi Y, Misumi T, Nakamura Y, Kawashima Y, et al. A randomized comparison of nivolumab versus nivolumab + docetaxel for previously treated advanced or recurrent ICI-naive non-small cell lung cancer: TORG1630. *Clin Cancer Res* (2022) 28:4402–9. doi: 10.1158/1078-0432.CCR-22-1687

71. Paik PK, Kim RK, Ahn L, Plodkowski AJ, Ni A, Donoghue MTA, et al. A phase II trial of albumin-bound paclitaxel and gemcitabine in patients with newly diagnosed stage IV squamous cell lung cancers. *Clin Cancer Res* (2020) 26:1796–802. doi: 10.1158/1078-0432.CCR-19-3060

72. Zhou C, Wu L, Fan Y, Wang Z, Liu L, Chen G, et al. Sintilimab plus platinum and gemcitabine as first-line treatment for advanced or metastatic squamous NSCLC: results from a randomized, double-blind, phase 3 trial (ORIENT-12). *J Thorac Oncol* (2021) 16:1501–11. doi: 10.1016/j.jtho.2021.04.011

73. Kenmotsu H, Yamamoto N, Yamanaka T, Yoshiya K, Takahashi T, Ueno T, et al. Randomized phase III study of pemetrexed plus cisplatin versus vinorelbine plus cisplatin for completely resected stage II to IIIA nonsquamous non-Small-Cell lung cancer. *J Clin Oncol* (2020) 38:2187–96. doi: 10.1200/JCO.19.02674

74. Tada H, Mitsudomi T, Misumi T, Sugio K, Tsuboi M, Okamoto I, et al. Randomized phase III study of gefitinib versus cisplatin plus vinorelbine for patients with resected stage II-IIIA non-Small-Cell lung cancer with EGFR mutation (IMPACT). *J Clin Oncol* (2022) 40:231–41. doi: 10.1200/JCO.21.01729

75. Tao Z, Le Blanc JM, Wang C, Zhan T, Zhuang H, Wang P, et al. Coadministration of trametinib and palbociclib radiosensitizes KRAS-mutant non-small cell lung cancers *In vitro* and *In vivo*. *Clin Cancer Res* (2016) 22:122–33. doi: 10.1158/1078-0432.CCR-15-0589

76. Thangavel C, Boopathi E, Liu Y, Mcnair C, Haber A, Perepelyuk M, et al. Therapeutic challenge with a CDK 4/6 inhibitor induces an RB-dependent SMAC-mediated apoptotic response in non-small cell lung cancer. *Clin Cancer Res* (2018) 24:1402–14. doi: 10.1158/1078-0432.CCR-17-2074

77. Liu H, Qin Y, Zhai D, Zhang Q, Gu J, Tang Y, et al. Antimalarial drug pyrimethamine plays a dual role in antitumor proliferation and metastasis through targeting DHFR and TP. *Mol Cancer Ther* (2019) 18:541–55. doi: 10.1158/1535-7163.MCT-18-0936

Glossary

3D PCA	3D principal component analysis
AUC	area under curve
BP	biological process
CAF	cancer associated fibroblast
CAR10	CAR intergenic 10
CC	cell component
CIBERSORT	Cell type Identification by Estimating Relative Subsets of RNA Transcripts
CIBERSORT-ABS	CIBERSORT-absolute mode
CMap	connectivity map
CTL	cytotoxic T lymphocyte
EGFR	epidermal growth factor
EPIC	Estimating the Proportions of Immune and Cancer cells
EREG-mRNAss	epigenetically regulated-mRNA expression-based stemness score
ESTIMATE	Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression data
FDR	false discovery rate
GEO	Gene Expression Omnibus
GO	Gene Ontology
GS	gene significance
GSEA	gene set enrichment analysis
GTEx	Genotype-Tissue Expression
IC50	semi-inhibitory concentration
ICI	immune checkpoint inhibitors
IFNG	interferon gamma
ImmPort	immunology database and analysis portal
InnateDB	systems biology of the innate immune response
KEGG	Kyoto Encyclopedia of Genes and Genomes
KM	Kaplan-Meier
lncRNA	long non-coding RNA
LUAD	lung adenocarcinoma
LUSC	lung squamous cell carcinomas
M	distant metastasis
MCODE	Molecular Complex Detection
MCPCOUNTER	Microenvironment Cell Populations-counter
MDSC	myeloid-derived suppressor cell
ME	module eigengene
MF	molecular function
MM	module membership
MSI	microsatellite instability
N	regional lymph nodes
NSCLC	non-small cell lung cancer
OS	overall survival
PPI	protein-protein interaction
QUANTISEQ	Quantifying Immune Contexture of Human Tumors
RNAss	RNA expression-based stemness score
ROC	receiver operating characteristic
T	primary tumor
TCGA	The Cancer Genome Atlas
TIDE	tumor immune dysfunction and exclusion
TIgeneSet	TMB-derived immune gene set
TILncSet	TMB-derived immune lncRNA set
TILncSig	TMB-derived immune lncRNA signature
TILPI	TMB-derived immune lncRNA prognostic index
TIME	Tumor immune microenvironment
TIMER	Tumor Immune Estimation Resource
TImSet	TMB-derived immune mRNA set
TIS	tumor inflammation signature
TISIDB	Tumor and Immune System Interaction Database
TLS	tertiary lymphoid structures signature
TMB	tumor mutation burden
TOM	Topological Overlap Matrix
TPM	transcripts per kilobase of exon model per million mapped reads
Tregs	T cell regulatory
TRRUST	transcriptional regulatory relationships unraveled by sentence-based text mining
WGCNA	weighted gene co-expression network analysis
XCELL	digitally portraying the tissue cellular heterogeneity landscape
YB-1	Y-box-binding protein 1.

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 20 July 2023
DOI 10.3389/fonc.2023.1224455

Check for updates

OPEN ACCESS

EDITED BY
Hongda Liu,
Nanjing Medical University, China

REVIEWED BY
Mang Yu,
Stanford University, United States
Yuting Ke,
Massachusetts Institute of Technology, United States
Song Xu,
Tianjin Medical University General Hospital, China

*CORRESPONDENCE
Yuqing Huang
huangyuqing555@gmail.com

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

RECEIVED 17 May 2023
ACCEPTED 03 July 2023
PUBLISHED 20 July 2023

CITATION
Pei G, Wang D, Sun K, Yang Y, Tang W, Sun Y, Yin S, Liu Q, Wang S and Huang Y (2023) Deep learning-enhanced radiomics for histologic classification and grade stratification of stage IA lung adenocarcinoma: a multicenter study. *Front. Oncol.* 13:1224455. doi: 10.3389/fonc.2023.1224455

COPYRIGHT
© 2023 Pei, Wang, Sun, Yang, Tang, Sun, Yin, Liu, Wang and Huang. 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.

Deep learning-enhanced radiomics for histologic classification and grade stratification of stage IA lung adenocarcinoma: a multicenter study

Guotian Pei[1†], Dawei Wang[2†], Kunkun Sun[3], Yingshun Yang[1], Wen Tang[2], Yanfeng Sun[2], Siyuan Yin[2], Qiang Liu[1], Shuai Wang[1] and Yuqing Huang[1*]

[1]Department of Thoracic Surgery, Beijing Haidian Hospital (Haidian Section of Peking University Third Hospital), Beijing, China, [2]Institute of Advanced Research, Infervision Medical Technology Co. Ltd., Beijing, China, [3]Department of Pathology, Peking University People's Hospital, Beijing, China

Background: Preoperative prediction models for histologic subtype and grade of stage IA lung adenocarcinoma (LUAD) according to the update of the WHO Classification of Tumors of the Lung in 2021 and the 2020 new grade system are yet to be explored. We aim to develop the noninvasive pathology and grade evaluation approach for patients with stage IA LUAD via CT-based radiomics approach and evaluate their performance in clinical practice.

Methods: Chest CT scans were retrospectively collected from patients who were diagnosed with stage IA LUAD and underwent complete resection at two hospitals. A deep learning segmentation algorithm was first applied to assist lesion delineation. Expansion strategies such as bounding-box annotations were further applied. Radiomics features were then extracted and selected followed by radiomics modeling based on four classic machine learning algorithms for histologic subtype classification and grade stratification. The area under the receiver operating characteristic curve (AUC) was used to evaluate model performance.

Results: The study included 294 and 145 patients with stage IA LUAD from two hospitals for radiomics analysis, respectively. For classification of four histological subtypes, multilayer perceptron (MLP) algorithm presented no annotation strategy preference and achieved the average AUC of 0.855, 0.922, and 0.720 on internal, independent, and external test sets with 1-pixel expansion annotation. Bounding-box annotation strategy also enabled MLP an acceptable and stable accuracy among test sets. Meanwhile, logistic regression was selected for grade stratification and achieved the average AUC of 0.928, 0.837, and 0.748 on internal, independent, and external test sets with optimal annotation strategies.

Conclusions: DL-enhanced radiomics models had great potential to predict the fine histological subtypes and grades of early-stage LUADs based on CT images, which might serve as a promising noninvasive approach for the diagnosis and management of early LUADs.

KEYWORDS

deep learning, artificial intelligence, radiomics, model, lung adenocarcinoma

Introduction

Lung cancer remained the leading cause of cancer death worldwide with annually 2.1 million new lung cancer cases and 1.8 million deaths (1). Unfortunately, approximately 70% of these patients are diagnosed with locally advanced stages and metastatic disease, which results in low survival rates (2). Thus, early detection and treatment of lung cancer are essential to reduce mortality. With the widespread development of low-dose chest CT screening programs, the detection of ground-glass nodules (GGNs) is rapidly increasing. Early-stage lung adenocarcinomas (LUADs) often manifest as pure ground-glass nodules (GGNs) and part-solid nodules (PSNs), and the prognosis is significantly related to pathological subtypes of LUADs (3, 4). Sublobar resection (including wedge resection and segmentectomy) could be considered for some stage I non-small cell lung cancer (NSCLC) patients with pre-invasive adenocarcinoma (adenocarcinoma in situ, AIS), minimally invasive adenocarcinoma (MIA), or lepidic predominant adenocarcinoma, owing to its favorable prognosis (5). However, some subtypes (solid, micropapillary, and complex glandular) of LUADs often have a poor prognosis (6), indicating the necessity of lobectomy for these patients. Therefore, the accurate pre-judgment of pathological subtypes and gradings would benefit the selection of surgery type, prognosis, and personalized postoperative follow-up of stage I LUADs.

Currently, many radiomics models have been developed to classify main histologic subtypes of lung cancer, such as the differentiation of non-small cell lung cancer (NSCLC) and small cell lung cancers (SCLC) (7), the classification of lung adenocarcinomas (ADC) and squamous cell carcinomas (SCC) (8), the differentiation of ADC, SCC, and SCLC (9). Of note, studies on LUADs also focused on the histologic subtype classification, and most studies simplified the problem by dividing LUADs into a 2-category classification (IAC; non-IAC) according to their invasiveness (10). In addition to the invasiveness, subtypes indicative of poor prognoses, such as the invasive mucinous adenocarcinoma (IMA), are still rarely included in classification studies, especially for stage IA LUADs. Additionally, although some reports studied the identification of high-grade LUADs via radiomics, the systematic stratification of IAC grades according to the 2020 new grade system from the International Association for the Study of Lung Cancer (IASLC) Pathology Committee (6) was yet to be explored.

In this study, we focused on patients with stage IA LUADs and aimed to develop two consecutive radiomics models for their non-invasive histologic subtype classification and grade stratification. Of note, Deep learning (DL)-based pre-annotation strategy and expansion annotation strategies were utilized to study the influence of ROIs delineation on the performance of radiomics. In combination with multiple machine learning algorithms, stable radiomics models were selected based on their performance on internal, independent, and external testing sets and further underwent subgroup analysis, validating their potential in supporting the clinical decisions in the era of precise and personal medicine.

Materials and methods

The retrospective study was approved by the Institutional Reviewing Board (IRB) of Beijing Haidian Hospital and Peking University People's Hospital and the informed consent was waived by IRBs since patient information was anonymized to ensure privacy.

Study population

Patients who underwent chest surgery and were diagnosed with stage IA LUAD were enrolled from two medical centers for radiomics model development and external validation according to the following including and exclusion criteria. Three cohorts were eventually included from two hospitals and constitute three datasets, including development set, independent test set, and external test set.

The first cohort, comprising 236 patients treated at our institution between February 27, 2017, and May 7, 2021, included 180 primary lung cancer (PLC) patients with a single lesion and 56 multiple primary lung cancer (MPLC) patients. This dataset was used for radiomics development and was divided into training, validation, and internal testing subsets at a ratio of 16: 4: 5. The second cohort

Abbreviations: NSCLC, non-small cell lung cancer; GGNs, ground-glass nodules; LUAD, lung adenocarcinoma; IAC, invasive non-mucinous adenocarcinoma; AUC, area under the curve; DL, deep learning; ML, machine learning; LR, logistic regression; PCC, Pearson correlation coefficient; PCA, principal component analysis.

included 58 eligible patients treated between May 10, 2021, and Nov 3, 2021, and was used as an independent test set. Of note, to further evaluate the robustness and generalization of proposed radiomics models, 145 eligible patients who underwent treatment at the other hospital between Sep 15, 2016, and Nov 1, 2021, were enrolled in cohort 3 and served as the external test set. Diagrams of patient enrollment and data partition details can be found in Figure 1.

The inclusion criteria were as follows: *a*) patients with stage IA lung adenocarcinoma; *b*) those who underwent complete surgical excision; *c*) those with preoperative thin-sliced chest CT images. Patients were excluded if *a*) histological subtype or clinical information was missing; *b*) their CT images were not in compliance with the Digital Imaging and Communications in Medicine (DICOM) standards; *c*) CT images were discontinuous, missing, or damaged; *d*) annotating radiologists could not confidently annotate images.

CT acquisition

All the enrolled patients underwent chest CT examinations before surgical excision. Particularly, multi-slice spiral CT low-dose scans were performed using instruments from GE Healthcare (Chicago, Ill, USA), Philips Healthcare (Amsterdam, Netherlands), and United Imaging (Shanghai, China). The key scanning parameters were as follows: tube voltage of 120KV; reconstruction slice thickness from 0.625 to 2mm. All CT scans were saved in the picture archiving and communication system.

Deep learning segmentation algorithm-aided annotation of pulmonary nodules

Given that deep learning (DL)-based auxiliary diagnosis systems for pulmonary nodules have been well developed and launched in clinical settings (11, 12), a modified Faster R-CNN model trained on more than 11,000 chest CT scans to detect different types of pulmonary nodules was utilized to aid the annotation of targeted nodules (12). Briefly, the employed modified Faster R-CNN first detected the targeted nodules and a U-Net segmentation algorithm output the contour. Then, senior radiologists further corrected the delineation of interested pulmonary nodules and deleted untargeted nodule lesions. In such a way, consumption of the medical labor force was significantly reduced, and the annotation efficiency was greatly improved. The credibility of the DL-based segmentation algorithm in annotating pulmonary nodules was examined by comparing it with manual-corrected lesion contours.

Expansion strategies for ROI annotation

Previous studies revealed that peritumoral information could improve the model performance on invasiveness prediction of ADC (13) and histological subtype stratification in patients with NSCLC (14). Another previous radiomics studies reported that bounding-box delineation of ROI could achieve equivalent performance to precisely annotated ones (15). Considering the potential advantages of peritumoral areas in histologic classification tasks, in addition to the DL-aided manual-correction annotation strategy, we further explored the pixel-expansion annotation strategy for radiomics modeling by expanding lesion contours based on manual corrected ones. Particularly, we performed 1-pixel, 3-pixel, 5-pixel, and bounding-box expansions after the manual correction was completed. The representation of annotated lesions was presented in the Supplementary Figure 1. Summarily, we selected different ROIs in this study, encompassing the precise lesion ROI, the expanded ROI, and the bounding-box ROI of the designated lesions. To ensure the accurate localization of the targeted lesion on

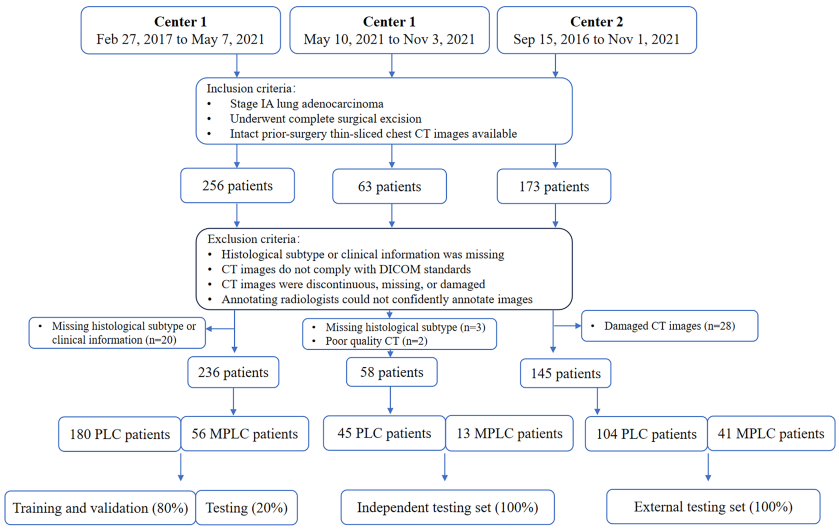


FIGURE 1
Diagram of patients enrollment and data partition. PLC = primary lung cancer, MPLC = multiple primary lung cancer, DICOM: Digital Imaging and Communications in Medicine.

CT images, a multidisciplinary team consisting of radiologists, thoracic surgeons, and pathologists collaborated in defining the targeted lesions. The impact of different annotation strategies on stage IA LUAD histologic subtype classification and invasive non-mucinous adenocarcinoma (IAC) grade stratification was analyzed in this study by comparing the performance of radiomics models.

Feature extraction

The PyRadiomics package (version 2.2.0) was called using Python (version 3.8.1) when performing radiomics feature extraction. Summarily, a total of 1454 features were extracted from the annotated ROIs, which belonged to 7 classes, including first-order (FOS), shape, Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM), Neighbouring Gray Tone Difference Matrix (NGTDM), and Gray Level Dependence Matrix (GLDM) features. Detailed information on extracted features was summarized in supplementary table 1.

Dimension reduction of extracted radiomics features

Pearson correlation coefficient (PCC) was first calculated and used to reduce the redundancy of the primary feature set, followed by the principal component analysis (PCA) approach which converted potentially correlated features into principal components that are linearly uncorrelated via orthogonal transformation (16). Features with a PCC <0.8 were retained after the first-round examination of feature redundancy. Subsequently, uncorrelated principal feature components were further obtained via PCA and utilized to develop radiomics models for histologic subtype classification and IAC grade stratification. Feature selection was accomplished by calling the scikit-learn (version 0.20.2) package.

Establishment of pathologic gold standard

Chest CT scans, pathological information, and clinical information was retrospectively collected from all included eligible patients and used to generate gold standard labels. Given the update of the WHO Classification of Tumors of the Lung in 2021 and IASLC grading system of IAC in 2020, histologic subtypes and IAC gradings of enrolled patients were all re-evaluated by an experienced pathologist before being utilized as the gold standard label in model development. In particular, histologic subtyping and grading were performed using the largest tumor sections in all cases, and the percentage of each histologic component was recorded in 5% increments according to the proposed IASLC grading system as follows: Grade 1, lepidic predominant tumors with no or less than 20% high-grade patterns (solid, micropapillary, and/or complex glandular patterns); Grade 2, acinar or papillary predominant tumors with no or less than 20% high-grade patterns; and Grade 3, any tumor with 20% or more of high-grade patterns.

Development and evaluation of radiomics models

Based on the five ROI annotation strategies mentioned above, four classic machine learning (ML) algorithms were utilized to develop radiomics models, including support vector machine (SVM), logistic regression (LR), and multi-layer perceptron (MLP), and eXtreme Gradient Boosting (XGBoost). The optimal hyper-parameters of ML algorithms were determined by the model performance on the validation set. The stable ML algorithm and potential practical annotation strategy were explored according to the model performance on the test datasets.

Radiomics models' performance was evaluated by classification sensitivity, specificity, precision, accuracy, F1 score, G-Mean, and area under the ROC curve (AUC). According to the study design, the first batch radiomics models focused on the classification of stage IA LUAD histological subtype classification, including precursor glandular lesions (PGL), MIA, IAC, and IMA. The second batch radiomics models were responsible for the stratification of IAC grade (6), which ranged from grade 1 to grade 3 (Figure 2).

Statistical analysis

Continuous variables were represented by the means ± SD while the categorical variables were expressed in terms of frequency and statistically analyzed by the Chi-square test. $P < 0.05$ was considered statistically significant. A two-sided 95% confidence interval for AUC was constructed following the approach of Hanley and McNeil (1982) (17). Cohen's Kappa coefficient was calculated in a confusion matrix to measure the agreement between pathological gold-standard and model predictions. All statistical analyses were performed with the R statistical package (The R Foundation for Statistical Computing, Vienna, Austria).

Results

Patient characteristics

From the two institutions, 256, 63, and 173 patients were initially eligible for the development set, independent test set, and external test set, respectively. However, due to missing histological subtype or clinical information, 20 (7.8%) and 3 (4.7%) patients were excluded. Additionally, 2 (3.2%) patients with motion-artifact induced poor quality CT scans and 28 (16.2%) patients with damaged CT scans were omitted. Thus, the final sample comprised 236, 58, and 145 patients in the development set, independent test set, and external test set (Figure 1).

In general, most of the included patients (79.04%, n=347) were non-smokers. Current (12.76%, n=56) and former smokers (8.20%, n=36) just count for a small portion of the studied population. Of note, 23.01% (n=101) of the population had a family history of cancer while 14.12% (n=62) of them had an alcohol intake history.

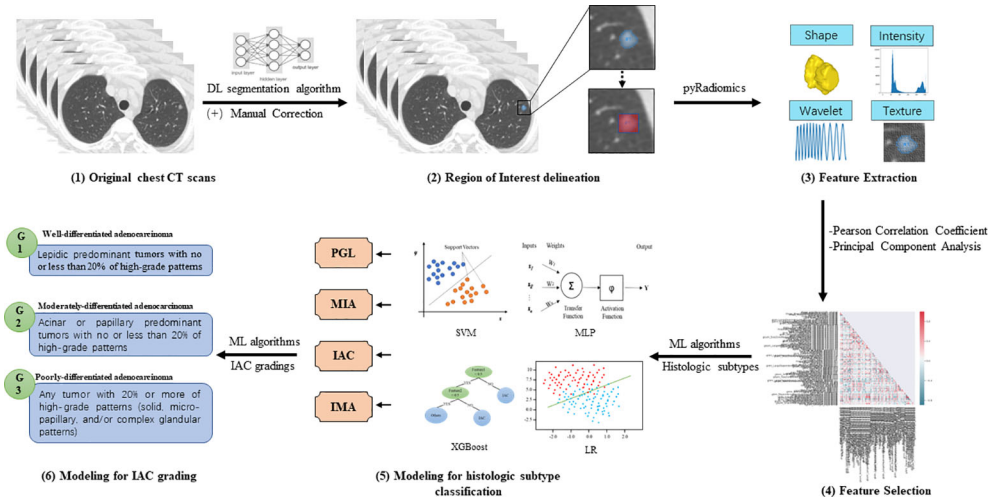


FIGURE 2
Illustration of the radiomics models for histologic subtype classification and IAC grading (1). Pre-operative chest CT scans were collected from enrolled patients for model development (2). Deep learning (DL)-based pulmonary nodule segmentation algorithm was utilized to pre-segment the target nodular lesions, followed by manual correction. Based on the manually edited region of interest (ROI), expansion strategies were applied to generate 1-pixel, 3-pixel, 5-pixel, and bounding-box masks of targeted lesions (3). PyRadiomics was utilized to extract radiomics features of different categories, including shape, intensity, wavelet, and texture features (4). Pearson correlation coefficient (PCC) and principal component analysis (PCA) were employed to reduce the dimensionality of extracted features (5). Classic machine learning (ML) algorithms were then used to develop radiomics models for classifying histologic subtypes of stage IA LUADs (6). Furthermore, ML algorithms were used to develop radiomics models for stratifying grades of invasive non-mucinous adenocarcinoma (IAC).

The most frequent surgical procedure was lobectomy (38.95%, n=171), followed by segmentectomy (26.65%, n=117) and wedge resection (21.41%, n=94); the rest of included patients (12.98%, n=57) received hybrid surgical procedures due to the presence of multiple primary lung cancer lesions. At adenocarcinoma lesion level, most of them presented as PSNs (53.22%, n=322), followed by GGN (30.58%, n=185), solid nodule (12.07%, n=73), and mass (4.13%, n=25). With respect to histologic subtypes, IAC (57.85%, n=350), MIA (30.91%, n=187), PGL (7.44%, n=45), and IMA (3.80%, n=23) were included. Additionally, most IAC lesions (84.57%, n=296) were categorized as Grade 2 according to the latest released grading system by the IASLC Pathology Committee.

Detailed characteristics of the included population in different datasets was summarized in Table 1. Notably, patients in the external test set were significantly older than those in the development set. Furthermore, family history of cancer was significantly less common among patients in external test set. It is also worth noting that the distribution of nodule types by density, histologic subtypes, and IAC gradings significantly varied across datasets due to different data collection timeframes. Notably, the independent test set lacked PGL and IMA lesions.

Analysis of radiomics features

A total of 1454 features were extracted from the annotated ROIs. A sum of 303 features with a Pearson correlation coefficient <0.8 was obtained after the first-round reduction of feature dimensionality. The correlation heatmap of selected features was presented in Supplementary Figure 2A. Subsequently, 40 principal feature components were preserved via PCA for the development of radiomics models. Principal component contribution rate was displayed in Supplementary Figure 2B. Details information about the extracted and selected features can be found in Supplementary Table 1.

Since PCA analysis selected feature components rather than certain features, we analyzed the significantly distinguished features (SDF) between each subtype based on PCC selected features in advance before developing the four-class histologic subtypes classification model and obtained 6 pairwise comparisons (PCs). Of the first-round selected 303 features, SDFs between each subtype were identified and grouped according to their identifying frequencies. Features were eventually divided into 7 groups, including SDFs in all PCs (n=46), 5PCs (n=19), 4PCs (n=17), 3PCs (n=17), 2PCs (n=16), 1PC (n=19), and none of the 6 PCs (n=169). These divided feature groups and their corresponding categories were displayed in the feature heatmap (Figure 3), and the details of features in each group were listed in Supplementary Table 2.

Selection of the optimal radiomics models for histologic subtypes classification and IAC grade stratification

DL-based nodule segmentation algorithms have enhanced the practicality of radiomics models. In the current study, we further employed five annotation strategies and four ML algorithms to develop two batches of models for LUAD diagnosis, including histologic subtype classification and IAC grade stratification. We first selected the optimal ML algorithms for both tasks by comparing the models' performance under different annotation

TABLE 1 Clinical characteristics of enrolled patients.

Groups Information	Model development set			Independent test set			External test set			Test
	Overall	PLC	MPLC	Overall	PLC	MPLC	Overall	PLC	MPLC	p value
Patients*	236	180	56	58	46	12	145	104	41	
Age (y)	57.0 (29 ~ 83)	56.6 (29 ~ 83)	58.2 (39 ~ 79)	59.8 (34 ~ 84)	60.2 (39 ~ 84)	57.9 (34 ~ 76)	61 (54~66)	61 (55~66)	61 (52~66)	0.018
Sex										0.086
Female	158 (66.95%)	119 (66.11%)	39 (69.64%)	41 (70.69%)	32 (69.57%)	9 (75%)	83 (57.24%)	62 (59.62%)	21 (51.22%)	
Male	78 (33.05%)	61 (33.89%)	17 (30.36%)	17 (29.31%)	14 (30.43%)	3 (25%)	62 (42.76%)	42 (40.38%)	20 (48.78%)	
Smoking history										0.849
Current	33 (13.98%)	27 (15%)	6 (10.71%)	6 (10.34%)	3 (6.52%)	3 (25%)	17 (11.72%)	11 (10.58%)	6 (14.63%)	
Former	18 (7.63%)	13 (7.22%)	5 (8.93%)	4 (6.9%)	4 (8.7%)	0 (0%)	14 (9.66%)	10 (9.62%)	4 (9.76%)	
Never	185 (78.39%)	140 (77.78%)	45 (80.36%)	48 (82.76%)	39 (84.78%)	9 (75%)	114 (78.62%)	83 (79.81%)	31 (75.61%)	
Family history of cancer										<0.010
Yes	64 (27.12%)	46 (25.56%)	18 (32.14%)	18 (31.03%)	15 (32.61%)	3 (25%)	19 (13.1%)	13 (12.5%)	6 (14.63%)	
No	172 (72.88%)	134 (74.44%)	38 (67.86%)	40 (68.97%)	31 (67.39%)	9 (75%)	126 (86.9%)	91 (87.5%)	35 (85.37%)	
Alcohol intake history										0.743
Yes	33 (13.98%)	25 (13.89%)	8 (14.29%)	10 (17.24%)	8 (17.39%)	2 (16.67%)	19 (13.1%)	15 (14.42%)	4 (9.76%)	
No	203 (86.02%)	155 (86.11%)	48 (85.71%)	48 (82.76%)	38 (82.61%)	10 (83.33%)	126 (86.9%)	89 (85.58%)	37 (90.24%)	
Lesion†	308	180	128	76	46	30	221	104	117	
Nodule type by density										<0.010
GGN	94 (30.52%)	38 (21.11%)	56 (43.75%)	41 (53.95%)	21 (45.65%)	20 (66.67%)	50 (22.62%)	4 (3.85%)	46 (39.32%)	
mGGN	175 (56.82%)	113 (62.78%)	62 (48.44%)	27 (35.53%)	21 (45.65%)	6 (20%)	120 (54.3%)	65 (62.5%)	55 (47.01%)	
Solid	34 (11.04%)	25 (13.89%)	9 (7.03%)	8 (10.53%)	4 (8.7%)	4 (13.33%)	31 (14.03%)	21 (20.19%)	10 (8.55%)	
mass	5 (1.62%)	4 (2.22%)	1 (0.78%)	0 (0%)	0 (0%)	0 (0%)	20 (9.05%)	14 (13.46%)	6 (5.13%)	
Histologic subtypes										<0.010
AIS	17 (5.52%)	5 (2.78%)	12 (9.38%)	0 (0%)	0 (0%)	0 (0%)	11 (4.98%)	1 (0.96%)	10 (8.55%)	
AAH	8 (2.6%)	0 (0%)	8 (6.25%)	0 (0%)	0 (0%)	0 (0%)	9 (4.07%)	0 (0%)	9 (7.69%)	
MIA	100 (32.47%)	52 (28.89%)	48 (37.50%)	36 (47.37%)	18 (39.13%)	18 (60%)	51 (23.08%)	8 (7.69%)	43 (36.75%)	
IAC	169 (54.87%)	111 (61.67%)	58 (45.31%)	40 (52.63%)	28 (60.87%)	12 (40%)	141 (63.8%)	87 (83.65%)	54 (46.15%)	

(Continued)

TABLE 1 Continued

Groups Information	Model development set			Independent test set			External test set			Test
	Overall	PLC	MPLC	Overall	PLC	MPLC	Overall	PLC	MPLC	p value
IMA	14 (4.55%)	12 (6.67%)	2 (1.56%)	0 (0%)	0 (0%)	0 (0%)	9 (4.07%)	8 (7.69%)	1 (0.85%)	
IAC gradings										0.027
Grade 1	16 (9.47%)	11 (9.91%)	5 (8.62%)	9 (22.5%)	4 (14.29%)	5 (41.67%)	12 (8.51%)	11 (12.64%)	1 (1.85%)	
Grade 2	148 (87.57%)	95 (85.59%)	53 (91.38%)	30 (75%)	23 (82.14%)	7 (58.33%)	118 (83.69%)	69 (79.31%)	49 (90.74%)	
Grade 3	5 (2.96%)	5 (4.5%)	0 (0%)	1 (2.5%)	1 (3.57%)	0 (0%)	11 (7.8%)	7 (8.05%)	4 (7.41%)	

Unless otherwise indicated, data are numbers of patients. AIS, adenocarcinoma in situ; AAH, atypical adenomatous hyperplasia; MIA, minimally invasive adenocarcinoma; IAC, invasive non-mucinous adenocarcinoma; IMA, invasive mucinous adenocarcinoma; GGNs, ground-glass nodules; PLC, primary lung cancer; MPLC, multiple primary lung cancer.
*Data are the median, and data in parentheses are the interquartile range.
†Data are numbers of nodules.

strategies on three test sets. As depicted in Figures 4A–C, MLP with 1-pixel annotation exhibited optimal performance on histologic subtype classification on the internal test set, and maintained consistent and excellent performance on independent and external test sets, regardless of annotation strategies. Notably, the bounding-box annotation strategy yielded comparable results for histologic subtype classification on the independent and external sets. Concurrently, LR displayed an overall superior performance on IAC grade stratification in terms of accuracy (Figures 4D–F). However, the performance of LR varied with different annotation strategies for IAC grade stratification.

Subsequently, impacts of annotations on selected ML algorithms were further evaluated on three test sets in terms of AUC, sensitivity, specificity, precision, F1-score, and G-Mean (Supplementary Figure 3). It was observed that MLP for histologic subtype classification had no preference for a specific annotation strategy, while LR for IAC grade stratification showed a preference for certain data labeling strategies. Regarding the performance of the radiomics models on each class, we noted inferior results for those classes with insufficient sample sizes.

Performance evaluation of selected radiomics model for histologic subtypes classification

We first evaluated the performance of the radiomics model on histologic subtype classification. The MLP with 1-pixel expansion was selected as the representative model. This model achieved an AUC of 0.903, 0.905, 0.951, and 0.661 for PGL, MIA, IAC, and IMA lesions, respectively, on the internal test set. On the external test set, it achieved an AUC of 0.929 and 0.914 for MIA and IAC lesions. On the external test set, it achieved an AUC of 0.691, 0.841,0.747, and 0.600 for PGL, MIA, IAC, and IMA lesions, respectively (Figures 5A–C). Notably, the performance of MLP was compromised on the external test set. Meanwhile, the kappa coefficient of MLP reached 0.696, 0.534, and 0.473, which presented a substantial and moderate agreement between model-predicted histologic subtypes and ground truth (Figures 5D–F). A decrease in the accuracy of MLP was also observed among the internal, independent, and external test sets (Table 2). This discrepancy could potentially be attributed to the prevalence of

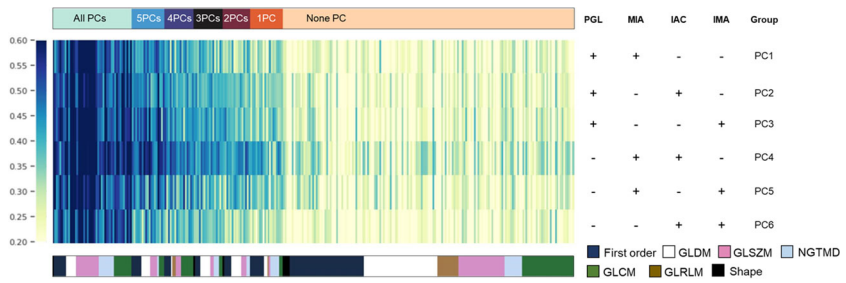


FIGURE 3
The most discriminative features for each histologic subtype. Based on PCC dimensionality reduction, distinguished features in a pair-wise comparison were analyzed to explain the potential key factors that distinguish them from each other. The detailed composition of each pair-wise comparison in each row is indicated in the right panel. Features were color-coded according to their category and listed from left to right based on their frequencies in pair-wise comparisons.

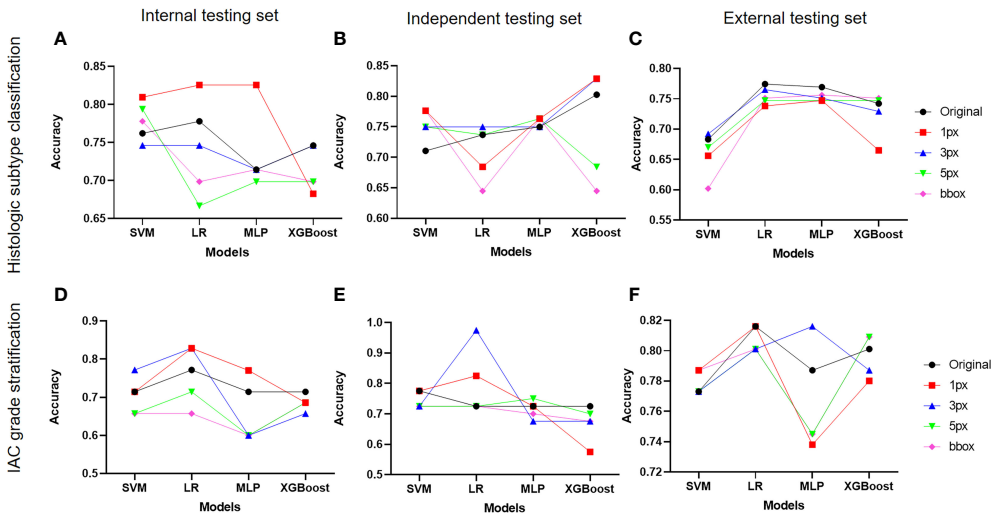


FIGURE 4
Impact of different annotation strategies on radiomics model performance. The performance of radiomics models developed on features from different annotation strategies were evaluated and compared in terms of accuracy. **(A-C)** displayed the accuracy of radiomics models for histologic subtype classification on the internal, independent, and external testing sets, respectively. **(D–F)** demonstrated the accuracy of radiomics models for IAC grade stratification on the internal, independent, and external testing sets.

challenging GGN lesions in the independent set and MPLC lesions in the external set. Of note, the accuracy of MLP remained stable on the external test sets (0.714 vs. 0.763 vs. 0.756) when the bounding-box annotation strategy was applied. The detailed performance metrics were summarized in Table 2.

Performance evaluation of optimal radiomics model for IAC grade stratification

We next evaluated the performance of the selected LR with optimal annotation strategies for IAC grade stratification. The LR model achieved an AUC of 0.911, 0.873, and 1.000 for grade 1, grade 2, and grade 3, respectively, on the internal testing set (Figure 5G), with a corresponding kappa coefficient of 0.547 (Figure 5J). However, on the independent test set, the LR model yielded a lower AUC of 0.771, 0.740, and 1.000 for grade 1, grade 2, and grade 3 respectively, and on the external test set, an AUC of 0.772, 0.644, and 0.878 for grade 1, grade 2, and grade 3, respectively. This suboptimal performance could be attributed to the imbalanced in sample size across the different grades (Figures 5H, I). The kappa coefficients of the LR model on the independent and external sets were 0.562 and 0.169, respectively (Figures 5K, L). Detailed performance metrics were summarized in Table 2.

Subgroup analysis of selected representative ML model performance on test sets

Notably, subgroup analyses of lesion numbers (PLC Vs MPLC), sex, nodule types by density (GGN vs PSNs vs solid), and age range were further performed (Figure 6). For histologic subtype classification, lower accuracy of MLP were observed on MPLC patients, significantly lower level was found on external test sets. Besides, significantly lower accuracy of MLP was also seen in GGN lesions on independent and external test sets. For IAC grade stratification, LR displayed significantly lower accuracy on male patients and solid nodules on the external test set. No significant difference of both two models was observed among other subgroups.

Discussion

Non-invasive preoperative prediction of pathological subtype and grade would greatly benefit the patients with stage IA LUADs in terms of the selection of surgery type, prognosis, and personalized postoperative follow-up. In this current study, we proposed two consecutive radiomics models for the diagnosis of patients with LUADs, including histologic subtype classification (PGL, MIA, IAC, and IMA) and IAC grade stratification (grade 1-3). Five annotation strategies and four ML algorithms were utilized for modeling. MLP and LR were selected as the optimal algorithms for histologic subtype classification and IAC grading stratification tasks, respectively, as supported by the overall better performance on different annotations on internal, independent, and external test sets. For histologic subtype classification, bounding-box annotation enabled an equivalent performance of MLP. Besides, distinguishing features between each pairwise comparison were revealed. Additionally, subgroup analyses validated the applicability of the radiomics models across cohorts with different sex, ages, and number of lesions.

Radiomics has been used since 2014 to solve clinical problems (18), and as its applications expand, efforts to streamline the process

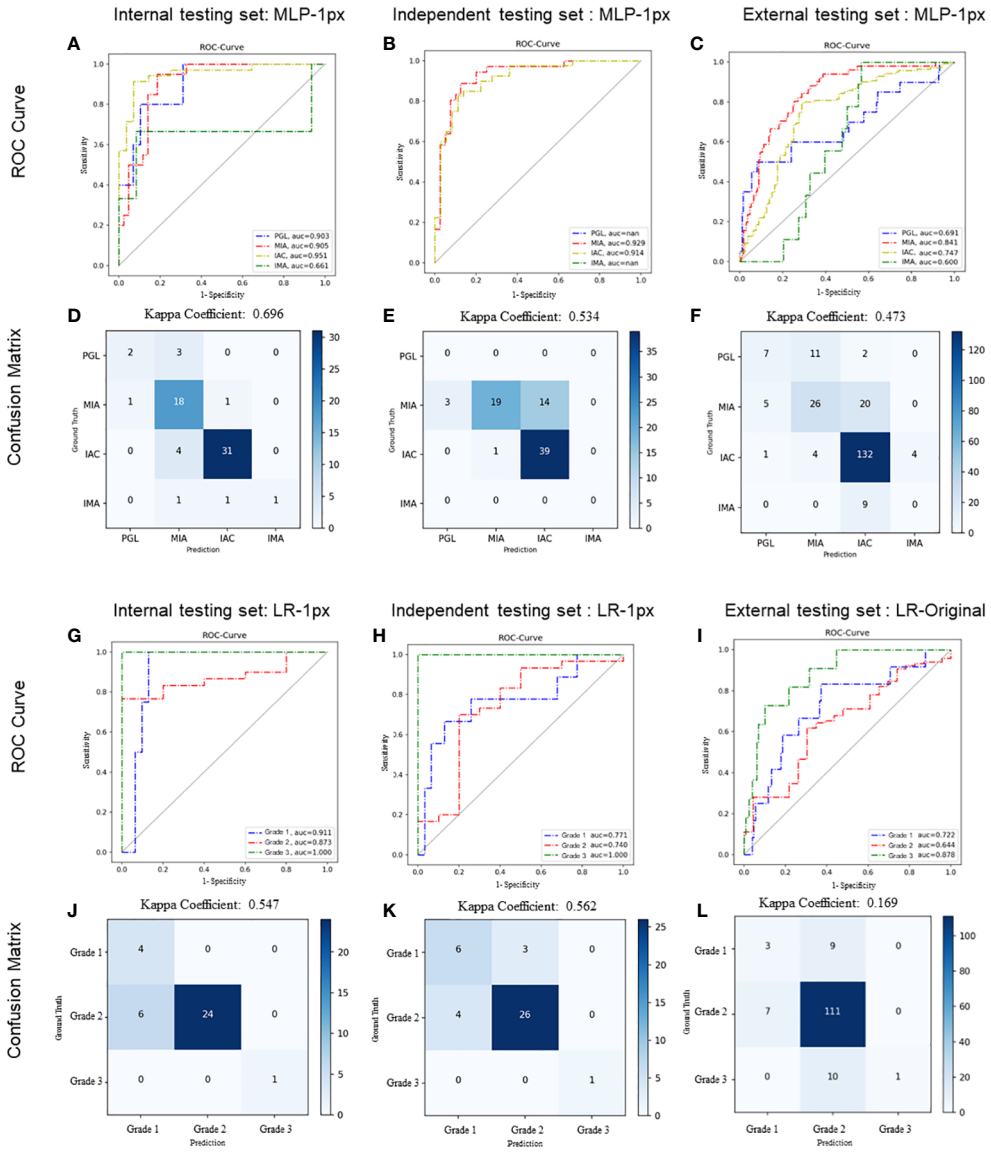


FIGURE 5
Performance of radiomics models on histologic subtype classification and IAC grading stratification. For histologic subtype classification, ROC curves were plotted to evaluate the performance of the Radiomic model in discriminating PGL, MIA, IAC, and IMA from the other three categories on internal **(A)**, independent **(B)**, and external **(C)** testing sets, respectively. Confusion matrices for four-category classification of PGL, MIA, IAC, and IM on internal **(D)**, independent **(E)**, and external **(F)** testing sets, respectively. For IAC grading stratification, ROC curves were plotted to evaluate the performance of the Radiomic model on internal **(G)**, independent **(H)**, and external **(I)** testing sets, respectively. Confusion matrices for the stratification of IAC grades (grade 1 to 3) on internal **(J)**, independent **(K)**, and external **(L)** testing sets, respectively. The exact number of true positives, false positives, true negatives and false negatives were listed. Kappa coefficients were calculated.

for clinical implementation are ongoing. Lesion annotation is often time-consuming and labor-intensive, limiting the clinical deployment of radiomics tools. Previous studies (19, 20) reported that semiautomatic lesion segmentation exhibited high agreement with manual delineations and could provide a significant reduction in interobserver variability. Some other studies utilized certain whole CT images (21), certain annotated slides (22), or bounding-box annotation (15) to develop models which could also avoid heavy annotation workload but might result in insufficient features. Given that DL segmentation algorithms for pulmonary nodules were well trained (11, 12), we then employed one to pre-segment the targeted lesions followed by a manual edition. The employed DL algorithm achieved an averaged Dice index of 0.94 (compared with manually edited contours), indicating the potential of the end-to-end or enhanced radiomics models by integrating DL segmentation algorithms into the classic radiomics modeling pathway. However, unlike the DL-enable end-to-end radiomics model in differentiating COVID-19 (22), we enrolled MPLC patients with other untargeted nodules that needed to be manually excluded before developing radiomics models. After all, as previously reported (23, 24), our hybrid approach avoided intensive labor force for lesion annotation.

10.3389/fonc.2023.1224455

TABLE 2 Detailed diagnostic metrics of radiomics models on internal, independent, and external test datasets.

Task	Test Sets	Subtypes/ Grades	Accuracy (95%CI)	AUC (95% CI)	Sensitivity	Precision	Specificity	F1-Score	G-Mean
Histological subtypes	Internal	PGL	0.825 (0.730-0.921)	0.903 (0.712-1.000)	0.400	0.667	0.983	0.500	0.516
		MIA		0.905 (0.824-0.970)	0.900	0.692	0.814	0.783	0.789
		IAC		0.951 (0.890-0.994)	0.886	0.939	0.929	0.912	0.912
		IMA		0.661 (0.049-1.000)	0.333	1.000	1.000	0.500	0.577
		Average		0.855	0.630	0.825	0.932	0.674	0.699
	Independent	PGL	0.763 (0.659-0.855)	-	-	-	0.961	-	-
		MIA		0.929 (0.863-0.978)	0.528	0.950	0.975	0.679	0.708
		IAC		0.914 (0.846-0.971)	0.975	0.736	0.611	0.839	0.847
		IMA		-	-	-	0.987	-	-
		Average		0.922	0.752	0.843	0.884	0.759	0.778
	External	PGL	0.747 (0.688-0.805)	0.691 (0.534-0.826)	0.350	0.538	0.970	0.424	0.434
		MIA		0.841 (0.779-0.894)	0.510	0.634	0.912	0.565	0.569
		IAC		0.747 (0.676-0.814)	0.936	0.810	0.612	0.868	0.870
		IMA		0.600 (0.494-0.712)	0.000	0.000	0.981	0.000	0.000
		Average		0.720	0.449	0.496	0.869	0.464	0.468
IAC grading	Internal	Grade 1	0.829 (0.714-0.943)	0.911 (0.750-1.000)	1.000	0.400	0.807	0.571	0.633
		Grade 2		0.873 (0.735-0.983)	0.800	1.000	1.000	0.889	0.894
		Grade 3		1.000 (0.500-1.000)	1.000	1.000	1.000	1.000	1.000
		Average		0.928	0.933	0.800	0.936	0.820	0.842
	Independent	Grade 1	0.825 (0.700-0.925)	0.771(0.553-0.943)	0.667	0.600	0.871	0.632	0.633
		Grade 2		0.740 (0.516-0.935)	0.867	0.897	0.700	0.881	0.882
		Grade 3		1.000 (0.500-1.000)	1.000	1.000	1.000	1.000	1.000
		Average		0.837	0.845	0.832	0.857	0.838	0.838
	External	Grade 1	0.816 (0.752-0.879)	0.722 (0.566-0.860)	0.250	0.300	0.946	0.273	0.274
		Grade 2		0.644 (0.512-0.756)	0.941	0.854	0.174	0.895	0.896
		Grade 3		0.878 (0.764-0.953)	0.090	1.000	1.000	0.167	0.302
		Average		0.748	0.427	0.718	0.707	0.445	0.491

Data in parentheses are 95% CIs. AUC, area under the receiver operating characteristic curve; PGL, precursor glandular lesions; MIA, minimally invasive adenocarcinoma; IAC, invasive non-mucinous adenocarcinoma; IMA, invasive mucinous adenocarcinoma.
"-" means not applicable (N/A).

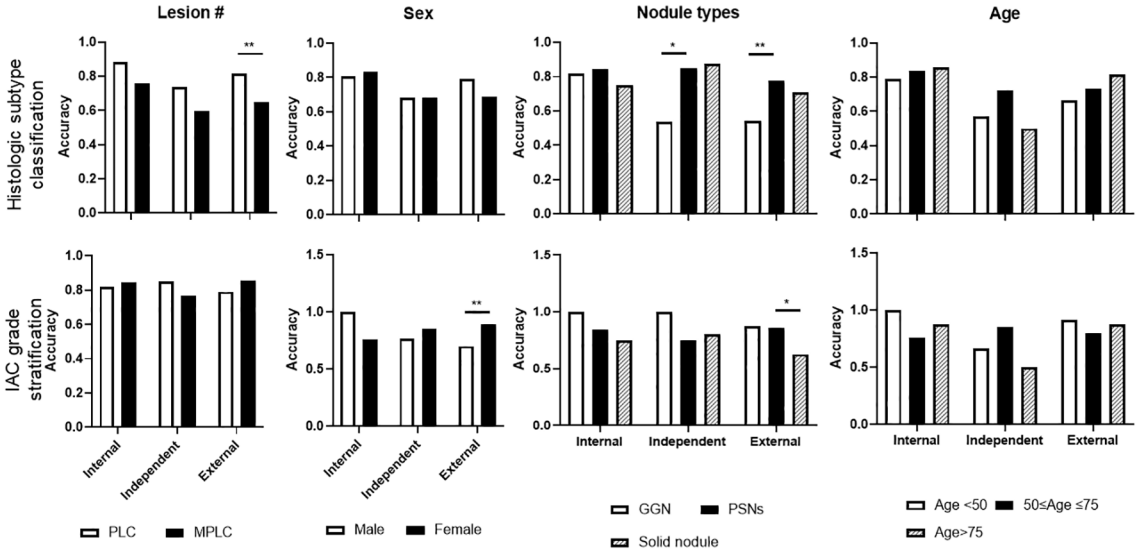


FIGURE 6
Subgroup analysis of selected representative ML model performance on test sets. Subgroup analyses were performed on histologic subtype classification and IAC grade stratification on internal, independent, and external testing sets, including target lesion numbers per patient, sex, nodule types, and age periods.

Since the easy-to-use bounding box annotation strategy was proved to be efficient in developing radiomics models for the diagnoses of gastric cancer and breast lesions (15, 25), we also examined the efficiency of an expansion strategy for the LUADs related tasks in our study by generating 1, 3, 5-pixel expanded and bounding-box (based on 5-pixel expansion) annotations. Notably, the 1-pixel expansion strategy, to some extent, enabled an overall stable performance of selected ML algorithms. An expansion strategy on cancerous lesions seemed to be a good option to enhance the model performance possibly by including more peritumoral features. Of course, the degree of expansion will need to be determined according to the situation. For histologic subtype classification, although the 1-pixel expansion strategy enabled an overall better performance, we also noticed the accuracy decline of MLP algorithm from internal to external test sets. Of note, accuracy of MLP remained acceptable and stable among test sets when applying the bounding-box strategy, indicating the practicality of the bounding-box strategy in this histologic subtype classification task. In contrast, the bounding-box strategy didn't perform well on the three-grade classification tasks in this study, indicating its applicability is algorithm- and context-dependent.

Another essential procedure for radiomics is dimensionality reduction which plays a key role in alleviating ML artifacts in the scenario of unbalanced datasets with small sample sizes (26). We utilized two classic approaches, PCC and PCA, to perform the dimensionality reduction in this study (27, 28). As an unsupervised method, PCA projected features into a dimensionally reduced set of uncorrelated variables called principal components via the linear orthogonal transformation, and outperformed the supervised technique in terms of generalizability capability (26). However, to solve the main problem of the variable's interpretation loss, we analyzed the distinguished features in a pair-wise comparison after PCC-based dimensionality reduction. The significant discriminating features between pair-wise comparisons may explain, to some extent, the key factors that distinguish them from each other.

Most previous related radiomics studies focused on binary classification in distinguishing NSCLC from SCLC, ADC from SCC, and IAC from other less invasive LUADs (7, 8). Given the update of the WHO Classification of Tumors of the Lung in 2021 and IASLC grading system of IAC in 2020 and the unique manifestations of IMA, we developed the first radiomics models for identifying four-category subtypes (PGL, IMA, IAC, and IMA) and three-category grades (grade1 to 3). We employed 4 classic ML algorithms and found that MLP and LR displayed an overall stable performance for four-category subtypes and three-category grades tasks, respectively. With the respect to identifying multi-class histologic subtypes, the selected representative MLP model in the current study achieved an average AUC of 0.855 and 0.922 on internal and independent testing sets, outperforming other models with an average AUC of 0.747 (4-category of NSCLC) (29), 0.833 (3-category subtypes of central lung cancer) (9), and 0.896 (4-category subtype of AAH, AIS, MIA, and IA) (30) in previous studies. Notably, the multiclass histological subtype classification model was not externally tested in previous studies, whereas the MLP achieved a mean AUC of 0.720 on external test set in this study. Meanwhile, few studies have reported the radiomics approach to stratify IAC grades according to the newly updated grading system. Instead, the radiomics approach was used to predict the micropapillary pattern that was reported to have a poor prognosis in a previous study (31). In comparison to multiparametric MRI-based radiomics approach for NSCLC grading (AUC 0.767) and contrast-enhanced CT-based radiomics signature for prediction of tumor differentiation degree (low and high degree, AUC 0.782) (32, 33), the selected representative LR algorithm for IAC grade stratification in this study achieved better

performance on both internal and independent testing sets (averaged AUC 0.928 and 0.837) and equivalent performance on external test set (averaged AUC 0.748), indicating the potential of CT-based radiomics approach in predicting histologic grades of IAC. Meanwhile, we noticed a dramatically decreased Kappa coefficient of LR algorithm on external test set, which caused by the miss classifications of grade1 and 3 into grade 2, suggesting the need of further improvement for IAC grading stratification algorithms by including more balanced data.

Of note, a previous study performed radiogenomic analyses of patients with stage I LUAD by an unsupervised consensus clustering approach to better classify patients with different prognoses, complementing the TNM system (34). In consistent, we developed supervised radiomic models on the patients with stage IA LUAD (not including IB) to enable the accurate differentiation of patients with poor prognosis at early stages according to histologic subtypes. To address the heterogeneity of LUAD, we further included the histologic type of IMA in the proposed model. IMA has different characteristics than non-mucinous adenocarcinoma in terms of histology, radiological and clinical features. Although IMA can show a lepidic growth pattern, invasive patterns are always present. Several studies have shown that IMA has a poor prognosis than non-mucinous adenocarcinoma (35–37). Additionally, IMA is commonly detected in the advanced stage and cannot be surgically treated. Therefore, our proposed radiomics models, to some extent, aided the accurate pre-judgment of patients' prognoses. Furthermore, they validated the revealed associations between CT-based radiomic features and known prognostic histologic factors, genomic drivers, and patient outcomes in the solid-type subgroup. In our subgroup analysis, the accuracy for differentiating histologic subtypes between GGN and PSNs lesions on both independent and external test sets were found to be significantly different.

There are some limitations to our study. The imbalance in histologic subtypes in the dataset compromised the performance of our proposed classification models, especially for PGL and IMA subtypes, and grade 3 lesions, which were less common in patients with operable clinical stage IA lung adenocarcinoma in clinical practice. The short follow-up of enrolled patients limited our ability to investigate the associations between radiomics and clinical features and the prognosis of patients with clinical stage IA LUAD. Although it is difficult for doctors to precisely classify those subtypes and grades, future work is also necessary to reveal the auxiliary effect of both models in promoting the diagnostic capabilities of these histologic subtypes, especially the identification of IMA, and IAC grades.

Despite these limitations, our results suggest that radiomics model, represented by MLP and LR, have great potential to predict the fine histological subtypes and grades of early-stage LUADs based on CT images, potentially providing a promising noninvasive approach for the diagnosis and management of early-stage LUADs.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by The Institutional Reviewing Board of Beijing Haidian Hospital and Peking University People's Hospital. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

YH designed the study and controlled the data used in this study. GP, KS, YY, QL, and SW participated in the collection of patients' data and manual correction of ROIs and provided clinical expertise. DW, WT, YS, and SY were responsible for modeling and testing. KS was responsible for quality control of the pathological samples. GP and DW prepared the main manuscript text. YH further polished the manuscript. All authors reviewed the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by Youth Foundation of Beijing Haidian Hospital (Grant No. KYQ2021002).

Acknowledgments

The abstract (Abstract Title: Classification of Stage IA Lung Adenocarcinoma into Histologic Subtypes on Computed Tomography Images Using Radiomics and Abstract Title: Grade stratification of stage IA invasive pulmonary adenocarcinoma on computed tomography images using radiomics) related to the content of this manuscript was presented as an e-poster at the Society of Thoracic Surgeons (STS) 2022 and presented as an EPOS Radiologist (educational) at the European Congress of Radiology (ECR) 2023 Annual Meeting, respectively.

Conflict of interest

Authors DW, WT, YS, and SY are employees of the company Infervision Medical Technology Co. Ltd.

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

Publisher's note

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.2023.1224455/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. Arriagada R, Dunant A, Pignon JP, Bergman B, Chabowski M, Grunenwald D, et al. Long-term results of the international adjuvant lung cancer trial evaluating adjuvant cisplatin-based chemotherapy in resected lung cancer. *J Clin Oncol* (2010) 28 (1):35–42. doi: 10.1200/JCO.2009.23.2272

3. Wang Y, Zheng D, Luo J, Zhang J, Pompili C, Ujiie H, et al. Risk stratification model for patients with stage I invasive lung adenocarcinoma based on clinical and pathological predictors. *Transl Lung Cancer Res* (2021) 10(5):2205–17. doi: 10.21037/tlcr-21-393

4. Yotsukura M, Asamura H, Motoi N, Kashima J, Yoshida Y, Nakagawa K, et al. Long-term prognosis of patients with resected adenocarcinoma *In situ* and minimally invasive adenocarcinoma of the lung. *J Thorac Oncol* (2021) 16(8):1312–20. doi: 10.1016/j.jtho.2021.04.007

5. Ito H, Suzuki K, Mizutani T, Aokage K, Wakabayashi M, Fukuda H, et al. Long-term survival outcome after lobectomy in patients with clinical T1 N0 lung cancer. *J Thorac Cardiovasc Surg* (2020) 11:S0022-5223(20)30054-4. doi: 10.1016/j.jtcvs.2019.12.072

6. Moreira AL, Ocampo PSS, Xia Y, Zhong H, Russell PA, Minami Y, et al. A grading system for invasive pulmonary adenocarcinoma: a proposal from the international association for the study of lung cancer pathology committee. *J Thorac Oncol* (2020) 15(10):1599–610. doi: 10.1016/j.jtho.2020.06.001

7. E L, Lu L, Li L, Yang H, Schwartz LH, Zhao B. Radiomics for classification of lung cancer histological subtypes based on nonenhanced computed tomography. *Acad Radiol* (2019) 26(9):1245–52. doi: 10.1016/j.acra.2018.10.013

8. Han Y, Ma Y, Wu Z, Zhang F, Zheng D, Liu X, et al. Histologic subtype classification of non-small cell lung cancer using PET/CT images. *Eur J Nucl Med Mol Imaging* (2021) 48(2):350–60. doi: 10.1007/s00259-020-04771-5

9. Li H, Gao L, Ma H, Arefan D, He J, Wang J, et al. Radiomics-based features for prediction of histological subtypes in central lung cancer. *Front Oncol* (2021) 11:658887. doi: 10.3389/fonc.2021.658887

10. Shi L, Shi W, Peng X, Zhan Y, Zhou L, Wang Y, et al. Development and validation a nomogram incorporating CT radiomics signatures and radiological features for differentiating invasive adenocarcinoma from adenocarcinoma *In situ* and minimally invasive adenocarcinoma presenting as ground-glass nodules measuring 5-10mm in diameter. *Front Oncol* (2021) 11:618677. doi: 10.3389/fonc.2021.618677

11. Wang Y, Yan F, Lu X, Zheng G, Zhang X, Wang C, et al. IILS: intelligent imaging layout system for automatic imaging report standardization and intra-interdisciplinary clinical workflow optimization. *EBioMedicine* (2019) 44:162–81. doi: 10.1016/j.ebiom.2019.05.040

12. Liu K, Li Q, Ma J, Zhou Z, Sun M, Deng Y, et al. Evaluating a fully automated pulmonary nodule detection approach and its impact on radiologist performance. *Radiol Artif Intell* (2019) 1(3):e180084. doi: 10.1148/ryai.2019180084

13. Wang X, Chen K, Wang W, Li Q, Liu K, Li Q, et al. Can peritumoral regions increase the efficiency of machine-learning prediction of pathological invasiveness in lung adenocarcinoma manifesting as ground-glass nodules? *J Thorac Dis* (2021) 13 (3):1327–37. doi: 10.21037/jtd-20-2981

14. Tang X, Huang H, Du P, Wang L, Yin H, Xu X. Intratumoral and peritumoral CT-based radiomics strategy reveals distinct subtypes of non-small-cell lung cancer. *J Cancer Res Clin Oncol* (2022) 148(9):2247–60. doi: 10.1007/s00432-022-04015-z

15. Liu D, Zhang W, Hu F, Wang L, Yin H, Xu X. A bounding box-based radiomics model for detecting occult peritoneal metastasis in advanced gastric cancer: a multicenter study. *Front Oncol* (2021) 11:777760. doi: 10.3389/fonc.2021.777760

16. Zhao Z, Xiao D, Nie C, Zhang H, Jiang X, Jecha AR, et al. Development of a nomogram based on preoperative bi-parametric MRI and blood indices for the differentiation between cystic-solid pituitary adenoma and craniopharyngioma. *Front Oncol* (2021) 11:709321. doi: 10.3389/fonc.2021.709321

17. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* (1982) 143(1):29–36. doi: 10.1148/radiology.143.1.7063747

18. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* (2014) 5:4006. doi: 10.1038/ncomms5006

19. Heye T, Merkle EM, Reiner CS, Davenport MS, Horvath JJ, Feuerlein S, et al. Reproducibility of dynamic contrast-enhanced MR imaging. part II. comparison of intra- and interobserver variability with manual region of interest placement versus semiautomatic lesion segmentation and histogram analysis. *Radiology* (2013) 266 (3):812–21.

20. Rios Velazquez E, Aerts HJ, Gu Y, Goldgof DB, De Ruysscher D, Dekker A. A semiautomatic CT-based ensemble segmentation of lung tumors: comparison with oncologists' delineations and with the surgical specimen. *Radiother Oncol* (2012) 105 (2):167–73. doi: 10.1016/j.radonc.2012.09.023

21. Wu X, Hui H, Niu M, Li L, Wang L, He B, et al. Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: a multicentre study. *Eur J Radiol* (2020) 128:109041. doi: 10.1016/j.ejrad.2020.109041

22. Zhang X, Wang D, Shao J, Tian S, Tan W, Ma Y, et al. A deep learning integrated radiomics model for identification of coronavirus disease 2019 using computed tomography. *Sci Rep* (2021) 11(1):3938. doi: 10.1038/s41598-021-83237-6

23. Zhang B, Ni-Jia-Ti MY, Yan R, An N, Chen L, Liu S, et al. CT-based radiomics for predicting the rapid progression of coronavirus disease 2019 (COVID-19) pneumonia lesions. *Br J Radiol* (2021) 94(1122):20201007. doi: 10.1259/bjr.20201007

24. Zhang M, Zeng X, Huang C, Liu J, Liu X, Xie X, et al. An AI-based radiomics nomogram for disease prognosis in patients with COVID-19 pneumonia using initial CT images and clinical indicators. *Int J Med Inform* (2021) 154:104545. doi: 10.1016/j.ijmedinf.2021.104545

25. Zhou J, Zhang Y, Chang KT, Lee KE, Wang O, Li J, et al. Diagnosis of benign and malignant breast lesions on DCE-MRI by using radiomics and deep learning with consideration of peritumor tissue. *J Magn Reson Imaging* (2020) 51(3):798–809. doi: 10.1002/jmri.26981

26. Ligero M, Torres G, Sanchez C, Diaz-Chito K, Perez R, Gil D. Selection of radiomics features based on their reproducibility. *Annu Int Conf IEEE Eng Med Biol Soc* (2019) 2019:403–8. doi: 10.1109/EMBC.2019.8857879

27. Chen C, Qin Y, Cheng J, Gao F, Zhou X. Texture analysis of fat-suppressed T2-weighted magnetic resonance imaging and use of machine learning to discriminate nasal and paranasal sinus small round malignant cell tumors. *Front Oncol* (2021) 11:701289. doi: 10.3389/fonc.2021.701289

28. Li Z, Zhang J, Song Y, Yin X, Chen A, Tang N, et al. Utilization of radiomics to predict long-term outcome of magnetic resonance-guided focused ultrasound ablation therapy in adenomyosis. *Eur Radiol* (2021) 31(1):392–402. doi: 10.1007/s00330-020-07076-1

29. Khodabakhshi Z, Mostafaei S, Arabi H, Oveisi M, Shiri I, Zaidi H. Non-small cell lung carcinoma histopathological subtype phenotyping using high-dimensional multinomial multiclass CT radiomics signature. *Comput Biol Med* (2021) 136:104752. doi: 10.1016/j.compbiomed.2021.104752

30. Deng J, Zhao M, Li Q, Zhang Y, Ma M, Li C, et al. Implementation of artificial intelligence in the histological assessment of pulmonary subsolid nodules. *Transl Lung Cancer Res* (2021) 10(12):4574–86. doi: 10.21037/tlcr-21-971

31. Song SH, Park H, Lee G, Lee HY, Sohn I, Kim HS, et al. Imaging phenotyping using radiomics to predict micropapillary pattern within lung adenocarcinoma. *J Thorac Oncol* (2017) 12(4):624–32. doi: 10.1016/j.jtho.2016.11.2230

32. Tang X, Bai G, Wang H, Guo F, Yin H. Elaboration of multiparametric MRI-based radiomics signature for the preoperative quantitative identification of the histological grade in patients with non-Small-Cell lung cancer. *J Magn Reson Imaging* (2022) 56(2):579–89. doi: 10.1002/jmri.28051

33. Chen X, Fang M, Dong D, Wei X, Liu L, Xu X, et al. A radiomics signature in preoperative predicting degree of tumor differentiation in patients with non-small cell lung cancer. *Acad Radiol* (2018) 25(12):1548–55. doi: 10.1016/j.acra.2018.02.019

34. Perez-Johnston R, Araujo-Filho JA, Connolly JG, Caso R, Whiting K, Tan KS, et al. CT-based radiogenomic analysis of clinical stage I lung adenocarcinoma with histopathologic features and oncologic outcomes. *Radiology* (2022) 303(3):664–72. doi: 10.1148/radiol.211582

35. Lee HY, Cha MJ, Lee KS, Lee HY, Kwon OJ, Choi JY, et al. Prognosis in resected invasive mucinous adenocarcinomas of the lung: related factors and comparison with resected nonmucinous adenocarcinomas. *J Thorac Oncol* (2016) 11:1064–73. doi: 10.1016/j.jtho.2016.03.011

36. Motono N, Matsui T, Machida Y, Usuda K, Uramoto H. Prognostic significance of histologic subtype in pStage I lung adenocarcinoma. *Med Oncol* (2017) 34:100. doi: 10.1007/s12032-017-0962-x

37. Kadota K, Nitadori J-I, Sima CS, Ujiie H, Rizk NP, Jones DR, et al. Tumor spread through air spaces is an important pattern of invasion and impacts the frequency and location of recurrences after limited resection for small stage I lung adenocarcinomas. *J Thorac Oncol* (2015) 10:806–14. doi: 10.1097/JTO.0000000000000486

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 20 July 2023
DOI 10.3389/fonc.2023.1196414

Check for updates

OPEN ACCESS

EDITED BY
Donghong Zhang,
Emory University, United States

REVIEWED BY
François Lucia,
Centre Hospitalier Regional Universitaire (CHU) de Brest, France
Han Yu,
Roswell Park Comprehensive Cancer Center, United States

*CORRESPONDENCE
Bertrand Routy
bertrand.routy@umontreal.ca

RECEIVED 29 March 2023
ACCEPTED 28 June 2023
PUBLISHED 20 July 2023

CITATION
Tonneau M, Phan K, Manem VSK, Low-Kam C, Dutil F, Kazandjian S, Vanderweyen D, Panasci J, Malo J, Coulombe F, Gagné A, Elkrief A, Belkaïd W, Di Jorio L, Orain M, Bouchard N, Muanza T, Rybicki FJ, Kafi K, Huntsman D, Joubert P, Chandelier F and Routy B (2023) Generalization optimizing machine learning to improve CT scan radiomics and assess immune checkpoint inhibitors' response in non-small cell lung cancer: a multicenter cohort study.
Front. Oncol. 13:1196414.
doi: 10.3389/fonc.2023.1196414

COPYRIGHT
© 2023 Tonneau, Phan, Manem, Low-Kam, Dutil, Kazandjian, Vanderweyen, Panasci, Malo, Coulombe, Gagné, Elkrief, Belkaïd, Di Jorio, Orain, Bouchard, Muanza, Rybicki, Kafi, Huntsman, Joubert, Chandelier and Routy. 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.

Generalization optimizing machine learning to improve CT scan radiomics and assess immune checkpoint inhibitors' response in non-small cell lung cancer: a multicenter cohort study

Marion Tonneau[1,2], Kim Phan[3], Venkata S. K. Manem[4,5], Cecile Low-Kam[3], Francis Dutil[3], Suzanne Kazandjian[6], Davy Vanderweyen[7], Justin Panasci[6], Julie Malo[1], François Coulombe[4], Andréanne Gagné[4], Arielle Elkrief[1,8], Wiam Belkaïd[1], Lisa Di Jorio[3], Michele Orain[4], Nicole Bouchard[9], Thierry Muanza[6,10], Frank J. Rybicki[3], Kam Kafi[3], David Huntsman[3], Philippe Joubert[4,11], Florent Chandelier[3] and Bertrand Routy[1,8]*

[1]Department of Cancer Research, Centre de Recherche du Centre Hospitalier Universitaire de Montréal (CRCHUM), Montreal, QC, Canada, [2]Université de Médecine, Lille, France, [3]Imagia Canexia Health, Montreal, QC, Canada, [4]Institut Universitaire de Cardiologie et de Pneumologie de Quebec, Université Laval, Québec City, QC, Canada, [5]Department of Mathematics and Computer Science, University of Quebec at Trois-Rivières, Trois-Rivières, QC, Canada, [6]Department of Medical Oncology, Jewish General Hospital, Montreal, QC, Canada, [7]Department of Radiology, Centre Hospitalier de Sherbrooke (CHUS), Sherbrooke, QC, Canada, [8]Hemato-Oncology Division, Centre Hospitalier de l'université de Montreal, Montreal, QC, Canada, [9]Department of Oncology, Centre Hospitalier de Sherbrooke (CHUS), Sherbrooke, QC, Canada, [10]Department of Radiation Oncology, Lady Davis Institute of the Jewish General Hospital, Montreal, QC, Canada, [11]Department of Pathology, Institut Universitaire de Cardiologie et de Pneumologie de Québec, Québec, QC, Canada

Background: Recent developments in artificial intelligence suggest that radiomics may represent a promising non-invasive biomarker to predict response to immune checkpoint inhibitors (ICIs). Nevertheless, validation of radiomics algorithms in independent cohorts remains a challenge due to variations in image acquisition and reconstruction. Using radiomics, we investigated the importance of scan normalization as part of a broader machine learning framework to enable model external generalizability to predict ICI response in non-small cell lung cancer (NSCLC) patients across different centers.

Methods: Radiomics features were extracted and compared from 642 advanced NSCLC patients on pre-ICI scans using established open-source PyRadiomics and a proprietary DeepRadiomics deep learning technology. The population was separated into two groups: a discovery cohort of 512 NSCLC patients from three

academic centers and a validation cohort that included 130 NSCLC patients from a fourth center. We harmonized images to account for variations in reconstruction kernel, slice thicknesses, and device manufacturers. Multivariable models, evaluated using cross-validation, were used to estimate the predictive value of clinical variables, PD-L1 expression, and PyRadiomics or DeepRadiomics for progression-free survival at 6 months (PFS-6).

Results: The best prognostic factor for PFS-6, excluding radiomics features, was obtained with the combination of Clinical + PD-L1 expression (AUC = 0.66 in the discovery and 0.62 in the validation cohort). Without image harmonization, combining Clinical + PyRadiomics or DeepRadiomics delivered an AUC = 0.69 and 0.69, respectively, in the discovery cohort, but dropped to 0.57 and 0.52, in the validation cohort. This lack of generalizability was consistent with observations in principal component analysis clustered by CT scan parameters. Subsequently, image harmonization eliminated these clusters. The combination of Clinical + DeepRadiomics reached an AUC = 0.67 and 0.63 in the discovery and validation cohort, respectively. Conversely, the combination of Clinical + PyRadiomics failed generalizability validations, with AUC = 0.66 and 0.59.

Conclusion: We demonstrated that a risk prediction model combining Clinical + DeepRadiomics was generalizable following CT scan harmonization and machine learning generalization methods. These results had similar performances to routine oncology practice using Clinical + PD-L1. This study supports the strong potential of radiomics as a future non-invasive strategy to predict ICI response in advanced NSCLC.

KEYWORDS

radiomics, Deeplearning, NSCLC, immunotherapy, DeepRadiomics

Introduction

The recent advent of radiomics by quantitative image analysis has been gaining interest in oncology as a novel strategy for cancer screening and predicting treatment response (1). Immune checkpoint inhibitors (ICIs) represent the standard of care for patients with advanced non-small cell lung cancer (NSCLC), and development of biomarkers represents a paramount interest (2–4). Nevertheless, primary resistance to ICIs remains unpredictable, reaching up to 60%, while the rate of secondary resistance approaches 100% (5–7). Assessment of PD-L1 expression in tumor tissue has been widely used to determine the therapeutic approach of either a single-agent anti-PD-1 inhibitor or the combination of platinum doublet with anti-PD-1 for patients with tumor PD-L1 expression ≥50% or <50%, respectively (4, 8).

Radiomics has been shown to predict $CD8^+$ T-cell infiltration and response to ICIs or radiotherapy (9–12). Additional studies attempted to determine PD-L1 expression, the only approved predictive biomarker in advanced NSCLC (13, 14). Nevertheless, validation of radiomic models requires large image datasets that include different cancer centers and a variety of computed tomography (CT) scanners. The necessary diversity is a major hurdle to validate published radiomics signatures in independent cohorts (15). Different image acquisition parameters and different reconstruction kernels with varying slice thicknesses alter the predictive potential of radiomics (16). Therefore, the development of signatures applicable across academic centers is a challenge that has stymied the adoption of radiomics in routine oncology clinical practice. Altogether, this highlights the importance of harmonizing image acquisition and reconstruction procedures to reduce multicenter variability before gathering data (17, 18). In recent years, research efforts have focused on developing a statistical harmonization strategy called ComBat (18–20). ComBat acts directly on already computed features, not on the original images. While a privacy advantage on one hand, this method is only capable of harmonizing for a single batch effect at a time; as further detailed (21), if variance in image acquisition and reconstruction protocols affects image properties, then different batches should be used for the same scanner corresponding to different settings. Furthermore, four assumptions have to be met for ComBat to generate valid results among which (1) covariates (if any) that might explain different distributions at two or more sites have to be identified and considered to redesign the original ComBat approach (2); the different sets of feature values to be realigned have to be independent, which challenges the very use of PyRadiomics known for many correlating features; and (3) determining a

single transformation with ComBat from data with different tissue or tumor types does not always lead to satisfactory data realignments, because different texture patterns are not necessarily affected identically by the image acquisition and reconstruction protocols (21). Taking these constraints altogether, the present study hypothesized that developing specific machine learning generalization methods rather than applying more complex feature harmonization strategies might prove more successful.

In this study, using conventional harmonization techniques together with AI generalization strategies, we showed that a radiomic signature generated in a discovery cohort from three independent cancer centers of NSCLC patients amenable to ICI to predict PFS at 6 months could be validated in a fourth cohort. This new method designed for generalizability rather than traditional performance has the potential to further the use of radiomics in routine oncology practice.

Methods

Study population

This retrospective study included 642 advanced NSCLC patients treated with anti-PD-1 alone or in combination with platinum-doublet chemotherapy between 2015 and 2021 in the chemotherapy-refractory or first-line settings. Signed, informed consent was obtained from each patient, and the study was approved by the Institutional Review Board [Human Ethics Committee (MP-02-2019-8091)] at four academic institutions where patient data were acquired: Centre Hospitalier de l'Université de Montréal (CHUM), Jewish General Hospital in Montreal (JGH), Institut Universitaire de Cardiologie et Pneumologie de Québec – Université Laval (IUCPQ-UL), and Centre Hospitalier de l'Université de Sherbrooke (CHUS). All patients with histology-proven stage III or IV NSCLC treated with ICI and with a pre-ICI CT scan were eligible for retrospective review. Response Evaluation Criteria in Solid Tumors (RECIST) criteria version 1.1 was used to assess tumor response, and all patients were followed until death or until the data were locked on 15 January 2022 (22).

Clinical data analysis

We separated the total population into two independent cohorts. All patients from CHUM, JGH, and IUCPQ comprised the discovery cohort (n = 512 patients) while patients from CHUS comprised the validation cohort (n = 130 patients). Baseline demographic and clinicopathological characteristics were compared between the discovery and validation cohort using chi-square or Fisher's exact test for categorical variables and Student's t-test or Mann–Whitney U test for continuous variables, as appropriate. Clinical outcome of PFS at 6 months was used as the stronger outcome marker for NSCLC patients amenable to ICIs as this clinical marker was found to be one of the most robust (3). All patients included had a PFS superior to 6 months or progressed before.

Harmonization process

CT scan normalization

Each primary lesion was manually annotated by a radiation oncologist or a radiologist for identifying the tumor's longest axis, on de-identified, pretreatment CT images. The following pre-processing steps were applied to all scans: resampling to 1-mm isometric voxels (to normalize pixel and slice thickness variation) followed by Hounsfield Unit (HU) truncation to a range −400 HU to 1,024 HU (to reduce the impact of artifacts on radiomics features), followed by image noise normalization using a Laplacian of Gaussian filter from the PyRadiomics library with hyperparameter values for sigma2 = {1,3} (20, 23) (see Figure 1).

Radiomics feature extraction

PyRadiomics features extraction

We used a three-stage process to determine the region of interest (ROI) used for extracting radiomic features. The first stage consisted of CT scan alignment achieved by principal component analysis (PCoA), followed by chest isolation through mathematical morphology-based denoising, and finally chest segmentation based on connected regions (24). In the second stage, the lung was automatically segmented based on the detected skin boundary, rough segmentation of lung contour, and pulmonary parenchyma refinement. Next, this lung segmentation was intersected with a clustering-based nodule mask to identify a nodule ROI agnostic of size, position, and spreading near or through the pleura, utilizing the relative symmetry of the lung (25). These ROIs were then assessed for clinical appropriateness in view of known shortcomings of segmentation techniques reported in the radiomics literature that extend beyond the objective of the current study but for which, nonetheless, we present an alternative in the form of a new DeepRadiomics method. From the segmentation, the PyRadiomics library v3.0.1, an open-source python package for the extraction of radiomics features from medical imaging, was used to extract 94 candidate radiomics features: 19 first-order features and 75 second-order features with a Laplacian of Gaussian (LoG) filter applied (26). The reproducibility of extracting PyRadiomics features from different segmentations has been well studied, and a high reproducibility was reported for first-order, Laplacian, Gaussian-filtered features and texture features, but low reproducibility for shape and wavelet features. Indeed, wavelet features were extracted but omitted from analyses due to the high association with the acquisition parameters.

DeepRadiomics features extraction

We proposed a new data-driven alternative to the traditional PyRadiomics method. We followed the emerging interest in deep learning models to provide suitable high-throughput extraction of quantitative imaging features from medical images (27, 28). A VGG16 backbone was pre-trained to learn image features followed by a SimCLR process, a self-learning framework for contrastive learning of visual representations (29, 30). The pre-

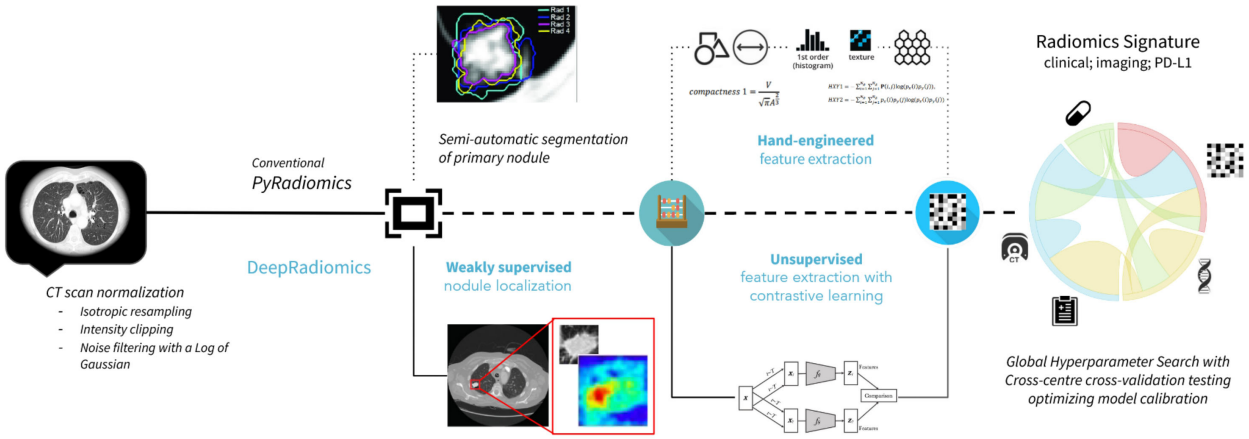


FIGURE 1
Radiomics workflows used in the study. The upper panel represents PyRadiomics pipeline including, sequentially, the segmentation input and hand-engineered feature extractions. The lower panel represents the DeepRadiomics pipeline with weakly supervised region extractions and automated feature learning extractions.

training datasets consisted of the public dataset [Lung Image Database Consortium (LIDC)], after applying the CT scan normalization procedure described in this article (31), in order to leverage a larger dataset of lung CT scans also obtained across multiple institutions and acquisition parameters. The pretraining was made on 2D patches of 48 × 48 pixels, centered on the nodules. For each image in a batch, we applied two different sets of data augmentation (random translation, rotation, flip, gaussian blur, and zoom), while training the model to correctly identify pairs of images representing the same nodule among other nodules in the same batch. In our proposed method, relevant images were processed using the pretrained SimCLR network, and automatically learned features were extracted from its last convolutional layer. Patch input could not be smaller than 32 × 32 due to the model architecture backbone, so we expanded context as needed for ROIs smaller than this. Finally, the full ROI bounding box derived from the annotation process was used instead of segmenting or delineating potential lesions as is usually required in radiomics, which we see as further contributing to clinical generalization.

Toward the generalizability of AI models across healthcare settings

We designed a global hyperparameter search framework (the "GHPS") to autonomously determine the final artificial intelligence (AI) model based on generalization-optimizing parameter configurations rather than performance-optimizing configurations. GHPS is ideally implemented by iteratively performing cross-center cross-validation testing over all combinations of parameters composed of the following four methods (32): (a) feature extraction, (b) feature selection, (c) model selection, and (d) model hyperparameter tuning. To reduce computational complexity, we elected to perform cross-validation testing, after processing all data with our normalization strategy, over all combinations of parameters consisting of (i) feature selection, (ii) model selection, and (iii) model hyperparameter tuning. Finally, our cross-validation testing allowed for refining the estimated final model with the best average performance remaining within small cross-center variability in performance, thus ensuring optimal generalizability.

We used a Sobol sequence for the randomized hyperparameter search to construct low discrepancy sets (33). The feature selection space was optimized for removing highly correlated features using a Spearman rho method with thresholds ranging from 0.8 to 1, as well as evaluating the optimal feature reduction method from among (a) *F*-test, (b) three relief-based algorithms from the open-source library ReBATE (34), or (c) a custom implementation of the Maximum Relevance – Minimum Redundancy strategy (35). Model selection space was optimized for identifying non-overfitting methods on our datasets (across folds) from logistic regression and XGboost, and then defining the optimal parametrization of such, considering (a) a metric of calibration with Nagekjerje's R index (average), (b) metrics of discrimination with the area under the curve (AUC), and (c) a metric of goodness of fit with the Brier's score; altogether, the agreement of these metrics is chosen as a proxy for generalizability due to their ability to capture (correlate with) the variance of the AUC (Supplementary Table 1). Moreover, we extended this observation by measuring the Youden's *J* statistics from both the discovery and validation models where the Clinical + DeepRadiomics model had a superior Youden score compared to Clinical + PD-L1 or Clinical + PyRadiomics, informing on the probability of a model to support an informed decision as opposed to a random guess, taking into account all model predictions (Supplementary Table 2).

The model with both best average and smallest variability in performance across folds was selected as the final model. The hyperparameter tuning space of the "selected model" was optimized for discovering the final 5 to 20 features best representing the complete information space (at each fold), inclusive of a set of 5 fixed features comprising lesion radius, ECOG score, age, smoking history (never/former/current), and first-line ICI (yes/no), with and without including PD-L1 status. In that manner, imaging features would only be identified in case of complementing clinically relevant features as per our primary objective.

We used the bootstrap 95% confidence interval of the model performance in the discovery cohort for establishing a judgment of generalizability to the validation cohort (36). We determined the success of a biomarker's generalizability test when its estimate of the AUC derived from the validation cohort fell within the confidence interval of estimate of the AUC derived from the discovery cohort.

Results

Cohort description

This study included a total of 910 eligible patients from four institutions. A first selection was made on availability of imaging within 3 months prior to initiation of ICI therapy, which resulted in the exclusion of 147 patients. Then, we excluded 121 patients for which a primary lesion could not be clearly delineated in the annotation process, to reach a final population of 642 patients (Supplementary Figure 1). A total of 512 patients from the discovery cohort had a median PFS of 5.5 months (95% CI [4.8–6.7]) and 130 patients from the validation cohort had a median PFS of 6.1 months (95% CI [5.1–7.5]) (p = 0.377). There were no statistical differences between the discovery and validation cohorts regarding the mean age, sex proportion, history of smoking, distribution of ECOG status, stage, or distribution of the PD-L1 group (all $p > 0.05$) (Table 1). However, a larger proportion of patients in the validation cohort received ICI as first-line therapy (72%) compared to the discovery cohort (39%) ($p < 0.001$) (Table 1). With respect to outcome, the proportion of patients in discovery and validation cohorts with PFS at 6 months were 51% and 46%, respectively (p = 0.377).

Benchmark of clinical outcome prediction with standard clinicopathological features

First, we sought to define the role of standard-of-care prognostic score using only clinical variables (age, ECOG status, smoking status, and line of treatment) alone or in combination with PD-L1 expression to establish a benchmark for clinical outcome prediction. The best clinical prognostic factor for PFS-6 was obtained with the combination of Clinical + PD-L1 expression

TABLE 1 Baseline characteristics of 642 patients segregated into the discovery and validation cohorts.

	Discovery cohort n=512	Validation cohort n=130	*p-value*
Age - median [IQR]	68.3 [62.3, 73.8]	67.1 [60.5, 72.8]	*0.232*
Sex - n (%)			
Male	258 (51)	69 (53)	*0.653*
Female	254 (49)	61 (47)	
Smoking history - n (%)			
Current or former	466 (92)	125 (97)	*0.067*
Never	46 (8)	5 (3)	
ECOG status - n (%)			
0	137 (27)	37 (29)	*0.514*
1	292 (59)	77 (59)	
>2	68 (14)	16 (12)	
Stage - n (%)			
III	52 (10)	11 (8)	*0.642*
IV	460 (90)	119 (92)	
Histology - n (%)			
Adenocarcinoma	404 (79)	89 (69)	*0.034*
Squamous	29 (6)	13 (10)	
Other	79 (15)	28 (21)	
PD-L1 status - n (%)			
<1%	97 (22)	36 (29.0)	*0.222*
1-49%	117 (27)	34 (27.4)	
≥50%	223 (51)	54 (43.5)	
Line of treatment - n (%)			
First line	197 (39)	94 (72)	*<0.001*
Second line or more	315 (61)	36 (28)	
Type of treatment - n (%)			
ICI alone	469 (91)	76 (59)	*<0.001*
Chemotherapy-ICI	43 (9)	54 (41)	
Progression-free survival - n (%)			
≥ 6 months	260 (51)	60 (46)	*0.377*
< 6 months	252 (49)	70 (54)	

Italic terms define the p-value.

with an AUC of 0.66 (95% CI [0.61–0.70]) in the discovery cohort and 0.62 (95% CI [0.53–0.72]) in the validation cohort (Supplementary Figures 2A, D). Clinical markers alone did not perform as well with an AUC of 0.64 (95% CI [0.59–0.69]) and 0.58 (95% CI [0.49–0.69]) in the discovery and validation cohort, respectively. Similarly, the AUC of the model with PD-L1 expression alone was 0.56 (95% CI [0.52–0.61]) in the discovery cohort and 0.59 (95% CI [0.48–0.70]) in the validation cohort (Supplementary Figures 2B, C, E, F).

Radiomics prediction without harmonization

Subsequently, we measured the predictive role of radiomics with no harmonization. We used PCoA to facilitate the projection and visualization of high-dimensional radiomic feature data. Before data harmonization, using PyRadiomic features, we observed clustering by CT scan slice thickness, manufacturer, kernel, and academic centers (Figure 2A). PCoA obtained from DeepRadiomics revealed similar clustering effect across the different medical centers, CT vendors, reconstruction kernels, and slice thicknesses (Figure 3A). These clusters were expected based on the important difference in CT acquisition parameters from each center (Supplementary Table 3). Without CT harmonization, a model combining Clinical + PyRadiomics or DeepRadiomics features to predict PFS-6 featured had an AUC of 0.69 (95% CI [0.64–0.74]) and 0.69 (95% CI [0.64–0.74]), respectively, in the discovery cohort (Supplementary Figure 3A). Nevertheless, AUC in the validation cohorts did not generalize; the AUC was 0.57 for Clinical + PyRadiomics and 0.52 for Clinical + DeepRadiomics both outside their respected interval CI obtained in the discovery cohorts (Supplementary Figure 3B).

Next, we aimed to include PD-L1 to these models to improve generalizability. First, PD-L1 did not increase the AUC in the discovery cohorts for both radiomic signatures with AUC reaching 0.71 compared to 0.69 without PD-L1 (Supplementary Figure 3C). Second, the addition of PD-L1 did not support the generalizability of performances in the validation cohorts (Supplementary Figure 3D).

Image processing-based harmonization of radiomics features

Following the normalization of raw CT scans data detailed in the "CT scan normalization" section, and processing first for PyRadiomics, we obtained a relatively homogeneous population for the clusters of slice thickness, manufacturer, and site (Figure 2B). Despite improvement in kernel distribution on the PCoA, visual clustering was still observed.

Next, the same normalization method alongside DeepRadiomics revealed a broad homogenization across the four parameters of interest including kernel (Figure 3B). Of note, we observed, after normalization, two populations across all PCoA. After further investigation, we confirmed that these populations were a result of the patch size of 48 × 48 pixels used during pretraining of the VGG16 backbone introduced in the "DeepRadiomics features extraction" section (Supplementary Figure 4A). To assess if the radius clusters were confounders for PFS-6 months, we represented a PCoA for DeepRadiomics features vs. PFS-6 months (Supplementary Figure 4B). We observed that

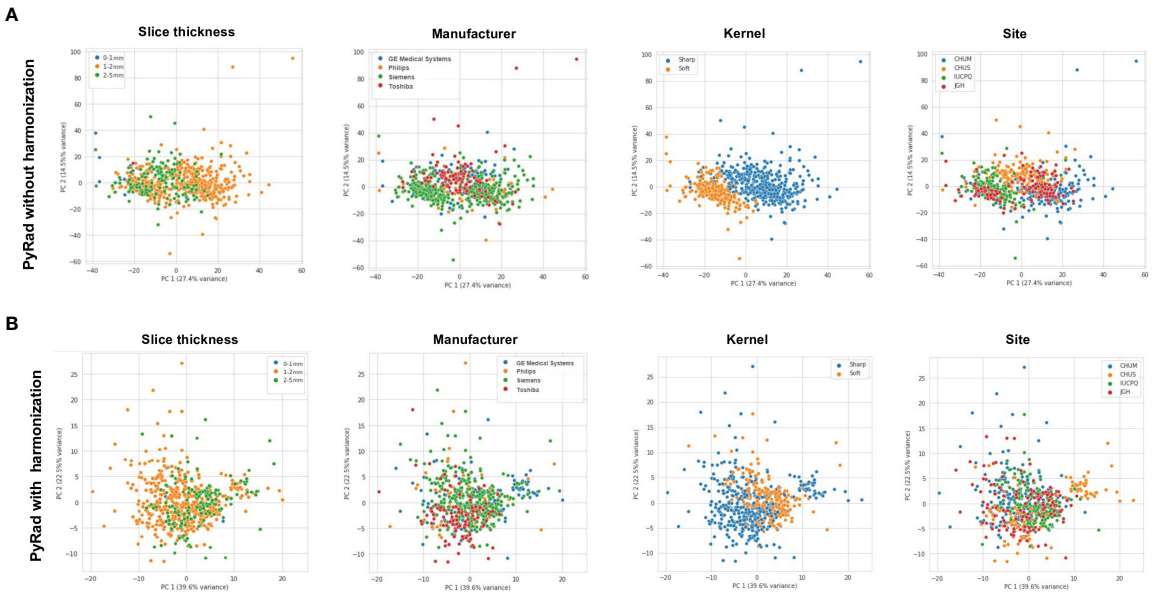


FIGURE 2
(A) Principal component analysis (PCoA) of PyRadiomics representation of various CT-scan acquisition parameters (Slide thickness, Manufacturer, Kernel, and Site) prior to image harmonization. **(B)** PCoA of PyRadiomics representation of various CT-scan acquisition parameters (Slide thickness, Manufacturer, Kernel, and Site) after image harmonization. PyRad, PyRadiomics.

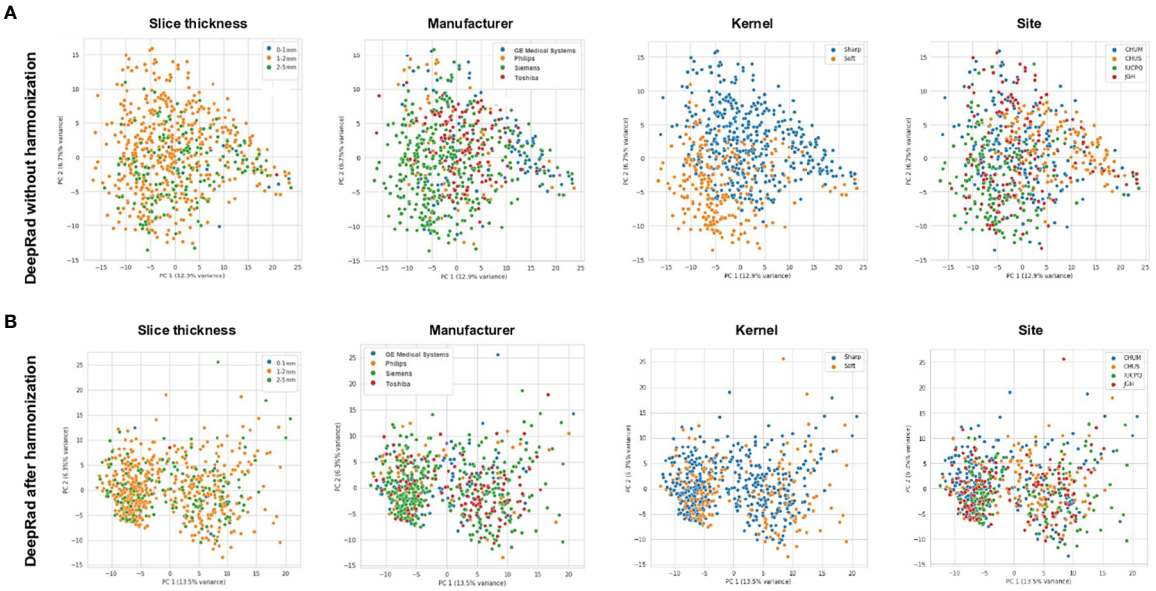


FIGURE 3
(A) Principal component analysis (PCoA) of DeepRadiomics representation of various CT-scan acquisition parameters (Slide thickness, Manufacturer, Kernel, and Site) prior to image harmonization. **(B)** PCoA of DeepRadiomics representation of various CT-scan acquisition parameters (Slide thickness, Manufacturer, Kernel, and Site) after image harmonization. DeepRad, DeepRadiomics.

there was no association of radius cluster with PFS-6 months. Our results showed that, after normalization, we were able to mitigate for variation across medical centers, CT vendors, acquisition variability, and reconstruction kernels.

Performance of AI-based radiomics signature designed for generalizability

Having laid out the foundation to construct a radiomic signature across centers, we sought to implement a global hyperparameter search framework (the "GHPS") to determine the optimal combination of machine learning and imaging features to establish a final radiomics biomarker to predict PFS-6.

First, using this construct, the combination of Clinical + PyRadiomics depicted an AUC of 0.66 (95% CI [0.61–0.70]) and 0.59 (95% CI [0.49–0.68]) in the discovery and validation cohorts, respectively, failing to meet the validation criteria for reproducibility (remaining within the discovery CI), *de facto* failing the generalizability objective (Figures 4A, B).

Second, the combination of Clinical + DeepRadiomics features reached an AUC of 0.67 (95% CI [0.63–0.73]) and 0.63 (95% CI [0.53–0.73]) in the discovery and validation cohort, respectively (Figures 4C, D). The validation cohort AUC of 0.63 also fell within its 95% CI estimate in the discovery cohort, meeting our generalizable objective. Interestingly, these results were comparable to Clinical + PD-L1 currently used in routine oncology practice. Moreover, both models also depicted a lower bound on the 95% CI of these models that was greater than 0.50 in the discovery cohort, confirming the predictive value of these models. Third, using both models, the addition of PD-L1 did not increase the performances (similar AUC) or the generalizability (Supplementary Figures 5A–D).

Finally, to provide valuable insights for future research to be leveraged as part of prior information for statistical study design, we conducted an exploratory assessment of non-inferiority using permutation analyses. Indeed, it is important to note that our study was not specifically designed or adequately powered for standard non-inferiority testing. Nonetheless, our findings indicate that the AUC of the Clinical + DeepRadiomics model was not statistically lower than the AUC of the Clinical + PD-L1 model (mean difference across permutations: 0.00035; *p*-value: 0.617).

Discussion

Radiomics represent a promising non-invasive biomarker for patients amenable to ICI; however, generalizability especially in various centers represent the major limitation (10, 19). In this large study of advanced NSCLC treated with ICI across four institutions, we demonstrated that a risk prediction model that combined Clinical + DeepRadiomics was generalizable and was non-inferior to the Clinical + PD-L1 model currently used by oncologists to predict PFS-6 months.

Importantly, our results showed that, after generalizability, DeepRadiomics methods had a better performance than the PyRadiomics pipeline. This could be explained by our proposed combination of traditional harmonization techniques, which, together with a generalization-optimizing AI framework, overcomes these limitations of previous models that did not generalize and enables clinical utility. Our AI framework involves

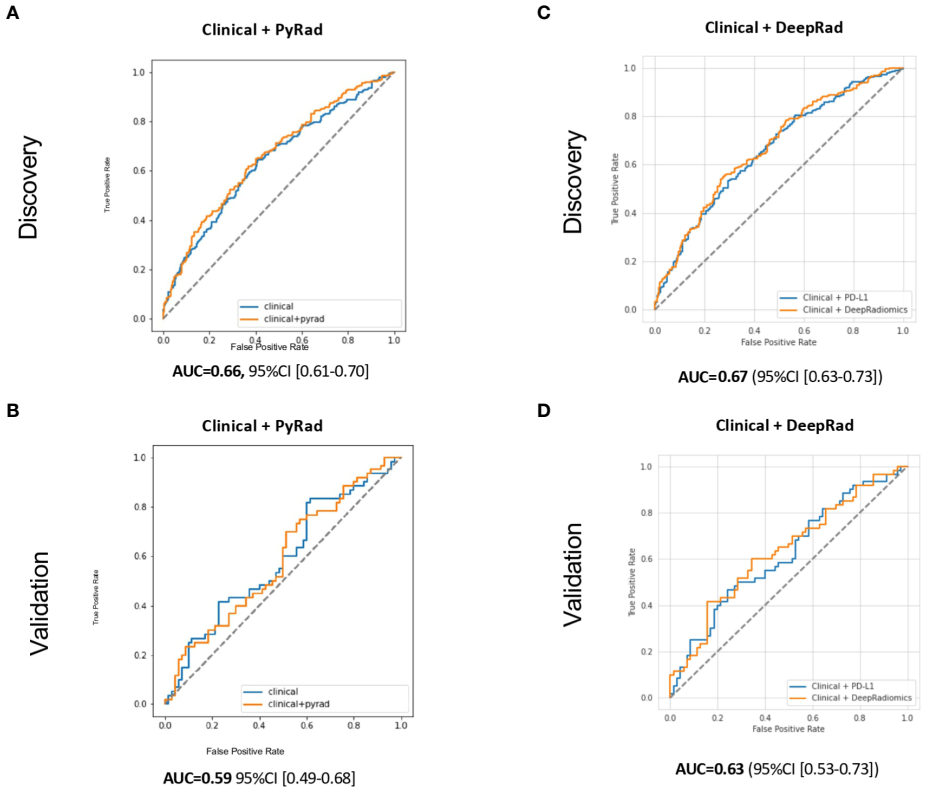


FIGURE 4
Receiver operating characteristic (ROC) curves for PFS-6 months prediction models with **(A)** Clinical (age, ECOG status, smoking status, and line of treatment) + PyRadiomics after harmonization, **(B)** Clinical + DeepRadiomics after harmonization in the discovery cohorts, **(C)** Clinical + PyRadiomics after harmonization, and **(D)** Clinical+ DeepRadiomics after harmonization in the validation cohorts. PyRad, PyRadiomics; DeepRad, DeepRadiomics.

two automated steps: the discovery of relevant imaging features using DeepRadiomics that delivers informative, reproducible, and stable compressed representation of an imaging data space, and a global hyperparameter search that iteratively loops over all combination of the three methods used in our modeling process: (a) feature selection that chooses the best algorithm that identifies features to include in the model, maintaining the most "informative" features, and removing noisy "non-informative," irrelevant and redundant features; (b) model selection that determines which machine learning estimator to use; and (c) hyperparameter tuning, which defines the optimum hyperparameter values to use for each estimator. While computationally intensive, this global search allows for the data-driven exploration of the somewhat unpredictable interplay between models and features (23). Consequently, we avoided radiomics features that were independently selected from other factors and/or not solely derived from the training portion of the data (in the machine learning training–validation–test sense), typically subject to an often-overlooked look-ahead bias and loss of future generalizability, addressing the problem of inappropriately applying cross-validation methods to feature selection (37).

However, the addition to our signature combining DeepRadiomics + Clinical was not improved by the implementation of a third parameter such as PD-L1. This could be explained by the limited discovery dataset available in this trial, which would otherwise be necessary to increase model parameters with sufficient examples for machine learning. Altogether, we obtained strong AUC in the discovery cohort independent of PD-L1 expression; however, there was no evidence of generalizability in the validation cohorts. This observation confirms that in the absence of homogeneous PCoA, radiomics validation is limited.

Furthermore, we acknowledge limitations in this study. First, although the CT scans were obtained from four institutions, only two physicians (a radiation oncologist and a radiologist) performed the image segmentations, reducing inter-observer variabilities. Second, the use of PyRadiomics is limited by the radiomic features being extracted from segmented ROI that required at least some degree of direct planimetry (and therefore additional physician time), subject to inter-annotator variability (38). This limitation was not present for the DeepRadiomics method we proposed, which does not require a segmentation input. Third, the validation cohort baseline characteristics had more patients treated with first line, which could decrease the performance of our model. Indeed, prior chemotherapy could impact the image features. Also, combination treatment such as chemotherapy with immunotherapy could modify the reproducibility of our model. Nevertheless, the DeepRadiomics method and the AI generalizability framework were able to mitigate this challenge.

Lastly, with a primary objective to assess model generalizability, we eventually lacked the power to undertake non-inferiority testing. Our current results would indeed motivate further investigation in that direction, on another larger cohort of patients.

In conclusion, this radiomics generalizability study was able to demonstrate that a DeepRadiomics signature with harmonization developed in a discovery cohort from various centers could overcome the negative impact of variable CT acquisition parameters and then could be validated in an independent cohort. This DeepRadiomics harmonization signature warrants further improvement and validation in external cohorts of patients with NSCLC treated with ICI and opens a new non-invasive biomarker strategy.

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 human participants were reviewed and approved by MP-02-2019-8091. The patients/participants provided their written informed consent to participate in this study.

Author contributions

MT annotated CT-scans, collected clinical data, and wrote and reviewed the manuscript. Imagia team (KP, CL-K, FD, LJ, FR, KK, DH, and FIC) completed statistical analysis, normalization, and radiomics features extraction. SK and JP collected clinical data. FrC annotated CT-scans. VM, DV JM, FIC, AG, WB, AE, MO, NB, TM, DH, PJ, and KP reviewed the manuscript. BR coordinated the project and reviewed the manuscript.

Funding

This work was supported by the Oncopole, the MEDTEQ Consortium, the TransMedTech Institute, and the Cancer Research Society, as part of the joint Onco-Tech Competition (BR, PJ, and Imagia), Marathon of Hope Montreal Cancer Consortium (BR), Institut du Cancer de Montréal NSCLC Biobank (BR), Terry Fox Clinician Scientist award (BR), Nuovo-Soldati Foundation for Cancer Research (MT), Bourse Maurice-Tubiana (MT), and Fonds de Recherche du Québec FRQS (BR and PJ). VM holds a salary support award from the Fonds de recherche du Québec – Santé (FRQS: Quebec Foundation for Health Research). This project was supported by the FRQS, a start-up fund from the Quebec Heart & Lung Institute Research Center, along with the Foundation grant from the Quebec Heart & Lung Institute Research Center.

Acknowledgments

We thank the Quebec Respiratory Health Research Network-IUCPQ site Biobank for providing clinical data and tissue from the IUCPQ cohort.

Conflict of interest

Author BR reports research funding with the company Imagia. Authors KP, CL-K, FD, LJ, FR, KK, DH, and FIC were Imagia employees at the time of the study.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.1196414/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Flow chart diagram of exclusion and final studied cohorts.

SUPPLEMENTARY FIGURE 2
Receiver operating characteristic (ROC) curves for the prediction models without radiomics and only with **(A)** clinical (age, ECOG status, smoking status, and line of treatment) + immunohistochemistry PD-L1 tumor expression, **(B)** representing clinical alone and **(C)** immunohistochemistry PD-L1 tumor expression in the discovery and **(D-F)** validation cohorts.

SUPPLEMENTARY FIGURE 3
(A) Receiver operating characteristic (ROC) curves for the prediction models with, clinical (age, ECOG status, smoking status, and line of treatment) + PyRadiomics before harmonization or clinical + DeepRadiomics before harmonization in the discovery cohorts **(B)** ROC curves of clinical + PyRadiomics before harmonization and clinical + DeepRadiomics before harmonization in the validation cohorts. **(C,D)** similar ROC curves but with the addiction of PD-L1 for the discovery and validation cohorts respectively.

SUPPLEMENTARY FIGURE 4
(A) Principal component analysis (PCoA) of DeepRadiomics features after normalization depicted by the VGG16 backbone network input of 24 pixels **(B)** PCoA of DeepRadiomics features stratified by outcome PFS-6 months.

SUPPLEMENTARY FIGURE 5
Receiver operating characteristic (ROC) curves for PFS-6 months prediction models with **(A)** Clinical (age, ECOG status, smoking status, and line of treatment) + PyRadiomics + PD-L1 after harmonization in the discovery cohort **(B)** Clinical + PyRadiomics + PD-L1 after harmonization in the validation cohort and **(C)** clinical + DeepRadiomics + PD-L1 after harmonization in the discovery

cohort **(D)** clinical + DeepRadiomics + PD-L1 after harmonization in the validation cohort. PyRad, PyRadiomics; DeepRad, DeepRadiomics.

SUPPLEMENTARY TABLE 1
– Generalizability of the performance of the prediction models using Nagelkerke's R, Bier's score and area under the curve (AUC) as measures of AI-model calibration.

SUPPLEMENTARY TABLE 2
– Determination of the Sensitivity and Specificity of the Model in discovery and validation cohorts using the Yonden's index.

SUPPLEMENTARY TABLE 3
- Baseline characteristics of computed tomography scan characteristics in the discovery and validation cohorts.

References

1. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* (2016) 278(2):563–77. doi: 10.1148/radiol.2015151169

2. Gandhi L, Rodríguez-Abreu D, Gadgeel S, Esteban E, Felip E, De Angelis F, et al. Pembrolizumab plus chemotherapy in metastatic non-Small-Cell lung cancer. *N Engl J Med* (2018) 378(22):2078–92. doi: 10.1056/NEJMoa1801005

3. Hellmann MD, Paz-Ares L, Bernabe Caro R, Zurawski B, Kim SW, Carcereny Costa E, et al. Nivolumab plus ipilimumab in advanced non-Small-Cell lung cancer. *N Engl J Med* (2019) 381(21):2020–31. doi: 10.1056/NEJMoa1910231

4. Kim HS, Cha H, Kim J, Park WY, Choi YL, Sun JM, et al. Genomic scoring to determine clinical benefit of immunotherapy by targeted sequencing. *Eur J Cancer* (2019) 120:65–74. doi: 10.1016/j.ejca.2019.08.001

5. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, et al. Five-year outcomes with pembrolizumab versus chemotherapy for metastatic non-Small-Cell lung cancer with PD-L1 tumor proportion score ≥ 50. *J Clin Oncol* (2021) 39 (21):2339–49. doi: 10.1200/JCO.21.00174

6. Gadgeel S, Rodríguez-Abreu D, Speranza G, Esteban E, Felip E, Dómine M, et al. Updated analysis from KEYNOTE-189: pembrolizumab or placebo plus pemetrexed and platinum for previously untreated metastatic nonsquamous non-Small-Cell lung cancer. *J Clin Oncol* (2020) 38(14):1505–17. doi: 10.1200/JCO.19.03136

7. Rodríguez-Abreu D, Powell SF, Hochmair MJ, Gadgeel S, Esteban E, Felip E, et al. Pemetrexed plus platinum with or without pembrolizumab in patients with previously untreated metastatic nonsquamous NSCLC: protocol-specified final analysis from KEYNOTE-189. *Ann Oncol* (2021) 32(7):881–95. doi: 10.1016/j.annonc.2021.04.008

8. Rizvi H, Sanchez-Vega F, La K, Chatila W, Jonsson P, Halpenny D, et al. Molecular determinants of response to anti-programmed cell death (PD)-1 and anti-programmed death-ligand 1 (PD-L1) blockade in patients with non-Small-Cell lung cancer profiled with targeted next-generation sequencing. *J Clin Oncol* (2018) 36 (7):633–41. doi: 10.1200/JCO.2017.75.3384

9. Limkin EJ, Sun R, Dercle L, Zacharaki EI, Robert C, Reuzé S, et al. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. *Ann Oncol* (2017) 28(6):1191–206. doi: 10.1093/annonc/mdx034

10. Sun R, Limkin EJ, Vakalopoulou M, Dercle L, Champiat S, Han SR, et al. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. *Lancet Oncol* (2018) 19(9):1180–91. doi: 10.1016/S1470-2045(18)30413-3

11. Castello A, Castellani M, Florimonte L, Urso L, Mansi L, Lopci E. The role of radiomics in the era of immune checkpoint inhibitors: a new protagonist in the jungle of response criteria. *J Clin Med* (2022) 11(6):1740. doi: 10.3390/jcm11061740

12. Sun R, Henry T, Laville A, Carré A, Hamaoui A, Bockel S, et al. Imaging approaches and radiomics: toward a new era of ultraprecision radioimmunotherapy? *J Immunother Cancer* (2022) 10(7):e004848. doi: 10.1136/jitc-2022-004848

13. Vanguri RS, Luo J, Aukerman AT, Egger JV, Fong CJ, Horvat N, et al. Multimodal integration of radiology, pathology and genomics for prediction of response to PD-(L)1 blockade in patients with non-small cell lung cancer. *Nat Cancer* (2022) 13:1–14. doi: 10.1038/s43018-022-00416-8

14. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* (2014) 5:4006. doi: 10.1038/ncomms5006

15. Ibrahim A, Primakov S, Beuque M, Woodruff HC, Halilaj I, Wu G, et al. Radiomics for precision medicine: current challenges, future prospects, and the proposal of a new framework. *Methods* (2021) 188:20–9. doi: 10.1016/j.ymeth.2020.05.022

16. He L, Huang Y, Ma Z, Liang C, Liang C, Liu Z. Effects of contrast-enhancement, reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule. *Sci Rep* (2016) 6:34921. doi: 10.1038/srep34921

17. Berenguer R, Pastor-Juan MDR, Canales-Vázquez J, Castro-García M, Villas MV, Mansilla Legorburo F, et al. Radiomics of CT features may be nonreproducible and redundant: influence of CT acquisition parameters. *Radiology* (2018) 288(2):407–15. doi: 10.1148/radiol.2018172361

18. Da-ano R, Masson I, Lucia F, Doré M, Robin P, Alfieri J, et al. Performance comparison of modified ComBat for harmonization of radiomic features for multicenter studies. *Sci Rep* (2020) 10(1):10248. doi: 10.1038/s41598-020-66110-w

19. Da-Ano R, Visvikis D, Hatt M. Harmonization strategies for multicenter radiomics investigations. *Phys Med Biol* (2020) 65(24):24TR02. doi: 10.1088/1361-6560/aba798

20. Horng H, Singh A, Yousefi B, Cohen EA, Haghighi B, Katz S, et al. Generalized ComBat harmonization methods for radiomic features with multi-modal distributions and multiple batch effects. *Sci Rep* (2022) 12(1):4493. doi: 10.1038/s41598-022-08412-9

21. Orlhac F, Eertink JJ, Cottereau AS, Zijlstra JM, Thieblemont C, Meignan M, et al. A guide to ComBat harmonization of imaging biomarkers in multicenter studies. *J Nucl Med* (2022) 63(2):172–9. doi: 10.2967/jnumed.121.262464

22. 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

23. Johnson MKK. Feature Engineering and Selection. *A practical approach for predictive models* (2022). Available at: https://bookdown.org/max/FES/.

24. Wang J, Guo H. Automatic approach for lung segmentation with juxta-pleural nodules from thoracic CT based on contour tracing and correction. *Comput Math Methods Med* (2016) 2016:2962047. doi: 10.1155/2016/2962047

25. Kuruvilla J, Gunavathi DK. Detection of lung cancer using morphological operations. *Inter J Sci Engineer Res* (2013) 4(8):4.

26. Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative. *Radiology* (2020) 295(2):328–38. doi: 10.1148/radiol.2020191145

27. Li Z, Wang Y, Yu J, Guo Y, Cao W. Deep learning based radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma. *Sci Rep* (2017) 7(1):5467. doi: 10.1038/s41598-017-05848-2

28. Zhang X, Zhang Y, Zhang G, Qiu X, Tan W, Yin X, et al. Deep learning with radiomics for disease diagnosis and treatment: challenges and potential. *Front Oncol* (2022) 12:12. doi: 10.3389/fonc.2022.773840

29. Chen T, Kornblith S, Norouzi M, Hinton G. *A simple framework for contrastive learning of visual representations [Internet]. arXiv* (2020). Available at: http://arxiv.org/abs/2002.05709.

30. Chen T, Kornblith S, Swersky K, Norouzi M, Hinton G. *Big self-supervised models are strong semi-supervised learners [Internet]. arXiv* (2020). Available at: http://arxiv.org/abs/2006.10029.

31. Armato SG, Roberts RY, McNitt-Gray MF, Meyer CR, Reeves AP, McLennan G, et al. The lung image database consortium (LIDC): ensuring the integrity of expert-defined "truth". *Acad Radiol* (2007) 14(12):1455–63. doi: 10.1016/j.acra.2007.08.006

32. Pudjihartono N, Fadason T, Kempa-Liehr AW, O'Sullivan JM. A review of feature selection methods for machine learning-based disease risk prediction. *Front Bioinf* (2022) 2:927312. doi: 10.3389/fbinf.2022.927312

33. Bergstra J, Bengio Y. Random search for hyper-parameter optimization. *J Machine Learning Res* (2012) 13:281–305.

34. *Home - ReBATE* (2022). Available at: https://epistasislab.github.io/ReBATE/.

35. Zhao Z, Anand R, Wang M. *Maximum relevance and minimum redundancy feature selection methods for a marketing machine learning platform [Internet]. arXiv* (2019). Available at: http://arxiv.org/abs/1908.05376.

36. Banjanovic ES, Osborne JW. *Confidence intervals for effect sizes: applying bootstrap resampling* (2022). Available at: https://scholarworks.umass.edu/pare/vol21/iss1/5/.

37. Hastie T, Tibshirani R, Friedman J. *Elements of statistical learning: data mining, inference, and prediction. 2nd edition. 2nd ed.* Springer. (2009). Available at: https://hastie.su.domains/ElemStatLearn/.

38. Primakov SP, Ibrahim A, van Timmeren JE, Wu G, Keek SA, Beuque M, et al. Automated detection and segmentation of non-small cell lung cancer computed tomography images. *Nat Commun* (2022) 13(1):3423. doi: 10.1038/s41467-022-30841-3

frontiers | Frontiers in Oncology

TYPE Review
PUBLISHED 21 July 2023
DOI 10.3389/fonc.2023.1209299

Check for updates

OPEN ACCESS

EDITED BY
Donghong Zhang,
Emory University, United States

REVIEWED BY
Orazio Fortunato,
National Cancer Institute Foundation (IRCCS), Italy
Erik Wiemer,
Erasmus Medical Center, Netherlands

*CORRESPONDENCE
Justyna Raczkowska
justyna.raczkowska@umb.edu.pl

RECEIVED 20 April 2023
ACCEPTED 28 June 2023
PUBLISHED 21 July 2023

CITATION
Raczkowska J, Bielska A, Krętowski A and Niemira M (2023) Extracellular circulating miRNAs as potential non-invasive biomarkers in non-small cell lung cancer patients.
Front. Oncol. 13:1209299.
doi: 10.3389/fonc.2023.1209299

COPYRIGHT
© 2023 Raczkowska, Bielska, Krętowski and Niemira. 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.

Extracellular circulating miRNAs as potential non-invasive biomarkers in non-small cell lung cancer patients

Justyna Raczkowska[1*], Agnieszka Bielska[1], Adam Krętowski[1,2] and Magdalena Niemira[1]

[1]Clinical Research Centre, Medical University of Białystok, Białystok, Poland, [2]Department of Endocrinology, Diabetology and Internal Medicine, Medical University of Białystok, Białystok, Poland

Non-small cell lung cancer (NSCLC) comprises 85% of all lung cancers and is a malignant condition resistant to advanced-stage treatment. Despite the advancement in detection and treatment techniques, the disease is taking a deadly toll worldwide, being the leading cause of cancer death every year. Current diagnostic methods do not ensure the detection of the disease at an early stage, nor can they predict the risk of its development. There is an urgent need to identify biomarkers that can help predict an individual's risk of developing NSCLC, distinguish NSCLC subtype, allow monitor disease and treatment progression which can improve patient survival. Micro RNAs (miRNAs) represent the class of small and non-coding RNAs involved in gene expression regulation, influencing many biological processes such as proliferation, differentiation, and carcinogenesis. Research reports significant differences in miRNA profiles between healthy and neoplastic tissues in NSCLC. Its abundant presence in biofluids, such as serum, blood, urine, and saliva, makes them easily detectable and does not require invasive collection techniques. Many studies support miRNAs' importance in detecting, predicting, and prognosis of NSCLC, indicating their utility as a promising biomarker. In this work, we reviewed up-to-date research focusing on biofluid miRNAs' role as a diagnostic tool in NSCLC cases. We also discussed the limitations of applying miRNAs as biomarkers and highlighted future areas of interest.

KEYWORDS

NSCLC, non-small cell lung cancer, miRNA, microRNA, biomarker, biofluids, serum, plasma

1 Introduction

Lung cancer remains one of the most frequent and deadly cancers worldwide. The International Agency for Research on Cancer estimated approximately 2.2 million of new cases and 1.8 million deaths in 2020. Among males, lung cancer appears to be the first diagnosed cancer, whereas it ranks third for females. Lung cancer continues to be the leading cause of cancer death being overtaken only by breast cancer as the most commonly

diagnosed one (1). Non-small cell lung cancer (NSCLC) is the first of two histologic lung cancer types, affecting approximately 85% of all cases. In this type, it is possible to distinguish three main subtypes: adenocarcinoma (ADC), relating to 40% of cases; squamous-cell carcinoma (SCC), which affects 25% of cases; and large-cell carcinoma (LCC), in 10% of cases (Figure 1) (2). The American Joint Committee on Cancer (AJCC) recommends using the TNM classification of lung cancer. This staging is established on the characteristics of the primary tumor (T), the degree of lymph node involvement (N) and metastasis status (M). Consequently, each patient is assigned a general stage (0, I, II, III, IV). The the most advanced stage is IV, while stages 0, I and II are considered to be early stages (3, 4).

The overall 5-year survival rate is 15% in NSCLC patients, which results from complex histology and genetics within subtypes of cancer, non-specific symptoms, and late diagnosis (5, 6). Unfortunately, lung cancer is often detected at late stages, when invasion and distant metastasis has already occurred. Nowadays, one of the initiating steps to diagnose lung cancer is a chest X-ray, which can expose a mass typical for cancer tumours. On that basis, more precise tests, such as bronchoscopy and liquid biopsy, can be performed to complete the diagnosis and establish the tumour subtype (7, 8). Currently, treatment is focused on surgery, chemotherapy, and radiotherapy, which is still insufficient to reduce NSCLC mortality (8). Molecular therapies are currently based mainly on tyrosine kinase inhibitors. Nevertheless, positive results can only be expected in NSCLC patients with *EGFR, BRAF* and *MET* mutations, as well as with rearrangement of the *ALK, ROS1* or *NTRK* genes. Likewise, immune checkpoint inhibitors (ICIs) provide a new and promising approach, but treatment is effective only in 20-30% of NSCLC cases. Despite the immunotherapy and recent development of targeted therapies, the World Health Organisation guidelines emphasized the importance of better subclassification of lung cancer (9). Accurate and detailed classification of subtype of NSCLC is often a challenge, due to limited diagnostic material or the need for additional techniques such as immunohistochemistry (10). New, accurate biomarkers could help distinguish the subtype of NSCLC at an early stage which would enable the implementation of early treatment. Moreover, a very important task is to identify biomarkers to predict treatment effects and response to therapies, which will allow the selection of personalised therapy (11).

MicroRNAs (miRNAs) appear to be one of the most abundant RNA in the cells, making them promising molecules for detecting NSCLC and other cancers. MiRNAs are a group of short (about 21-23 nucleotide-long) and single-stranded particles that function through negative post-transcriptional regulation of gene expression. MiRNAs genes are primarily located in the inter-gene regions. They can exist as independent transcription units between protein-coding sequences. About 25% of the miRNA coding sequences in humans are located in introns. MiRNA genes are also found in exons and can form polycistronic clusters having identical regulatory sequences (12–14). The first stage of miRNAs' canonical formation occurs in the cell nucleus, where the miRNAs are transcribed into pri-miRNAs (Figure 2). The primary transcripts, which can be longer than 1000 nucleotides, include an approximately 70-nucleotide double-stranded fragment in the form of a "hairpin". This fragment is recognised and transformed by the Drosha-Pasha complex into pre-miRNAs, having a phosphate group at the 5' end and two unpaired nucleotides at the 3'end. The obtained pre-miRNAs are then transported to the cytoplasm by the nuclear transport protein - exportin 5. Subsequently, the enzyme Dicer converts pre-miRNA into double-stranded duplexes (dsRNAs), which are loaded into the RNA Induced Silencing Complex (RISC). Only one miRNA strand is retained and bound to the RISC component - protein argonaut (AGO2) as mature miRNA, while the second is usually degraded. The target mRNA sequences are recognised by mature miRNAs mainly in the

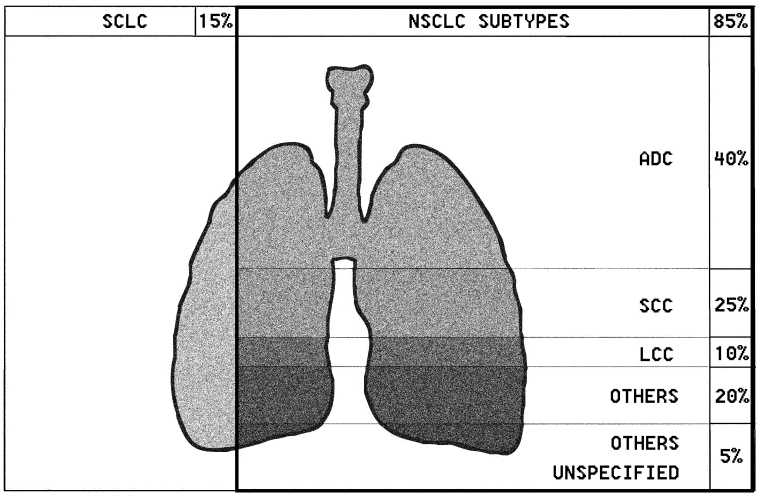


FIGURE 1
Lung cancer (lung carcinoma) is divided into two histologic types: non-small cell lung carcinoma (NSCLC) and small-cell lung carcinoma (SCLC). There are three main subtypes of NSCLC, including the following: adenocarcinoma (ADC), squamous-cell carcinoma (SCC), and large-cell carcinoma (LCC). Other subtypes include pulmonary enteric adenocarcinoma.

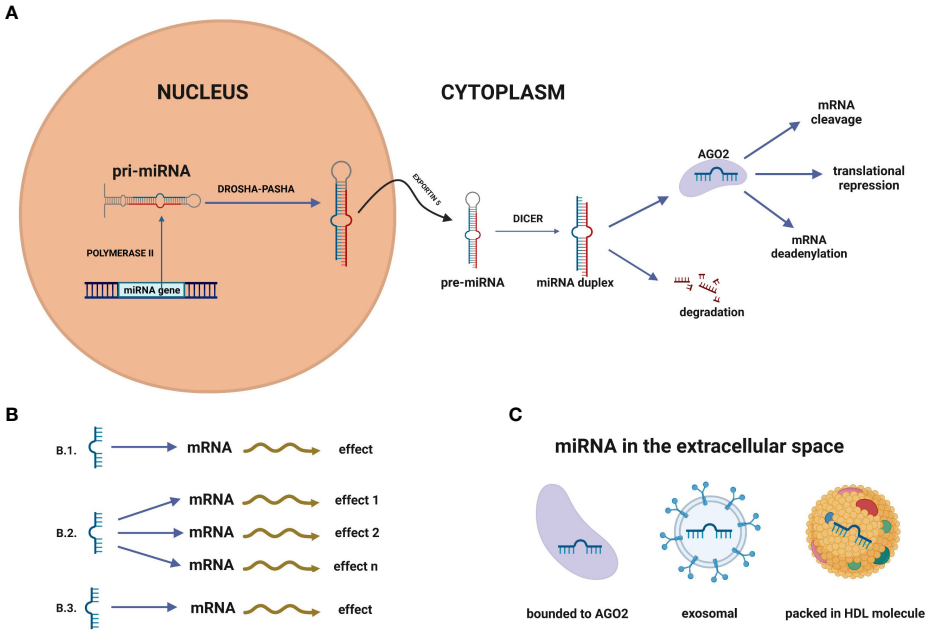


FIGURE 2
(A) MicroRNA biogenesis. MiRNA genes are transcribed in the nucleus by polymerase II to primary RNA (pri-miRNA). Pri-miRNA is transformed by the Drosha-Pasha complex into pre-miRNAs. In the next step, pre-miRNAs are transported to the cytoplasm by the nuclear transport protein – exportin 5. Subsequently, the enzyme Dicer converts pre-miRNA into double-stranded duplexes, which are loaded into the RNA Induced Silencing Complex (RISC). Only one miRNA strand is retained and bound to the RISC component - protein argonaute (AGO2) as mature miRNA, while the second is usually degraded. **(B)** MiRNA networks. **(B.1)** Single miRNA-mRNA interplay, **(B.2)** One miRNA can affect many miRNAs, **(B.3)** Single miRNA can impact different miRNAs to gain the biological effect. **(C)** In the extracellular space, miRNAs can be packed in vehicles (exosomes, microvesicles, apoptotic bodies) but also into high-density lipoprotein (HDL) molecules or bound to proteins, such as Ago2.

3' untranslated region (3' UTR), although recognition is sometimes found in the 5 'untranslated region (5'UTR). Consequently, these targeted mRNAs are silenced by cleavage of the miRNA, translational repression, mRNA deadenylation, or a combination of these processes (15–17).

It is observed that miRNAs are tissue-specific, and their concentration changes are seen in affected and healthy tissues (18–21). The multifunctionality and multidirectional activity of miRNA molecules underline their essential role in organism functioning. They play an important role in many processes, such as proliferation, differentiation, gene silencing, carcinogenesis, apoptosis, and cell survival (22–25). Furthermore, it has been found that specific miRNAs are associated with metastasis, inter alia, through the regulation of epithelial-mesenchymal transition (26).

A solid tumour releases miRNAs into the body fluids such as blood, serum, plasma, amniotic fluid, saliva, urine and peritoneal fluid (22, 23, 27). In the extracellular space, miRNAs can be packed in vehicles (exosomes, microvesicles, apoptotic bodies) but also into high-density lipoprotein (HDL) molecules (24) or bound to proteins, such as Ago2 (Figure 2B) (28, 29). It has been found that circulating miRNAs are protected from endogenous RNase activity (23), and have a low susceptibility to extreme temperatures and pH changes (30). All these features make them stable even after repeated freezing and thawing cycles of diagnostic material. Collecting biofluids from patients by liquid biopsy is minimally invasive and easy; thus, diagnostics could be more accessible and less expensive (31). MiRNAs can act as oncogenes (oncomiRs) by negatively regulating tumor suppressor genes involved in apoptosis and differentiation. Some miRNAs act as tumor suppressors, so under-expression of the gene leads to cancer development (32). MiRNA can act as an oncomiR and suppressor in specific conditions. For example, miR-196 was examined to be up-regulated in the oesophagus, glioblastoma and colon cancer, but other studies have verified it to act as a tumour suppressor (33).

MiRNAs can act as oncogenes (oncomiRs) by negatively regulating tumor suppressor genes involved in apoptosis and differentiation. Some miRNAs act as tumor suppressors, so under-expression of the gene leads to cancer development (32).

Many bioinformatic tools for analysis require a good understanding of biological processes. The increasing use of molecular techniques has allowed researchers to identify miRNAs and expand the knowledge about the targets of these molecules. The most commonly used process for expression analysis of miRNA is quantitative reverse transcriptase polymerase chain reaction (qPCR/RT-qPCR) which is a gold standard. Other methods widely used include northern blotting, next-generation RNA sequencing, and microarrays (11). One of the most advanced methods of detecting miRNAs is the nCounter platform (Nanostring Technology). This method offers direct RNA detection without an amplification process, so it is based on direct multiplexed measurement of gene expression (34). The freely available database miRBase (http://www.mirbase.org) is constantly enriched in information about miRNAs. It provides

knowledge about mature and hairpin miRNA sequences, their location and biogenesis precursors. There are also links to literature references, annotations and databases that contain predicted and experimentally validated targets of miRNAs (35). Discovering the miRNA sequences provides bioinformatics tools to predict its targets. It is possible due to the recognition of mRNA matches to miRNA seed regions which are 6-8 pairs long. Statistics methods have revealed that a single miRNA could have nearly hundred targets, which may result in a sum of effects determining a typical phenotype (36). In addition, miRNA networks show that one miRNA can impact different miRNAs to act synchronically (Figure 2C) (37). Clusters of two or more miRNAs are commonly related to several cell functions. For example, the miR-17-92 cluster is connected with the development of the lungs, immune system, and cancer development (38).

Many studies investigating miRNA activity in cancer have focused on its potential as diagnostic biomarkers, predictors of prognosis and drug efficiency. It could also effectively classify cancer subtypes (39) and predict metastatic outcomes (40). Research evidence indicates the role of biofluid miRNAs in many cancers, such as ovarian cancer (41), prostate cancer (42), liver cancer (43), and NSCLC (30). Lin et al. (44), as early as 2010, published a review focusing on the significant role of miRNAs in lung cancer. It is well known that miR-21 is involved in many cancer processes, such as tumorigenesis, progression and metastasis (45, 46). Numerous studies demonstrated that miR-21 is up-regulated in hepatocellular carcinoma (47), gastric cancer (48), prostate cancer (49), and also in lung cancer (45). It was shown that this miRNA is highly expressed in NSCLC patients, and inhibition of miR-21 reduces the proliferation, migration and invasion of cancer cells (45). Yu et al. (50) indicated that miR-10a is up-regulated in NSCLC patients' tissue. Furthermore, authors showed that this miRNA might promote the progression of cancer cells by targeting *PTEN/AKT/ERK* signalling pathway. Interestingly, in the previous study by Markou et al. (51), plasma miR-10a was described as a promising biomarker for NSCLC. It has been proven that miRNA-200a-3p showed under-expression in the NSCLC tumour tissues and A549 cell lines. Insulin receptor substrate 2 (*IRS2*) was shown to be a direct target for this miRNA (52).

This review focuses on recent advances in the role of biofluid-circulating miRNAs in NSCLC. We organize and present here the latest reports, which could support the search for new diagnostic, prognostic and predictive biomarkers for NSCLC. This manuscript includes recently studied miRNAs from exosomes (with "exosomal" prefix in the text, with "ex" in tables), protein-bounded miRNAs (without the prefix in tables) and vehicle-derived ("sEVs") miRNAs.

The paragraphs of this manuscript are divided according to the potential application of the studied miRNAs. Notably, the results of many studies point to more than one possible use for miRNAs, so the results of one paper are present in more than one paragraph. However, in the tables, we decided to list the latest reports concerning miRNAs in chronological order: serum (Table 1), plasma (Table 2), and sputum (Table 3), to avoid repetition in the tables. The most promising miRNAs for the diagnosis of NSCLC are presented in Table 4.

We searched the Pubmed database with keywords: miRNA, NSCLC, biomarkers AND serum/plasma/sputum/saliva/urine to

TABLE 1 Latest reports concerning miRNAs in serum of NSCLC patients.

miRNA	Expression in Research Group vs Controls	Number of cases	Significant Findings	Methods	Application in NSCLC	Year of Publication	Reference
miR-30	down	108 NSCLC/ 108 controls	levels of this miRNA correlated with lymph node metastasis, tumour size, TNM stage, and differentiation degree	qRT-PCR	diagnosis, prognosis	2020	(53)
exosomal miR-1246	up	105 NSCLC/ 50 NMLD patients/50 controls	miR-1246 may help distinguish NSCLC patients from controls and NMRD-diseased patients	qRT-PCR	diagnosis, prognosis	2020	(54)
exosomal miR-620	down	235 NSCLC/ 231 controls	miR-620 was associated with alcohol consumption, distant metastasis and chemotherapeutic effect	TEM, qNano, and western blot; validation: qPCR	diagnosis, prediction	2020	(55)
miR-762	up	148 NSCLC/ 60 controls	high levels of miR-762 were associated with histological grade, clinical stage and lymph node metastasis	qRT-PCR	diagnosis, prognosis	2020	(56)
miR-519d	down	130 NSCLC/ 50 controls	miR-519d under-expression was associated with lymph node	qRT-PCR	diagnosis, prognosis	2020	(57)

(Continued)

TABLE 1 Continued

miRNA	Expression in Research Group vs Controls	Number of cases	Significant Findings	Methods	Application in NSCLC	Year of Publication	Reference
			metastases, clinical stage, and a poor prognosis				
ex-miR-5684, ex-miR-125b-5p	down	330 NSCLC/ 312 controls	levels of that miRNAs were associated with metastasis (P < 0.0001), chemotherapeutic effect (P = 0.007)	TEM, qNano, western blot, microarrays, qPCR	diagnosis, prediction	2020	(58)
miR-4687-3p	up	training cohort: 20 NSCLC/20 controls, validation cohort: 30 NSCLC/30 controls	this miRNA plays an important role in NSCLC development	microarray, qRT-PCR	diagnosis	2020	(59)
miR-145, miR-20a	down: miR-145, up: miR-20a	32 NSCLC/33 other lung cases/23 controls	miR-145 and miR-20a may be NSCLC biomarkers	qRT-PCR	diagnosis	2020	(60)
miR-942, miR-601	up	125 NSCLC/ 40 benign lung diseases/ 60 controls	these miRNAs performed better than CEA, CYFRA21-1, and the SCC antigen in the early detection of NSCLC	qRT-PCR	diagnosis, prognosis	2020	(61)
miR-15a-5p, miR-320a, miR-25-3p, miR-192-5p, miR-let-7d-5p, miR-let-7e-5p, miR-148a-3p, miR-92a-3p	down	75 NSCLC/40 controls	miR-320a, miR-25-3p, miR-192-5p, miR-let-7d-5p, miR-148a-3p, and miR-let-7e-5p were significantly correlated with gender of patients	small RNA seq., validation: qPCR	diagnosis	2020	(62)
miR-375, miR-10b-5p	down only in SCC	75 NSCLC/40 controls	miR-375 levels were significantly correlated with lymph node involvement, while both miRNAs with pleural effusion	small RNA seq., validation: qPCR	prognosis	2020	(63)
ex-miR-20b-5p, ex-miR-3187-5p	down	276 NSCLC/ 282 controls	integrating those miRNAs could distinguish the patients at an early stage from healthy controls	qNano, TEM, western blot, qRT-PCR, microarrays	early-stage diagnosis	2020	(64)
miR-17-92 gene cluster: miR-17, miR-18a, miR-19a, miR-19b-1, miR-20a and miR-92a	down: miR-17 (stage II, III), miR-18a (stage II), miR-19a/b-1 (stage II, III), miR-20a (stage I, II, III), miR-92a-1 (stage II, III); up in stage IV: miR-17, miR-18a, miR-20a, miR-92a-1	74 NSCLC/23 controls	miR-17-92 gene cluster might be helpful as a diagnostic marker in NSCLC	qRT-PCR	diagnosis	2020	(65)
miR–518b	up	118 NSCLC/ 60 controls	higher miR-518b expression was associated with larger tumour size, lymph node metastasis, advanced TNM stage, and with poor overall survival in NSCLC patients	qRT-PCR	diagnosis, prognosis,	2020	(66)
ex-miR-378	up	103 NSCLC/ 32 NMLD	miR-378 could be potentially used as an	qRT-PCR	diagnosis, prediction	2020	(67)

(Continued)

TABLE 1 Continued

miRNA	Expression in Research Group vs Controls	Number of cases	Significant Findings	Methods	Application in NSCLC	Year of Publication	Reference
		patients/60 controls	indicator of radiotherapeutic response in NSCLC				
miR-185	down	146 NSCLC/ 50 carcinoma in situ/25 NMLD/80 controls	serum miR-185 expression was progressively decreased in healthy controls, patients with NMLD, carcinoma *in situ* and NSCLC	qRT-PCR	early detection, prognosis	2020	(68)
miR-181a-5p, miR-1228-3p	down: miR-181a-5p, up: miR-1228-3p,	50 NSCLC/30 controls	miR-1228-3p and miR-181a-5p levels were significantly associated with OS	qRT-PCR	diagnosis, prognosis	2020	(69)
miR-3195,	down	75 NSCLC/40 controls	miR-3195 as an independent prognostic factor for OS	qRT-PCR	prognosis	2020	(70)
miR-1249-3p,	down	75 NSCLC/40 controls	miR-1249-3p might predict therapeutic response to chemotherapy	qRT-PCR	therapeutic response	2020	(70)
ex-miR-1269a, ex-miR-9-3p, ex-miR-205-5p and ex-miR-210-5p	down	147 NSCLC/ 139 controls	panel of four tested exosomal miRNAs may serve as noninvasive diagnostic biomarker for NSCLC	qRT-PCR	diagnosis	2020	(71)
miR-186	down	62 NSCLC/60 controls	miR-186 and IL-1β might serve as potential biomarkers for NSCLC	qRT-PCR	diagnosis	2020	(72)
miR–130a, miR–25 and miR–191	up	84 NSCLC/42 controls	these miRNAs may have an oncogenic role in the radiation–mediated metastasis of NSCLC	qRT-PCR	prediction	2020	(73)
ex-miR-21/miR-let-7a ratio	up	75 NSCLC/23 benign pulmonary nodules/18 pulmonary inflammation diseases/24 controls	this ratio can be used to distinguish NSCLC from benign pulmonary disease patients	qRT-PCR	diagnosis	2020	(74)
ex-let-7e	down	105 NSCLC/ 50 controls	higher levels of that miRNA had an anticarcinogenic properties against NSCLC	TEM, cell culture, qRT-PCR, western blot	prognosis	2020	(75)
miR-21-5p, miR-141-3p, miR-222-3p, and miR-486-5p, ex-miR-146a-5p, ex-miR-486-5p	up	48 NSCLC/32 benign lung lesion/48 controls	these miRNAs could further improve early diagnosis of NSCLC	TEM, Western blot, qRT-PCR	early diagnosis	2020	(76)
miR-184, miR-191	down: miR-184, up: miR-191	100 NSCLC/ 59 pneumonia/51 controls	these miRNAs might be used as biomarkers to diagnose and predict the curative effect of treatment in patients with NSCLC	qRT-PCR	diagnosis, prediction	2020	(77)

(Continued)

TABLE 1 Continued

miRNA	Expression in Research Group vs Controls	Number of cases	Significant Findings	Methods	Application in NSCLC	Year of Publication	Reference
panel of: miR-let-7a-5p, miR-375, miR-1-3p, miR-1291, and miR-214-3p	down: miR-let-7a-5p and miR-375; up: miR-1-3p, miR-1291, miR-214-3p	744 NSCLC/ 944 controls	the five-miRNA panel could detect early-stage (I and II) NSCLC	qRT-PCR	prognosis	2020	(78)
miR-93-5p, miR-18a	up	107 NSCLC/ 42 controls	combination of these miRNAs had a high diagnostic value	qPCR	diagnosis, prognosis	2021	(79)
ex-miR-382	down	126 NSCLC/ 60 controls	level of this miRNA was increased after surgery in most of the patients	qRT-PCR	prognosis	2021	(80)
miR-891a-5p	up	120 NSCLC/ 68 controls	this miRNA had a high diagnostic value (AUC = 0.904) in NSCLC patients	cell cultures, qRT-PCR	diagnosis, prognosis,	2021	(81)
miR-23a, miR-let7i	down	31 NSCLC/21 controls	miR-23a and miR-let7i might serve as non-invasive markers in NSCLC	qPCR	diagnosis	2021	(82)
ex-miR-7	up	35 NSCLC/35 controls	up-regulation of this miRNA was associated with longer survival of patients and with a solid reaction to gefitinib treatment	cell cultures, western blot, qRT-PCR	prediction	2021	(83)
miR-629	up	166 NSCLC/ 70 nonmalignant lung diseases/ 100 controls	miR-629 had better performance for distinguishing NSCLC patients from healthy ones than CYFRA 21-1 and CEA	qRT-PCR	prognosis	2021	(84)
ex-miR-192 and ex-miR-194	down	100 NSCLC/ 100 controls	lower levels of miR-192 and miR-194 were associated with NSCLC TNM stage, metastases and histopathological grade	qRT-PCR	prognosis	2021	(85)
ex-miR-184, ex-miR-3913-5p	up (after the onset of osimertinib resistance)	67 NSCLC/10 controls	these miRNAs may serve as biomarkers to indicate osimertinib resistance	cell cultures, TEM, RNA –seq., qRT-PCR,	prediction	2021	(86)
ex-miR-1169, ex-miR-260	up	64 NSCLC/20 controls	those miRNAs could be biomarkers to distinguish wild-type EGFR and mutant EGFR NSCLC patients in early-stage	TEM, cell cultures, western blot, small RNA seq., validation: qRT- PCR	prediction	2021	(87)
miR-205-5p	up	75 NSCLC/62 controls	increased level of serum miR-205-5p was associated with patients' gender, clinical stage, and drinking status	qRT-PCR	diagnosis, prognosis	2021	(88)
ex-miR-433	down	33 NSCLC/19 controls	this miRNA might help distinguish chemotherapy sensitive patients from resistant ones	TEM, western blot, qRT-PCR	prediction	2021	(89)

(Continued)

TABLE 1 Continued

miRNA	Expression in Research Group vs Controls	Number of cases	Significant Findings	Methods	Application in NSCLC	Year of Publication	Reference
miR-138-5p	down	18 NSCLC/18 controls	ROC analysis revealed that miR-138-5p might be a promising diagnostic biomarker	qRT-PCR	diagnosis	2022	(90)
miR-31-3p	up	140 NSCLC (metastatic NSCLC group – 58; non-metastatic group – 82)/ 60 controls	higher miR-31-3p levels, TNM classification, and lymphatic metastasis might be used as risk factors and independent predictors of bone metastasis	qRT-PCR	prognosis	2022	(91)
miR-30a-5p	down	20 ADC/20 controls	this miRNA correlated with progression and immune infiltration	qRT-PCR	prognosis	2022	(92)

ADC, adenocarcinoma; JCC, American Joint Committee on Cancer; CEA, carcinoembryonic antigen; CYFRA21-1, cytokeratin 19 fragments; EGFR, epidermal growth factor receptor; ex-, exosomal; HOXA5, homeobox A5; IL-1b, interleukin-1b; NSCLC, non-small cell lung cancer; NMLD, non-malignant lung diseases; OS, overall survival; qNano, TRPS (Tunable Resistive Pulse Sensing) particle analyzer; qPCR/qRT-PCR, quantitative polymerase chain reaction/real-time quantitative reverse transcription PCR; SCC, squamous-cell carcinoma; small RNA seq., small RNA sequencing; TEM, transmission electron microscope; TNM, tumour node metastasis; ZEB2, zinc finger E-box binding homebox 2.

TABLE 2 Latest reports concerning miRNAs in plasma of NSCLC patients.

miRNA	Expression in Research Group vs. Controls	Number of cases	Significant Findings	Methods	Application in NSCLC	Year of Publication	Reference
miR-590-5p	down	80 NSCLC/80 controls	negative correlation of that miRNA was observed in NSCLC prognosis	qRT-PCR	prognosis	2020	(93)
ex-miR-1273a	down	49 NSCLC/Cel-miR-39	miR-1273a level was correlated with worse therapeutic outcomes in advanced NSCLC patients receiving platinum-based chemotherapy	TEM, western blot, qRT-PCR	prediction	2020	(94)
ex-miR-320b, ex-miR-320c, ex-miR-320d, ex-miR-125b-5p	down: miR-125b-5p; up: miR-320b, miR-320c, miR-320d	30 NSCLC/7 controls	patients with down-regulated miR-125b-5p might have greater T-cell function and respond well to immunotherapy	TEM, small RNA seq	prediction	2020	(95)
miR-21, miR-128, miR-155, miR-181a	up	128 NSCLC/19 controls	miR-128 and miR-155 could be independent predictors in platinum-based chemotherapy: of worse OS (miR-128) and worse OS in SCC (miR-155)	qRT-PCR	prediction	2020	(96)
miR-1247-5p, miR-301b-3p, miR-105-5p	up	154 NSCLC/146 controls	these miRNAs can serve as potential NSCLC early diagnosis biomarkers	microarrays, validation: qPCR	early diagnosis	2021	(97)
miR-21	up (in patients who achieved a partial/complete treatment response)	39 NSCLC	this miRNA might be used to monitor NSCLC patients during the EGFR-TKIs treatment	qRT-PCR	prediction	2021	(98)

(Continued)

TABLE 2 Continued

miRNA	Expression in Research Group vs. Controls	Number of cases	Significant Findings	Methods	Application in NSCLC	Year of Publication	Reference
ex-miR-1246, ex-miR-96	up	52 NSCLC (27 radioresistant NSCLC, 25 radiosensitive NSCLC)/45 healthy controls	These miRNAs may serve as NSCLC biomarkers	qRT-PCR	diagnosis	2021	(99)
ex-miR-96	up	52 NSCLC (27 radioresistant NSCLC, 25 radiosensitive NSCLC)/45 healthy controls	exosomal miR-96 could be a biomarker of radioresistant NSCLC	TEM, western blot, qRT-PCR	prediction	2021	(99)
miR-183, miR-210, miR-182, miR-144	down:miR-144 (in HPV-DNA-positive NSCLC subjects); up: miR-183, miR-210, miR-182,	100 NSCLC/52 controls	these miRNAs might serve as biomarkers to identify circulating HPV-DNA-positive NSCLC	qRT-PCR	diagnosis of HPV-DNA-positive NSCLC	2021	(100)
miR-216b	down	80 NSCLC/40 controls	miRNA-216b levels were associated with tumour staging and negatively correlated with 18F-FDG uptake in NSCLC	qRT-PCR	diagnosis, prediction	2021	(101)
miR-320a	down	80 NSCLC/80 controls	levels of that miRNA were associated with tumour size, TNM stage and lymph node metastasis	qRT-PCR	prognosis	2021	(102)
miR-210, miR-1290, miR-150, miR-21-5p	up	training set: 40 NSCLC/20 controls, validation: 88 NSCLC/ 50 benign lung disease/40 controls	up-regulation of miR-210 and miR-150 were associated with worse disease–free survival time in NSCLC patients	qRT-PCR	diagnosis, prognosis (only miR-210 and miR-150)	2021	(103)
ex-miR-574-5p, ex-miR-328-3p, ex-miR-423-3p	down: miR-574-5p, up: miR-328-3p, miR-423-3p	30 *EGFR/ALK genes* positive NSCLC/14 controls	these miRNAs may serve as predictors of bone metastasis in NSCLC patients	small RNA seq.	prognosis	2021	(104)
miR-202, miR-26a	up	125 NSCLC/33 controls	miR-202 may be an independent prognostic factor for shorter progression-free and OS	qRT-PCR	prognosis of treatment outcomes	2021	(105)
ex-miR-1260b	up	48 NSCLC/48 controls	miR-1260b up-regulation was associated with high-grade disease, metastasis, and poor survival	TEM, western blot, qRT-PCR	prognosis	2021	(106)
miR-200c, miR-34a	down: miR-34a, up: miR-200c	69 NSCLC/33 controls	levels of these miRNas were associated with the response and outcome in advanced NSCLC patients treated with anti-PD1 immunotherapy	qRT-PCR	prediction	2022	(107)
sEVs-miR-152-3p, sEVs-miRNA-1277-5p	up	10 NSCLC/10 SCLC/ 15 controls	sEVs plasma miRNA-152-3p and miRNA-1277-5p might be used to diagnose early-stage NSCLC patients	TEM, western blot, small RNA seq., validation: qRT-PCR	diagnosis	2022	(108)
miR-340, miR-450b-5p	down: miR-340, up: miR-450b-5p	120 NSCLC/120 controls	these miRNAs might be independent biomarkers of	qRT-PCR	diagnosis, prognosis	2022	(109)

(Continued)

TABLE 2 Continued

miRNA	Expression in Research Group vs. Controls	Number of cases	Significant Findings	Methods	Application in NSCLC	Year of Publication	Reference
			survival in non-metastatic NSCLC				
miR-22-3p, EV-let-7b-5p, EV-miR-184	down: EV-miR-22-3p, EV-miR-184; up: EV-let-7b-5p	20 NSCLC/6 false-positive NSCLC/14 non-cancer lung nodules	these miRNAs could distinguish NSCLC patients from high-risk subjects	small RNA seq., western blot, qRT-PCR	diagnosis	2022	(110)
miR-34c-5p	down	19 NSCLC males/34 males controls	this miRNA could potentially be a diagnostic NSCLC biomarker	qRT-PCR	diagnosis	2022	(111)
ex-miR-130a-3p	up (in relapsed patients compared to non-relapsed)	67 non-metastatic NSCLC/30 metastatic NSCLC	combining of miR-130a-3p and fibrinopeptide A expression was significantly associated with disease-free survival	microarrays, droplet digital PCR, qPCR, nCounter platform, cell cultures	prediction	2022	(112)
EV-miR-625-5p	down	22 SCC/45 ADC	levels of that miRNA could distinguish patients with longer survival	nCounter platform	prediction of survival and response	2022	(113)

18F-FDG, 18F-Fluorodeoxyglucose; EGFR-TKIs, epidermal growth factor receptor tyrosine kinase inhibitors; EV, extracellular vesicle; ex-, exosomal; HPV, human papilloma virus; NSCLC, non-small cell lung cancer; OS, overall survival; RNA seq., RNA sequencing; sEVs, small extracellular vesicles; STAT3, signal transducer and activator of transcription 3; TEM, transmission electron microscope.

TABLE 3 Latest reports concerning miRNAs in sputum of NSCLC patients.

miRNA	Expression in Research Group vs. Controls	Number of cases	Significant Findings	Methods	Application in NSCLC	Year of Publication	Reference
miR-21, miR-32, miR-210	up	training set: 46 NSCLC/55 controls; testing set: 57 NSCLC/62 controls	these miRNAs and two sputum small nucleolar RNAs (SNORD66, SNORD78) had a synergistic effect for lung cancer detection	qRT-PCR	early diagnosis	2016	(114)
miR-31, miR-210	up	discovery cohort: 68 NSCLC/66 cancer-free smokers, validation cohort: 49 NSCLC/50 cancer-free smokers	combination of these two miRNAs with 1 PBMC miRNA (miR-19b-3p) had higher diagnostic value than the individual panel	qRT-PCR	early diagnosis	2018	(115)
miR-145, miR-126, miR-7	down	30 NSCLC/30 controls	these miRNA could be potential biomarker for NSCLC due to their high AUC value and stability in the sputum	qRT-PCR	diagnosis	2018	(116)
miR-31-5p, miR-210-3p	up	76 NSCLC/72 cancer-free smokers	combining these miRNAs and miR-21-5p from plasma had the highest diagnostic value	qRT-PCR	early detection	2019	(117)
miR-21, miR-31,	down: miR-486; up: miR-21, miR-31, miR-210	32 NSCLC/33 cancer-free smokers	these miRNAs could potentially be NSCLC biomarkers	small RNA seq.,	diagnosis, therapy	2020	(118)

(Continued)

TABLE 3 Continued

miRNA	Expression in Research Group vs. Controls	Number of cases	Significant Findings	Methods	Application in NSCLC	Year of Publication	Reference
miR-210, miR-486				validation: qRT-PCR			
miR-31-5p, miR-210-3p	up	40 NSCLC/36 cancer-free smokers	combination of these miRNAs with sputum DNA methylation RASSF1A, plasma miR-21-5p and miR-126 had the highest diagnostic value	droplet digital PCR	diagnosis	2021	(119)

NSCLC, non-small cell lung cancer; PBMC, peripheral blood mononuclear cells; qPCR/qRT-PCR, quantitative polymerase chain reaction/real-time quantitative reverse transcription PCR; RASSF1A, ras association domain family member 1; small RNA seq., RNA sequencing.

TABLE 4 The most promising miRNAs in NSCLC diagnosis.

miRNA/panel	Source of miRNA	AUC	Sensitivity (%)	Specificity (%)	Application in NSCLC	Year of Publication	Reference
miR-762	serum	0.92	81.8	93.33	distinguishing NSCLC patients at the clinical stage I from healthy controls	2020	(56)
let-7d-5p	serum	0.91	76	100	distinguishing NSCLC patients from healthy controls	2020	(62)
miR-17 + miR-20a	serum	0.94	98	0.90	2-miRNA panel for stage I-III NSCLC diagnosis	2020	(65)
miR-518b	serum	0.91	88.1	81.7	distinguishing NSCLC patients from healthy controls	2020	(66)
miR-185 + CEA	serum	0.93	97.3	78.7	distinguishing patients with carcinoma from healthy controls	2020	(68)
ex-miR-1228-3p	serum	0.91	0.77	0.89	distinguishing NSCLC patients from healthy controls	2020	(71)
miR-21-5p, miR-141-3p, miR-222-3p, miR-486-5p) and two exosomal ones (miR-146a-5p and miR-486-5p	serum	0.96	85.42	92.5	early diagnosis of NSCLC patients	2020	(76)
miR-184 + miR-191	serum	0.929	89	84.75	distinguishing NSCLC patients from pneumonia subjects	2020	(77)
miR-184 + miR-191	serum	0.925	87	86.25	distinguishing NSCLC patients from pneumonia subjects	2020	(77)
miR-891a-5p	serum	0.904	82.5	80.9	distinguishing the patients with NSCLC from the healthy controls	2021	(81)
miR-210 + miR-144	plasma	0.98	93.7	98	detecting HPV-DNA-positive NSCLC patients	2021	(100)
panel of: miR-210, miR-1290, miR-150 and miR-21-5p	plasma	0.98	93.75	98	differentiating patients with stage II-IIIA NSCLC from healthy individuals	2021	(103)
miR-31 and miR-210 and miR-19b-3p	sputum, miR-19b-3p (from PBMC)	0.92	82.3	87.9	distinguishing the patients with NSCLC from the healthy controls	2021	(120)

MiRNAs with high AUC>0.9, with high specificity or/and specificity (>80%).

find out recent research studies. In the case of serum and plasma, we selected research studies from 2020 to 2023. Due to the small amount of research on sputum, we selected research studies from 2016 to 2022. Moreover, we excluded studies from journals with an Impact Factor below 2.0 and those with the small study groups (<10).

2 Circulating miRNAs as diagnostic biomarkers for NSCLC

The high mortality rate from NSCLC is related to diagnosing patients in advanced stages. Early detection might improve patient outcomes and prevent complications. Diagnostic biomarkers are needed to determine whether a patient has NSCLC, discriminate it from other disease entities, and help determine its exact subtype of NSCLC. The importance of diagnostic markers is invaluable, as their results can be potentially used for later predicting the disease's progression and the patient's treatment response. Several studies have been carried out that have indicated the diagnostic potential of miRNAs in NSCLC.

In 2020, Yu (53) and Sui indicated that serum miR-30 might be a diagnostic biomarker of NSCLC patients. Levels of serum and tissue miR-30 were observed as significantly decreased compared to healthy controls. A receiver operating characteristic curve (ROC) analysis showed a high diagnostic accuracy with an area under the curve (AUC) value was 0.802, with 75.9% sensitivity and 76.0% specificity. It is worth conducting more extensive research on larger cohorts of NSCLC patients since miR-30 family members are proven to be involved in oncogenesis and invasion in different types of tumours (121). One of the recent studies focused on the clinical significance of serum exosomal miR-1246 in NSCLC. The authors showed for the first time a significant up-regulation of that miRNA in NSCLC patients than in non-malignant respiratory disease patients and healthy controls. Moreover, ROC analysis revealed that this miRNA differentiated early-stage NSCLC patients from healthy and non-malignant respiratory disease patients with an AUC value of 0.827 and 0.757, respectively (54). According to the research of Tang et al. (55), exosomal-derived serum miR-620 may be another promising diagnostic biomarker of NSCLC. Microarrays and subsequent validation via qPCR were performed using serum exosomes from 235 NSCLC patients and 231 healthy donors. The under-expression of miR-620 levels has been observed in NSCLC and early-stage patients' exosomes compared to healthy individuals. The strengths of this study were the validation performed and the size of the study group. In addition, the study and control groups were characterized by the absence of any immunological, endocrine or metabolic diseases in the study and control groups. The results of another study by Chen et al. (56), showed up-regulation of serum miR-762 in NSCLC patients with the advanced clinical stage (III-IV), positive lymph node metastasis and poorly differentiated tumors towards patients at the early clinical stage. The combination of carcinoembryonic antigen (CEA), and cytokeratin 19 fragments (CYFRA21-1) with serum miR-762 enhanced the diagnostic efficiency for NSCLC. It was the first research to demonstrate that the level of serum miR-762 was significantly up-regulated in NSCLC patients, but validation on larger cohorts is needed to confirm those reports. It has been shown that miR-519b in NSCLC tissue and serum samples was markedly lower. ROC analysis demonstrated that serum miR-519d levels could discriminate NSCLC patients from healthy controls with an AUC of 0.855 ($p < 0.0001$). When the cut-off value was 0.22, the sensitivity and specificity were 98.1% and 91.8% (57). Altered expression of miR-519d was previously observed in other cancers (122, 123), so more accurate high-throughput cohort studies are needed to assess the diagnostic value of miR-519 for NSCLC. The study of Zhang et al. (58) has shed light on the potential utility of miR-5684 and miR-125b-5p in NSCLC diagnosing. Firstly, the authors used transmission electron microscopy (TEM), particle analyzer (qNano) and western blots to characterize the exosomes, then screened them out by microarrays to finally be verified by qPCR. The highest diagnostic accuracy (AUC = 0.863) of tested miRNAs was reached after combining it with two tested markers – 19 CYFRA21-1 and CEA. Interestingly, miR-5684 expression was markedly lower in patients in the T1-T4 compared to healthy subjects, whilst miR-125-5p was lower at T2 and higher stages. Moreover, miR-125b-5p allowed for the precise distinction of patients with the early stage (I and III) and advanced stage (III and IV). Another research was conducted using microarrays to screen 2.549 miRNAs in serum samples of NSCLC patients. Subsequently, qRT-PCR validation helped select miR-4687-3p, which next showed the highest diagnostic accuracy of NSCLC from the other five miRNAs (miR-1915-5p, miR-432-3p, miR-4488, miR-520a-5p, and miR-6087). On that basis, a comparison with information from The Cancer Genome Atlas (TCGA) database was made, what had confirmed that the level of miR-4687-3p was significantly higher in NSCLC tissues than in normal lung tissues ($p < 0.05$) (59). There is a need to identify the underlying molecular mechanisms of miR-484, therefore the authors plan to implement it in their further research. Some studies have focused on integrating results concerning miRNA expression from different body fluids to improve NSCLC detection (124). The participation of miR-145 has been repeatedly proven by scientists to be involved in NSCLC (63, 125, 126). An interesting and comprehensive systemic review and meta-analysis of that miRNA was published in 2020. The authors found nine studies (with 11 data sets) concentrated on miR-145 role in NSCLC, and as a result, the summary ROC of the miR-145 for the detection of NSCLC amounted to AUC = 0.83. Due to the authors, miR-145 can potentially be a diagnostic biomarker of NSCLC, but further studies are still required (60). Another group investigated serum miRNAs from the dataset GSE24709, which were: miR-432, miR-942, miR-29c-5p, and miR-601. Data validation was carried out on 20 NSCLC patients and 20 healthy volunteers allowing a selection of miR-942 and miR-601 as the most up-regulated in NSCLC patients. Further analysis on a larger cohort allowed the authors to confirm significantly higher levels of these miRNAs in NSCLC patients than in healthy controls. Furthermore, these miRNAs performed better than CEA, CYFRA21-1, and the SCC antigen in the early detection of NSCLC (61). One of the recent studies on Indian NSCLC patients aimed to validate serum miRNAs, which were selected as differentially expressed after small RNA sequencing. The authors

found miR-15a-5p, miR-320a, miR-25-3p, miR-192-5p, miR-let-7d-5p, miR-let-7e-5p, miR-148a-3p, and miR-92a-3p dysregulated in serum of NSCLC patients. Additionally, miR-375 and miR-10b-5p were significantly down-regulated in SCC patients compared to controls. Let-7d-5p showed the highest diagnostic value with AUC = 0.917 and a sensitivity of 76% at 100% specificity (62). It is worth confirming these studies on larger groups of patients from other populations to assess the diagnostic value of the miRNAs included in this study. Microarray analysis indicated that the serum exosomal miR-20b-5p and miR-3187-5p were remarkably lower in NSCLC patients compared to healthy subjects. Next, it was verified via RT-qPCR on larger groups containing 276 NSCLC patients and 282 healthy controls. Combining both miRNAs with CEA or CYFRA21-1 showed AUC = 0.905 for miR-20b-5p and AUC = 0.894 for miR-3187-5p. Furthermore, integrating those two miRNAs could discriminate patients at an early stage from healthy controls (64). An undoubted advantage of this study was the size of the tested groups. Nevertheless, the authors point out the need to examine the potential utility of these miRNAs to distinguish NSCLC patients from patients with benign lung diseases. Several studies proved that the miR-17-92 gene cluster participates in cancer development by targeting mRNAs involved in distinct pathways that promote or inhibit carcinogenesis (127–130). This cluster consists of six miRNAs: miR-17, miR-18a, miR-19a, miR-19b-1, miR-20a, and miR-92a-1. The latest research by Yang et al. (65) aimed to examine the potential diagnostic utility of the miR-17-92 gene cluster in NSCLC. ROC analysis proved that the miR-17-92 gene cluster might be helpful as a diagnostic NSCLC marker. Furthermore, the high positive correlation between miR-17 and miR-20a expression was observed, indicating these miRNAs could be used together as a panel to improve the accuracy of the NSCLC diagnosis. Serum miR−518b was up-regulated in NSCLC serum, tumor tissues and cell lines (A549, H1299, H1975, and PC9) compared to coherent healthy controls. The diagnostic accuracy of miR-518b was supported by ROC analysis, where the AUC value was 0.910, with 88.1% sensitivity and 81.7% specificity (66). A recent study by Zhang and Xu (67) indicated that combining exosomal miR-378 with CEA could accurately differentiate NSCLC patients from healthy donors. The role of miR-185 in oral cancer (131), prostate cancer (132) and in NSCLC has been previously evaluated (133, 134). In 2020, Liu et al. (68) demonstrated that serum miR-185 might be a promising biomarker for NSCLC early detection. Their work showed significantly reduced miR-185 expression in NSCLC patients with lymph node metastasis at the advanced stage or with poor differentiation. Moreover, serum miR-185 demonstrated better diagnostic accuracy than CEA for distinguishing patients with carcinoma from disease-free controls. Xue et al. (69) identified 12 miRNAs profiling studies and analyzed differentially expressed miRNAs according to GEO2R online tool and RRA method from R. Performing validation via RT-qPCR revealed reduced miR-1228-3p serum levels (P = 0.006) and increased miR-181a-5p serum levels (p = 0.030) compared to healthy controls. ROC analysis results indicate the potential utility of miR-1228-3p and miR-191a-5p as NSCLC diagnostic biomarkers. Wang et al. selected four miRNAs (miR-1269a, miR-205-5p, miR-210-5p, and miR-9-3p) from TCGA database, which occur abundantly in serum exosomes from NSCLC patients. Their further analysis proved that these miRNAs may serve as novel diagnostic panel distinguishing NSCLC patients from healthy subjects (71). Another authors found lower miR-186 levels in serum and exhaled breath condensate (EBC) of NSCLC patients compared to healthy donors. The analysis showed that the SCC group had a higher serum interleukin-1β level than the ADC. Moreover, a larger AUC was obtained when the interleukin-1β and miR-186 levels were combined (72). Yang et al. (74) indicated the potential diagnostic role of exosomal miR-21/Let-7a ratio, which was up-regulated in NSCLC patients compared to healthy controls, patients with pulmonary inflammation diseases, and benign pulmonary nodules. Analysis revealed that combined levels of both miRNAs could precisely differentiate NSCLC patients from pulmonary nodules diseases and healthy controls. This study highlights that combining more than one miRNA may have better diagnostic value than testing individual miRNAs' alone. A limitation of this study was the relatively small number of patients in the groups. However, it has the advantage of comparing NSCLC (n = 75) not only with healthy controls (n = 24) but also with benign pulmonary nodules (n = 23) and pulmonary inflammation diseases (n = 18). Analysis of larger patient cohorts is needed to more precisely assess the value of exosomal miR-21/Let-7a ratio to serve as NSCLC biomarker (74). Wu et al. (76) conducted a groundbreaking study exploring the same eight serum and exosome-derived miRNAs. Significant miRNA up-regulation was observed in four early-stage NSCLC serum particles (miR-21-5p, miR-141-3p, miR-222-3p, and miR-486-5p) and two exosomal ones (miR-146a-5p and miR-486-5p) compared to patients with benign lung lesions and healthy donors. Combining the results of over-expression of these miRNAs gave the best results in terms of the diagnostic power of the test. (AUC = 0.960, with a sensitivity of 85.42% and a specificity of 92.50%). These findings encourage an exploration of the potential utility of combined serum and exosomal miRNAs, which can improve NSCLC early diagnosis. Results from the study conducted by Ding et al. (77), indicated serum miR-184 down-regulation and miR-191 up-regulation in NSCLC patients compared to controls and pneumonia patients. Further analysis proved that these miRNAs may potentially serve as diagnostic biomarkers in NSCLC. Performing a ROC curve analysis showed better diagnostic efficacy when both miRNAs were combined. In turn, levels of miR-93-5p and miR-18a were significantly up-regulated in NSCLC patients compared to disease-free controls. The AUC value for both mRNAs in NSCLC diagnosis was high (0.89). According to the authors, further analysis of larger cohorts is needed to understand the role of these miRNAs in NSCLC (79). Another group recognized miR-891a-5p to have diagnostic value in NSCLC. Increased levels of miR-891a-5p were notably increased in serum, tissue and NSCLC cell lines (H1299, HCC827, H460, and A549). Expression levels of serum miR-891a-5p could accurately distinguish NSCLC patients from the healthy controls with the AUC = 0.904 (81). The study of Kryczka et al. (82) in 2021 was focused on finding diagnostic NSCLC biomarkers in serum exosomes. Among four surveyed miRNAs, the authors found miR-23a and miR-let7i significantly down-regulated in NSCLC compared to healthy controls. All vehicle-derived miRNA

included in the research (miR-23a, miR-361, miR-1228 and miR-let7i) were connected in panels, and results showed better diagnostic values with AUC of 0.705 for miR-23a and miR-let7. Zhao et al. (88) indicated that serum miR-205-5p could accurately differentiate patients in early NSCLC stage (I and II) and advanced stage (III and IV) when compared to healthy controls, with the AUC of 0.8141 and 0.8045, respectively. However, some studies on miR-miR-205-5p have been contradictory, so there is a need to precisely define the role and utility of this miRNA in diagnosing NSCLC (135). MiR-138-5p was found under-expressed in NSCLC cell lines (A549, H1975, and PC9), serum and lung ADC tissues. NSCLC could be distinguished accurately from healthy controls by miR-138-5p with an AUC of 0.922, which supports this miRNA potential utility in NSCLC diagnosing (90). In another study, microarray analysis of miRNA from four NSCLC patients and five healthy subjects helped to select miR-1247-5p, miR-301b-3p, and miR-105-5p. Next, those miRNAs were validated via qPCR on 154 NSCLC patients and 146 controls. All those miRNAs were highly expressed in diseased patients' plasma, and further analysis proved its role in NSCLC tumorigenesis. After combining those three miRNAs with CEA, ROC analysis showed higher sensitivity and specificity than CEA alone (97). A study conducted by Zheng Q et al. (99) suggests that exosomal miR-1246 and miR-96 from plasma may be helpful in diagnosing patients with NSCLC. However, miR-96 showed significantly higher diagnostic potential (AUC = 0.9735) than miR-1246 (AUC = 0.6761). Wu et al. (100) focused on human papillomavirus (HPV) infection, which seems to be a dangerous risk factor in cancer patients (136–138). They found that miR-183, miR-210, and miR-182 were significantly higher, and miR-144 was markedly lower in HPV-DNA-positive NSCLC than in HPV-DNA-negative NSCLC patients. Notably, miR-210 combined with miR-144 had the best prediction performance in diagnosing HPV-DNA-positive NSCLC patients. In the subsequent study from 2021, a reduction of plasma miR-216b was observed in early, advanced ADC and SCC patients. Lower levels of that miRNA were also observed in early ADC and SCC tissues compared to adjacent ones. Results of this study demonstrated that miR-216b reduced expression was correlated with the tumor staging, indicating this miRNA potential utility in NSCLC patients classification (101). Jiang et al. (103) selected 12 previously found dysregulated miRNAs in NSCLC, investigated them via RT–qPCR in the training group, and selected four miRNAs (miR–210, miR –1290, miR–150, and miR–21–5p), which were subsequently validated on the testing set. Expression of all four miRNAs proved to be over-expressed in NSCLC patients compared to patients with benign lung disease and healthy donors. Combining all tested miRNAs as a panel showed a higher diagnostic performance than individual miRNAs alone. The limitation of the study was too short follow-up and a relatively small number of included patients. A study by Jiang (108) aimed to investigate the role of plasma circulating small extracellular vesicles (sEVs) in early-stage NSCLC and SCLC. Next Generation Sequencing revealed six dysregulated miRNAs (miR-483-3p, miR-152-3p, miR-1277-5p, miR-130b-3p, miR-25-5p, and miR-4429), which were further validated via RT-qPCR. The results showed that sEVs plasma miRNA–152–3p and miRNA–1277–5p might be used to diagnose early–stage NSCLC patients. Wu et al. (109) found plasma miR-340 expression decreased and plasma miR-450b-5p increased in NSCLC patients compared to healthy donors. AUC values were 0.740 and 0.808, respectively. When both miRNAs were combined, the AUC value was 0.862 with 78.33% specificity and 77.5% sensitivity. It is worth conducting more detailed analyses to assess the value of both these plasma miRNAs as a panel. Vadla et al. (110) drew attention to the problem of false positive NSCLC diagnosis, which leads to unnecessary testing and treatment. The authors uncovered that combining plasma miR-22-3p, vesicle-derived miR–184, and miR-let–7b–5p distinguished NSCLC patients from high-risk subjects. This valuable research highlights the potential role of those miRNAs in improving the precision of NSCLC diagnosis. The validation of miR-34c-5p via qRT-PCR showed its up-regulation in NSCLC plasma samples from males compared to healthy donors (p = 0.004, AUC = 0.8467). These findings were consistent with the microarray results on tissues and plasma samples (111). In the newest meta-analysis by Wang et al. (139), it is concluded that miR-21 from serum or sputum can be a promising biomarker in the identification of patients with lung cancer, including NSCLC, with high precision. In this paper, four studies tested the miR-21 expression in sputum and 13 in serum samples. In both cases, the specificity and sensitivity for miR-21 were relatively high.

Several studies have reported that NSCLC diagnosing can be aided by determining miRNAs in the patient's sputum. Sputum is a kind of mucus that is an expectorated secretion of the respiratory system. It includes epithelial cells from the lungs and lower respiratory tract, microbial products, inflammatory cell components, and traces of saliva. It is readily available, easy to collect, cost-effective, and can reflect the conditions of the lungs (140–142). Sputum collection may be facilitated by inhalation of nebulized hypertonic saline solution (143).

In the meta-analysis combining results from 14 articles, the authors suggest that the integration of miR-210, miR-21, and miR-31 from sputum may be a good biomarker for NSCLC diagnosis (144). The next meta-analysis study shows the high diagnostic significance of miR-210 in the serum and sputum of NSCLC patients. The authors analyzed data from GEO and TCGA. Indicated AUC values for this miRNA were 0.82 in the serum and 0.81 in the sputum (145). A recent study by Su et al. (114) revealed that the combination of miRNA with other non-coding RNAs might be profitable in the diagnosis of NSCLC patients. Integrating three sputum miRNAs (miR-21, miR-32, and miR-210) and two sputum small nucleolar RNAs (SNORD66, SNORD78) synergistically affect lung cancer detection. It showed higher sensitivity and specificity than miRNA or those small nucleolar RNAs analyzed alone. In 2018 (115), the same authors integrated two sputum miRNAs (miR-31 and miR-210) and miR-19b-3p from peripheral blood mononucleated cells (PBMC) in the NSCLC group. This panel had higher sensitivity and specificity than these miRNAs analyzed separately. The above studies were the continuation of previous research by these scientists, which have proven the potential utility of sputum miR-21, miR-31, and miR-210 as NSCLC biomarkers (120, 146). The logistic regression model based on miR-145, miR-126, and miR-7 expression, obtained by

qRT-PCR reaction, showed 0.93 AUC in the lung cancer group, which is a promising result for NSCLC diagnosis. Furthermore, the miRNA showed stability in the sputum even after one year of storage. The main limitation of this study was the relatively small specimen number, which consisted of 30 NSCLC patients and 30 healthy controls (116). Li et al. (119) indicated using digital droplet PCR, a combination of biomarkers sputum miR-31-5p and miR-210-3p, sputum DNA methylation *RASSF1A*, and two plasma miR-21-5p and 126 in NSCLC diagnosis. The integration of these three components showed AUC = 0.956, which is higher than a single type of biomarker. The results were confirmed in a validation cohort, which indicated the high potential of this combination in NSCLC diagnosis. Another study of these authors showed that combining sputum miR-31-5p and miR-210-3p with plasma miR-21-5p synergistically affected early NSCLC diagnosis (117).

Sputum appears to be a very promising source of miRNAs diagnosing NSCLC. However, the number of publications focusing on miRNAs in sputum is limited compared to those conducted on serum and plasma. We have not found any studies demonstrating a clear prognostic and predictive potential of sputum miRNAs in NSCLC. The studies presented here require further validation and testing in more significant numbers of patients. It is worth noting that miR-21, miR-31, and miR-210, in particular, are repeated in several studies demonstrating their efficacy in diagnosing patients with NSCLC. However, most of these studies were conducted by the same authors, so the groups of patients may have come from a similar population. Independent research centres should confirm these results on independent patient groups. Analyzing the studies, the conclusion is that miRNAs assembled in panels have more significant diagnostic potential than those miRNAs analyzed separately. This usually increases the accuracy and precision of the diagnostic test.

3 Circulating miRNAs as prognostic biomarkers for NSCLC

TNM Classification of Malignant Tumors plays a significant role in NSCLC prediction and prognosis, allowing for the estimation of patient performance status, survival time, histologic tumour grade, and relapse rate. For instance, the presence of any malignant pleural effusion, distant disease or contralateral nodule is associated with 5-year survival rates of less than 6%. However, searching for new molecular and clinical biomarkers is essential since heterogeneity in outcomes is perceived among the same TNM groups. Identifying prognostic biomarkers for NSCLC is of great importance to refine risk stratification and guide treatment planning (147, 148).

In 2020 (53), serum and tissue miR-30 levels were found down-regulated in NSCLC patients compared to healthy donors. Importantly, a statistically significant difference was found in the median overall survival (OS) between patients with under-expression of miR-30 (23.0 months) and patients with over-expression of that miRNA (36.0 months). Levels of miR-30 were correlated with lymph node metastasis, tumor size, tumor node metastasis (TNM) stage, and differentiation degree. Huang and Qu (54) demonstrated that the patients with higher levels of serum exosomal miR-1246 had poorer OS and disease-free survival. Further analysis showed that the level of that miRNA, as well as, the TNM stage and lymph node metastasis, were proven to be independent prognostic factors in NSCLC. Another study aimed to determine whether circulating serum miR-762 can improve the NSCLC diagnosis and prognosis. The results showed up-regulation of this miRNA in NSCLC patients with the advanced clinical stage (III and IV), positive lymph node metastasis and poorly differentiated tumors towards patients at the early clinical stage. Furthermore, it has been observed that the level of miR-762 was statistically associated with histological grade, clinical stage, and lymph node metastasis (56). It has been shown that miR-519b expression in NSCLC tissue and serum samples was markedly lower. The association between lymph node metastasis and clinical stage have been observed, as well as significantly shorter OS rates than patients with higher serum miR-519d levels. Notably, the cell-based part of the research has revealed that miR-519d regulates NSCLC progression by targeting human epidermal growth factor receptor 3 (57). Zhou et al. (61) indicated serum miR-942 and miR-601 as over-expressed in NSCLC patients compared to control. Their further analysis showed the potential role of both miRNAS as NSCLC prognostic biomarkers. Higher levels of one tested miRNA (miR-942 or miR-601) were connected with adverse clinical variables and poor survival of NSCLC patients. Interestingly, the most favourable outcome could be observed in NSCLC patients with low levels of miR-942 and miR-601. A recent study conducted on Indian NSCLC patients (62) pointed to a potential prognostic role of serum miR-375 and miR-10b-5p. Levels of that miRNAs were down-regulated in SCC patients compared to controls. Authors observed significant correlation of miR-375 with lymph node involvement ($p = 0.0224$) and with pleural effusion ($p = 0.0148$), while miR-10b-5p association was seen only with the pleural effusion ($p = 0.0037$). In 2020, X. Zhang et al. (66) found increased levels of serum miR-518b, which was observed also in tissues and cell lines compared to corresponding biological material. Over-expression of miR-518b was significantly associated with tumor size ($p = 0.042$), advanced TNM stage ($p = 0.006$) and lymph node metastasis ($p = 0.039$). Significantly, miR-518b over-expression was likewise associated with shorter survival of patients ($p = 0.009$). J. Liu et al. (68) as first demonstrated that decreased level of serum miR-185 was markedly associated with worse survival and adverse clinicopathological parameters in NSCLC patients. Performing univariate and multivariate Cox regression analysis indicated that this miRNA was an independent prognostic indicator for NSCLC. Another study investigated the levels of serum miR-1228-3p and miR-181a-5p in 50 NSCLC patients. They showed that increased levels of miR-1228-3p and decreased levels of miR-181a-5p were significantly associated with OS. MiR-1228-3p was characterized as independent factor for poor prognosis in Cox regression analysis (69). In the study by Kumar et al. (70), miR-3195 from serum was characterized as an independent prognostic factor for OS. Patients with higher levels of that miRNAs are predisposed to significantly longer OS ($p = 0.0298$). In a study conducted by Xu et al. (75), serum exosomal let-7e could differentiate NSCLC patients from

healthy controls. This tendency has also been noticed in tissues. Kaplan–Meier method showed that miR-let-7e expression and higher *SUV39H2* expression were correlated with lower survival rates of NSCLC patients. Shao et al. reported a significant increase in levels of miR-93-5p and miR-18a in 107 NSCLC patients compared to osimertinib controls (n = 42). Both miRNAs' expressions were associated with tumor differentiation degree, TNM stage, lymph node metastasis and lymph-vascular space invasion. Another study was focused on exosomal miR-382, which turned out to be significantly down-regulated in NSCLC patients. Interestingly, further research revealed that most of the patients, after surgical resection, had increased miR-382 expression (80). The results of another study showed up-regulated expression of miR-891a-5p in serum and tissues from NSCLC patients compared to NSCLC cell lines. According to the authors, miR-891a-5p might serve as a prognostic NSCLC biomarker, inferred from association with differentiation, tumour, node and metastases stages (81). Serum miR-629 was another example of up-regulated miRNA in NSCLC compared to controls. Dysregulation of that miRNA was positively related with lymph node metastasis, differentiation and clinical stage. Moreover, patients with higher miR-629 levels had poorer OS and disease-free survival than subjects with lower levels of miR-629 (84). Wang et al. found that miR-192 and miR-194 were remarkably lower in NSCLC patients. Both miRNA levels were linked to the TNM, distant metastases, and pathological stages. Moreover, miR-192 was as well correlated with the pathological stage (85). In another study from 2021 (88), serum levels of miR-205-5p were found to increase and it was associated with patients' gender, drinking status, and clinical stage. In turn, exosomal miR–433 was down-regulated in chemotherapy-resistant NSCLC plasma patients compared with chemotherapy-sensitive NSCLC and normal serum. Expression of this miRNA was negatively associated with large tumour size, distant metastasis, advanced TNM stage and poor prognosis (89). The research of Zeng et al. (91) showed that the expression miR-31-3p level was significantly higher in serum and NSCLC tissue. Further analysis showed that high expression of miR-31-3p with the TNM classification, and lymphatic metastasis could be used as risk factors and independent predictors of bone metastasis, while tumour size could be used as a risk factor for bone metastasis. The role of miR-30a-5p in ADC was raised in a recent study by Jiang et al. (92). According to the results of their work, expression of that miRNA was significantly decreased in ADC cell lines, tissues and patients' serum. Lower levels of miR-30a-5p were, inter alia, linked with the TNM stage, pathologic stage, gender, and smoking (92). Another study showed that plasma miR-590-5p levels were dramatically down-regulated in NSCLC patients compared to healthy controls. Importantly, patients with stages III and IV had marginal down-regulation of that miRNA compared to healthy controls. Furthermore, Kaplan-Meier and log-rank analyses revealed a negative correlation of miR-590-5p in the prognosis of NSCLC patients (93). The recent Khandelwal et al. (102) study detected that plasma miR-320a was under-expressed in NSCLC patients. Furthermore, levels of that miRNA were associated with tumor size, TNM stage, and lymph node metastasis. In another study, the authors highlighted the problem of bone metastasis in NSCLC patients. The purpose of their research was to find high-accuracy biomarkers to be alternatives for skeletal scintigraphy, computerized tomography (CT), positron emission tomography-computed tomography (PET-CT), and magnetic resonance imaging (MRI). Within one of the tested miRNA clusters, the authors found significant under-expression of plasma-derived exosomal miR-574-5p, over-expression of plasma-derived miR-328-3p and miR-423-3p in patients with bone metastasis compared with subjects without that ailment (stage IV) (104). Jiang et al. (103) investigated the role of four plasma miRNAs (miR-210, miR-1290, miR-150, and miR-21-5p) in NSCLC prognosis and diagnosis. Kaplan–Meier survival analysis demonstrated that two miRNAs (miR-210 and miR-150) were significantly correlated with a shorter disease-free survival time in patients without SCC lung cancer, but not in SCC patients. Significant differences were not observed in plasma levels of miR−210 and miR−150 between patients with SCC and patients without SCC lung cancer. Kim et al. (106) analyzed plasma exosomal miR-1260b role in NSCLC. Their results showed its up-regulation in tissues, plasma and human umbilical vein endothelial cells. These levels were connected with high-grade disease, metastasis, and poor survival. Another study characterized plasma miR-340 and miR-450b-5p as independent biomarkers of survival in non-metastatic NSCLC patients. Besides, plasma miR-340 was also negatively correlated with tumor grade (109).

4 Circulating miRNAs as predictors of NSCLC treatment response

Drug- and radio-resistance in NSCLC patients are major causes of therapeutic failure. Hence, choosing the right way of treatment is crucial to prevent tumor recurrence, disease progression and unnecessary side effects. Some authors indicate that routine treatment with EGFR-TKI should be avoided in molecularly unstudied populations (149, 150). Accumulating evidence has shown that miRNA concentrations might be varied between treatment-sensitive and treatment-resistant NSCLC patients. This paragraph presents recent studies focused on the search for miRNAs, which have predictive potential in NSCLC treatment response.

Tang et al. (55) revealed the potential role of exosomal-derived serum miR-620 as a predictive biomarker in NSCLC patients. Results showed an association of lower levels of that miRNA with chemotherapeutic effect (p = 0.044), which can be helpful in the prediction of response to chemotherapy. Moreover, significant down-regulation of that miRNA was observed in patients with metastatic NSCLC, showing its potential in the prediction of future metastasis. According to the research of Zhang et al. (58), exosomal miR-5684 and miR-125b-5p might serve as prognostic biomarkers in NSCLC. Under-expression of these miRNAs was observed in diseased serum samples compared to healthy controls. Importantly, levels of exosomal miR-5684 and miR-125b-5p were significantly associated with metastasis ($p < 0.0001$), chemotherapeutic effect ($P = 0.007$) and survival ($P = 0.008$). This study was characterized by a relatively large study group what makes the results convincing. However, the authors emphasize the importance of further examining the insights mechanism of exosomal miR-5684 and

miR-125b-5p. A recent study by Zhang and Xu (67) reported that the serum exosomal miR-378 was notably up-regulated in NSCLC patients. These levels were significantly related to positive lymph node metastasis and advanced TNM stage. Among the many interesting observations, the level of miR-378 was decreased after radiotherapy and could be potentially used as an indicator of radiotherapeutic response. Kumar et al. (70) as first investigated the role of miR-3692-3p, miR-3195, and miR-1249-3p as circulating NSCLC biomarkers. The results indicated the potential usefulness of miR-1249-3p in prediction of therapeutic response to chemotherapy. Moreover, it was found that levels of miR-1249-3p were significantly higher in ADC versus SCC (p = 0.0178). The subsequent interesting study has indicated serum-derived miR −130a, miR−25, and miR−191 up-regulation in NSCLC patients compared to disease-free controls. These levels were associated with advanced age (≥ 60 years), radiotherapy, histological type (SCC), low survival rate and low median survival time. These miRNAs could potentially serve as predictors of radiotherapy outcome in NSCLC patients (73). Gefitinib is the first selective inhibitor of epidermal growth factor receptor's (EGFR) tyrosine kinase domain that was the first targeted drug that entered clinical practice for treating lung cancer (151). Ding et al. point to potential usefulness of serum miR-184 and miR-191 as biomarkers for predicting gefitinib efficacy in NSCLC patients. Results showed that increased miR-191 and decreased miR-184 expressions in NSCLC patients were associated with a higher risk of ineffective gefitinib treatment (77). Chen et al. (83) also indicated the potential prognostic potential role of miRNAs in treatment predicting by gefitinib. Exosomal miR-7 from serum was found dysregulated and it was associated with a longer survival rate of patients, early tumor size and a solid reaction to gefitinib treatment. Osimertinib is the first third-generation epidermal growth factor receptor (EGFR) tyrosine kinase inhibitor (TKI) approved for treating locally advanced or metastatic NSCLC in patients with epidermal growth factor receptor (EGFR T790M) resistance mutation (152). Li et al. (86) found up-regulation of exosomal miR-184 and miR-3913-5p in NSCLC patients after the onset of osimertinib resistance. MiR-184 amounts were correlated with lactate dehydrogenase levels, whether expression of miR-3913-5p with TNM stage, platelet count, CEA and metastases. This study suggests that these two miRNAs may serve as biomarkers for detecting NSCLC patients being resistant to osimertinib. In another study, serum exosomal miR-1169 and miR-260 have been presented as potential biomarkers discriminating between wild-type epidermal growth factor receptor (EGFR) and mutant EGFR NSCLC at an early stage (87). In turn, serum exosomal miR−433 was down-regulated in chemotherapy-resistant NSCLC patients' plasma compared with sensitive NSCLC and serum of healthy donors. Expression of this miRNA was negatively associated with large tumor size, distant metastasis, advanced TNM stage and poor prognosis. Platinum-based chemotherapeutics, mainly cisplatin and carboplatin, are routinely used for the treatment of lung cancer (153, 154). Plasma exosomal miR-1273a was found significantly under-expressed in the non-responder NSCLC patients after cisplatin therapy. Furthermore, a significant association was observed between miR-1273a low levels and worse therapeutic outcomes of advanced NSCLC subjects receiving platinum-based chemotherapy. This study demonstrates that miR-1273a is related to cisplatin resistance and might be helpful in NSCLC prognosis and therapy (94). Another research by Peng et al. (95) revealed higher levels of plasma miR-320d, miR-320c, and miR-320b in the progressive NSCLC disease group compared with the achieved partial response group at the beginning of the therapy. Furthermore, decreased plasma exosomal miR-125b-5p levels were observed in post-treatment patients, which might help to gain greater T-cell function and respond well to immunotherapy. Papadaki et al. (96) researched miRNAs (plasma miR-21, miR-128, miR-155, and miR-181a) which involvement in damage response (155) and tumor responsiveness to platinum was previously studied (156, 157). According to the results, these miRNAs were significantly up-regulated in NSCLC patients compared to healthy controls. Among interesting findings, miR-128 was associated with worse OS, whereas miR-155 with OS in SCC with platinum-based chemotherapy. Those two miRNAs may function as independent predictors in platinum-based chemotherapy. Leonetti et al. (98) conducted a comprehensive study to explore the role of plasma miRNAs (miR-21, miR-27a, and miR-181a) which could potentially serve as substitute for EGFR-TKI treatment. The level of miR-21 was observed as up-regulated in NSCLC patients who partially/completely responded to the treatment, when compared to patients with disease stability/progression. Moreover, sixth-month lasting clinical benefits patients have shown higher basal levels of circulating miR-21 (p = 0.039). Another study aimed to examine the possible utility of sixth exosomal miRNAs (miR-21, miR-1246, miR-let-7g, miR-210, miR-214, and miR-96) from plasma, but only miR-1246 and miR-96 were significantly up-regulated in NSCLC patients. Interestingly, exosomal miR-96 in patients with radioresistant NSCLC was observed significantly up-regulated and was correlated with poor prognosis (99). Another study by Zuo et al. (101) showed down-regulation of plasma miR-216b in early and advanced SCC patients, what corresponded to the tissue results. Moreover, plasma miR-216b was negatively correlated with 18F-fluorodeoxyglucose (18F-FDG) uptake in NSCLC. According to the results, 18F-FDG could potentially serve as a predictor of therapeutic response in the application of miR-216b-based cancer treatment. In turn, up-regulation of plasma miR-202 and miR-26a was found in advanced NSCLC patients treated with platinum-based chemotherapy. High level of miR-202 was associated with disease progression and was shown as an independent prognostic factor for shorter progression-free and OS. Whilst miR-26a was correlated with shorter OS in SCC (105). Another study presented the role of plasma miR-200c and miR-34a advanced NSCLC treated with anti-PD1 immunotherapy. Over-expression of miR-200c was characterized as an independent prognostic factor for inferior OS in all NSCLC and most non-SCC patients. In turn, the low miR-34a level was connected with shorter OS in non-SCC (107). Marconi et al. first showed that combining exosomal plasma miRNA with peptidome might serve as an NSCLC biomarker to identify patients with a higher risk of recurrence after surgery. They investigated the role of plasma exosomal miR-130a-3p and fibrinopeptide A (FpA), and a significant correlation with DFS was observed in NSCLC patients. According to the authors, these findings may help predict

early-stage NSCLC patients predisposed to relapse after surgery (112). Pantano et al. (113) conducted a large-scale profiling of plasma extracellular miRNA vesicles. They characterized plasma EV-miR-625-5p as an independent biomarker of response and survival in NSCLC patients treated with ICIs, particularly those with the programmed death ligand -1 expression ≥ 50%.

5 The challenges of circulating miRNAs as cancer biomarkers

The discovery of miRNA molecules allowed science and medicine to expand the knowledge about their functions and mechanisms of action in the human body. MiRNAs were found to be differentially expressed in various tissue and cell types, which indicated their participation in many essential processes, such as proliferation or carcinogenesis (158). Many studies have indicated their potential utility as diseases biomarker, including cancer, but none of the miRNAs studied so far has been successfully introduced into routine clinical practice. A biomarker, by definition, is "objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention". Moreover, biomarkers should be non-invasive, easy to measure, detectable from multiple sources, and have high specificity, allowing for early diagnosis (159).

MiRNAs have many features of a promising biomarker, so extensive research into their potential utility in this area is well-funded. Unfortunately, many studies reveal contradictory results which may be caused i.a. by the lack of standardized procedures. They consist of the pre-analytical phase, including sample collection and preparation; the analytical phase, which determines the miRNA expression levels; and the post-analytical phase, handling data extraction and normalization (160). Collecting of material for research is a challenge, even for typical materials like serum and plasma. The concentration of serum miRNAs in higher than in plasma, which may indicate that the coagulation process may affect the change of the miRNA profile (161). The limitation in using plasma as a tested material is adding the heparin, which may cause inhibition of the Taq polymerase during the amplification (162).

The selection of patient groups also has an undoubted influence on the results. Apart from the disease state, miRNA expression can be affected by many other factors, such as diet, physical activity, and chemotherapy treatments. Many of these changes are difficult to assess, so it is essential to standardize the groups and the results should be confirmed on large independent cohorts. If several studies on the same miRNAs come from similar populations, selected groups may not have been sufficiently diverse (163). Dysregulation of miRNAs can often result not only from a disease state but also from the activity of the immune system (164). It could be the explanation for why the altered expression of the same miRNAs is observed in many studies. An important aspect is selecting an appropriate large group to obtain sufficient study power. The possibility of getting random results can be reduced by eliminating bias in the study design using techniques such as randomization and blinding (165).

Many methods and kits are used to isolate miRNAs, which makes comparing results less reliable. Moreover, miRNA detection can be affected by inhibitors of the qPCR, which may be present after isolation procedure (166). Mestdagh et al. compared nine different platforms for measuring serum miRNA expression. The high accuracy with a strong sensitivity was observed in the qPCR method. These findings indicate the need to confirm the results with at least two methods (167).

Despite its many limitations, miRNA has considerable potential to become a biomarker for many diseases, including NSCLC. Above all, developing standard operating procedures at each stage of miRNA processing can minimize the disadvantages of using miRNA as a biomarker.

6 Concluding remarks

The latest studies searching for biomarkers for NSCLC in body fluids have focused mainly on serum and plasma. A highly interesting and still not very popular direction in the search for lung cancer biomarkers is evaluating the expression of miRNAs isolated from patients' sputum. However, other biofluids, such as urine and saliva, should be considered in future research. We found only one study which proved that urine exosomal miR-4466 levels in lung cancer patients could be a biomarker for predicting an increased risk of metastatic disease among smokers (168). Interestingly, many of the studies we have cited revealed that miRNA levels are not significantly related to age, gender, tumour characteristics and tumour size (56, 66, 86, 115, 145). These observations suggest the potential usefulness of miRNAs in screening, regardless of the patient's age or gender. It still needs to be determined, which biofluid provides the best chance of finding accurate NSCLC biomarker, especially the research results are inconsistent.

MiRNAs play a significant role in many processes leading to the development of NSCLC and its consequences. The studies above proved that these particles have many features of an ideal potential biomarker for molecular diagnosis, prognosis, and monitoring of NSCLC. Their use provides a promising alternative to other diagnostic methods, which are invasive and often imprecise. Nevertheless, there are also some limitations in miRNAs reliability for practical clinical application. For instance, dysregulated levels of the same miRNAs are observed in different types of cancer (169).

Among the research papers described in this review, we selected miRNAs with the highest diagnostic value and presented them in Table 4. Several studies indicate that miRNAs combined into panels or combined with an additional parameter, such as CEA and CYFRA21-1, had a higher diagnostic value than these biomarkers analyzed separately. These miRNAs and panels are worth considering in future studies to improve the diagnosis of NSCLC. Unfortunately, many studies did not have a validation group, and the authors only used an unbiased screen and separate discovery and validation cohorts in three studies. Most of authors used qPCR/qRT-PCR in their studies, which is a valid approach for testing

miRNAs. However, researchers should also confirm their findings with another available method (167).

Our review suggest that miRNA extracted through the use of liquid biopsy appears to be an ideal biomarker for the diagnosis, prognosis and prediction of NSCLC. However, researchers should be aware of the numerous challenges of using miRNAs in these applications. Scientists should aim to develop standard operating procedures for material collection, storage, choice of detection platform, analysis of miRNA profile, and data normalization (166, 170). All of those factors may have an influence to the levels of circulating miRNAs. Moreover, levels of miRNAs may depend on age, gender and previous therapies used by patients, causing misinterpretation of the results (171). The most sensitive and high-throughput miRNA examing methods, such as Next Generation Sequencing, digital PCR and Nanostring Technology, prove to be invaluable in dispelling many doubts about the usefulness of some miRNAs as NSCLC and other disease biomarkers. To avoid false positive results, it is worth finding miRNAs that differ as much as possible in the expression level between NSCLC patients and healthy individuals (172).

Researchers should constantly strive to standardize and normalize the sampling, storage, isolation and subsequent detection of miRNAs to avoid misleading findings. Therefore, there is a need to validate all promising results on larger groups of patients before miRNAs begin to be used in diagnostic and clinical practice. The new direction of current and recent research on miRNAs in the context of NSCLC and other cancers is an important target for future research.

Author contributions

Conceptualization, JR, MN, AB and AK; supervision, MN and AK; writing—original draft preparation, JR; writing—review and editing, AB, MN and AK. All authors contributed to the article and approved the submitted version.

Funding

The study was supported by the funds of the Ministry of Education and Science within the project “Excellence Initiative - Research University”.

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

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. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. *Nature* (2018) 553(7689):446–54. doi: 10.1038/nature25183

3. Tsim S, O’Dowd CA, Milroy R, Davidson S. Staging of non-small cell lung cancer (NSCLC): a review. *Respir Med* (2010) 104(12):1767–74. doi: 10.1016/j.rmed.2010.08.005

4. Goebel C, Louden CL, McKenna R, Onugha O, Wachtel A, Long T. Diagnosis of non-small cell lung cancer for early stage asymptomatic patients. *Cancer Genomics Proteomics* (2019) 16(4):229–44. doi: 10.21873/cgp.20128

5. Alberg AJ, Ford JG, Samet JM. Epidemiology of lung cancer: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest* (2007) 132(3 SUPPL.):29S–55S. doi: 10.1378/chest.07-1347

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

7. Duma N, Santana-Davila R, Molina JR. Non–small cell lung cancer: epidemiology, screening, diagnosis, and treatment. *Mayo Clin Proc* (2019) 94 (8):1623–40. doi: 10.1016/j.mayocp.2019.01.013

8. Alexander M, Kim SY, Cheng H. Update 2020: management of non-small cell lung cancer. *Lung* (2020) 198(6):897–907. doi: 10.1007/s00408-020-00407-5

9. Osmani L, Askin F, Gabrielson E, Li QK. Current WHO guidelines and the critical role of immunohistochemical markers in the subclassification of non-small cell lung carcinoma (NSCLC): moving from targeted therapy to immunotherapy. *Semin Cancer Biol* (2018) 52(Pt 1):103–9. doi: 10.1016/j.semcancer.2017.11.019

10. Rossi G, Pelosi G, Barbareschi M, Graziano P, Cavazza A, Papotti M. Subtyping non-small cell lung cancer: relevant issues and operative recommendations for the best pathology practice. *Int J Surg Pathol* (2013) 21(4):326–36. doi: 10.1177/1066896913489346

11. Villalobos P, Wistuba II. Lung cancer biomarkers guideline. *Hematol Oncol Clin* (2017) 31(1):13–29. doi: 10.1016/j.hoc.2016.08.006

12. Millar AA, Waterhouse PM. Plant and animal microRNAs: similarities and differences. *Funct Integr Genomics* (2005) 5(3):129–35. doi: 10.1007/s10142-005-0145-2

13. Huang W. MicroRNAs: Biomarkers, diagnostics, and therapeutics. *Bioinf MicroRNA Res* (2017) 1617:21. doi: 10.1007/978-1-4939-7046-9

14. Lytle JR, Yario TA, Steitz JA. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5′ UTR as in the 3′ UTR. *Proc Natl Acad Sci USA* (2007) 104(23):9667–72. doi: 10.1073/pnas.0703820104

15. Ha M, Kim VN. Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol* (2014) 15(8):509–24. doi: 10.1038/nrm3838

16. Budakoti M, Panwar AS, Molpa D, Singh RK, Büsselberg D, Mishra AP, et al. Micro-RNA: the darkhorse of cancer. *Cell Signal* (2021) 83(March):109995. doi: 10.1016/j.cellsig.2021.109995

17. Felekkis K, Touvana E, Stefanou C, Deltas C. MicroRNAs: a newly described class of encoded molecules that play a role in health and disease. *Hippokratia* (2010) 14(4):236–40.

18. Lujambio A, Calin GA, Villanueva A, Ropero S, Sánchez-Céspedes M, Blanco D, et al. A microRNA DNA methylation signature for human. *Proc Natl Acad Sci USA* (2008) 105(36):13556–61. doi: 10.1073/pnas.0803055105

19. Hill M, Tran N. miRNA interplay: mechanisms and consequences in cancer. *Dis Model Mech* (2021) 14(4):dmm047662. doi: 10.1242/dmm.047662

20. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, et al. Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci USA* (2002) 99(24):15524–9. doi: 10.1073/pnas.242606799

21. Tiscornia G, Izpisúa Belmonte JC. MicroRNAs in embryonic stem cell function and fate. *Genes Dev* (2010) 24(24):2732–41. doi: 10.1101/gad.1982910

22. Weber JA, Baxter DH, Zhang S, Huang DY, Huang KH, Lee MJ, et al. The microRNA spectrum in 12 body fluids. *Clin Chem* (2010) 56(11):1733–41. doi: 10.1373/clinchem.2010.147405

23. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, et al. Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci USA* (2008) 105(30):10513–8. doi: 10.1073/pnas.0804549105

24. Vickers KC, Palmisano BT, Shoucri BM, Shamburek RD, Remaley AT, Mi, et al. MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. *Nat Cell Biol* (2011) 13:423–33. doi: 10.1038/ncb2210

25. He B, Zhao Z, Cai Q, Zhang Y, Zhang P, Shi S, et al. Mirna-based biomarkers, therapies, and resistance in cancer. *Int J Biol Sci* (2020) 16(14):2628–47. doi: 10.7150/ijbs.47203

26. Zhang J, Ma L. MicroRNA control of epithelial-mesenchymal transition. *Cancer Metastasis Rev* (2013) 31(0):653–62. doi: 10.1007/s10555-012-9368-6

27. Schwarzenbach H, Nishida N, Calin GA, Pantel K. Clinical relevance of circulating cell-free microRNAs in cancer. *Nat Rev | Clin Oncol* (2014) 11:145–56. doi: 10.1038/nrclinonc.2014.5

28. Hu G, Drescher KM, Chen XM. Exosomal miRNAs: biological properties and therapeutic potential. *Front Genet* (2012) 3(APR):1–9. doi: 10.3389/fgene.2012.00056

29. Wesołowska A, Piwocka K. Egzosomalne mikroRNA jako element komunikacji międzykomórkowej w nowotworach. *Postepy Biochem* (2017) 63(2):110–8.

30. Świtlik WZ, Szemraj J. Circulating miRNAs as non-invasive biomarkers for non-small cell lung cancer diagnosis, prognosis and prediction of treatment response. *Postepy Hig Med Dosw (Online)* (2017) 71:649–62. doi: 10.5604/01.3001.0010.3845

31. Świtlik WZ, Karbownik MS, Suwalski M, Kozak J, Szemraj J. Serum miR-210-3p as a potential noninvasive biomarker of lung adenocarcinoma: a preliminary study. *Genet Test Mol Biomarkers* (2019) 23(5):353–8.

32. Croce CM. 27 causes and consequences of microRNA dysregulation in cancer. *Eur J Cancer Suppl* (2010) 8(5):8. doi: 10.1016/S1359-6349(10)70836-3

33. Acunzo M, Romano G, Wernicke D, Croce CM. MicroRNA and cancer - a brief overview. *Adv Biol Regul* (2015) 57:1–9. doi: 10.1016/j.jbior.2014.09.013

34. Hong LZ, Zhou L, Zou R, Khoo CM, Chew ALS, Chin CL, et al. Systematic evaluation of multiple qPCR platforms, NanoString and miRNA-seq for microRNA biomarker discovery in human biofluids. *Sci Rep* (2021) 11(1):1–11. doi: 10.1038/s41598-021-83365-z

35. Kozomara A, Birgaoanu M, Griffiths-Jones S. MiRBase: from microRNA sequences to function. *Nucleic Acids Res* (2019) 47(D1):D155–62. doi: 10.1093/nar/gky1141

36. Brennecke J, Stark A, Russell RB, Cohen SM. Principles of microRNA-target recognition. *PloS Biol* (2005) 3(3):0404–18. doi: 10.1371/journal.pbio.0030085

37. Hayes J, Peruzzi PP, Lawler S. MicroRNAs in cancer: biomarkers, functions and therapy. *Trends Mol Med* (2014) 20(8):460–9. doi: 10.1016/j.molmed.2014.06.005

38. Mendell JT. miRiad roles for the miR-17-92 cluster in development and disease. *Cell* (2008) 133(2):217–22. doi: 10.1016/j.cell.2008.04.001

39. Yiu AJ, Yiu CY. Biomarkers in Colorectal Cancer. *Anticancer Res* (2016) 36 (3):1093–338.

40. Hu ZQ, Zhou SL, Li J, Zhou ZJ, Wang PC, Xin HY, et al. Circular RNA sequencing identifies CircASAP1 as a key regulator in hepatocellular carcinoma metastasis. *Hepatology* (2020) 72(3):906–22. doi: 10.1002/hep.31068

41. Ghafouri-Fard S, Shoorei H, Taheri M. miRNA profile in ovarian cancer. *Exp Mol Pathol* (2020) 113. doi: 10.1016/j.yexmp.2020.104381

42. Sharma N, Baruah MM. The microRNA signatures: aberrantly expressed miRNAs in prostate cancer. *Clin Transl Oncol* (2019) 21(2):126–44. doi: 10.1007/s12094-018-1910-8

43. Lee YR, Kim G, Tak WY, Jang SY, Kweon YO, Park JG, et al. Circulating exosomal noncoding RNAs as prognostic biomarkers in human hepatocellular carcinoma. *Int J Cancer* (2019) 144(6):1444–52. doi: 10.1002/ijc.31931

44. Lin PY, Yu SL, Yang PC. MicroRNA in lung cancer. *Br J Cancer* (2010) 103 (8):1144–8. doi: 10.1038/sj.bjc.6605901

45. Bica-Pop C, Cojocneanu-Petric R, Magdo L, Raduly L, Gulei D, Berindan-Neagoe I. Overview upon miR-21 in lung cancer: focus on NSCLC. *Cell Mol Life Sci* (2018) 75(19):3539–51. doi: 10.1007/s00018-018-2877-x

46. Pfeffer SR, Yang CH, Pfeffer LM. The role of MIR-21 in cancer. *Drug Dev Res* (2015) 76(6):270–7. doi: 10.1002/ddr.21257

47. Li D, Zhang J, Li J. Role of miRNA sponges in hepatocellular carcinoma. *Clin Chim Acta* (2020) 500:10–9. doi: 10.1016/j.cca.2019.09.013

48. Shin VY, Chu KM. MiRNA as potential biomarkers and therapeutic targets for gastric cancer. *World J Gastroenterol* (2014) 20(30):10432–9. doi: 10.3748/wjg.v20.i30.10432

49. Abramovic I, Ulamec M, Bojanac Katusic A, Bulic-Jakus F, Jezek D, Sincic N. MiRNA in prostate cancer: challenges toward translation. *Epigenomics* (2020) 12 (6):543–58. doi: 10.2217/epi-2019-0275

50. Yu T, Liu L, Li J, Yan M, Lin H, Liu Y, et al. MiRNA-10a is upregulated in NSCLC and may promote cancer by targeting PTEN. *Oncotarget* (2015) 6(30):30239–50. doi: 10.18632/oncotarget.4972

51. Markou A, Sourvinou I, Vorkas PA, Yousef GM, Lianidou E. Clinical evaluation of microRNA expression profiling in non small cell lung cancer. *Lung Cancer* (2013) 81 (3):388–96. doi: 10.1016/j.lungcan.2013.05.007

52. Tan T, Xu XH, Lu XH, Wang XW. MiRNA-200a-3p suppresses the proliferation, migration and invasion of non-small cell lung cancer through targeting IRS2. *Eur Rev Med Pharmacol Sci* (2020) 24(2):712–20. doi: 10.26355/eurrev_202001_20050

53. Yu J-F, Sui X-J. MiRNA-30 play an important role in non-small cell lung cancer (NSCLC). *Clin Lab* (2020) 66(4):529–35. doi: 10.7754/Clin.Lab.2019.190712

54. Huang D, Qu D. Early diagnostic and prognostic value of serum exosomal miR-1246 in non-small cell lung cancer. *Int J Clin Exp Pathol* (2020) 13(7):1601–7.

55. Tang Y, Zhang Z, Song X, Yu M, Niu L, Zhao Y, et al. Tumor-derived exosomal miR-620 as a diagnostic biomarker in non-small-cell lung cancer. *J Oncol* (2020) 2020. doi: 10.1155/2020/6691211

56. Chen L, Li Y, Lu J. Identification of circulating miR-762 as a novel diagnostic and prognostic biomarker for non-small cell lung cancer. *Technol Cancer Res Treat* (2020) 19:1–9. doi: 10.1177/1533033820964222

57. Wang A, Zhang H, Wang J, Zhang S, Xu Z. MiR-519d targets HER3 and can be used as a potential serum biomarker for non-small cell lung cancer. *Aging (Albany NY)* (2020) 12(6):4866–78. doi: 10.18632/aging.102908

58. Zhang Z, Tang Y, Song X, Xie L, Zhao S, Song X. Tumor-derived exosomal miRNAs as diagnostic biomarkers in non-small cell lung cancer. *Front Oncol* (2020) 10 (October). doi: 10.3389/fonc.2020.560025

59. Liu M, Si Q, Ouyang S, Zhou Z, Wang M, Zhao C, et al. Serum MiR-4687-3p has potential for diagnosis and carcinogenesis in non-small cell lung cancer. *Front Genet* (2020) 11(November):1–10. doi: 10.3389/fgene.2020.597508

60. Tao S, Ju X, Zhou H, Zeng Q. Circulating microRNA-145 as a diagnostic biomarker for non-small-cell lung cancer: a systemic review and meta-analysis. *Int J Biol Markers* (2020) 35(4):51–60. doi: 10.1177/1724600820967124

61. Zhou C, Chen Z, Zhao L, Zhao W, Zhu Y, Liu J, et al. A novel circulating miRNA-based signature for the early diagnosis and prognosis prediction of non–small-cell lung cancer. *J Clin Lab Anal* (2020) 34(11):1–9. doi: 10.1002/jcla.23505

62. Kumar S, Sharawat SK, Ali A, Gaur V, Malik PS, Kumar S, et al. Identification of differentially expressed circulating serum microRNA for the diagnosis and prognosis of Indian non–small cell lung cancer patients. *Curr Probl Cancer* (2020) 44(4):100540. doi: 10.1016/j.currproblcancer.2020.100540

63. Liu K, Chen H, You Q, Ye Q, Wang F, Wang S, et al. miR-145 inhibits human non-small-cell lung cancer growth by dual-targeting RIOK2 and NOB1. *Int J Oncol* (2018) 53(1):257–65. doi: 10.3892/ijo.2018.4393

64. Zhang ZJ, Song XG, Xie L, Wang KY, Tang YY, Yu M, et al. Circulating serum exosomal miR-20b-5p and miR-3187-5p as efficient diagnostic biomarkers for early-stage non-small cell lung cancer. *Exp Biol Med* (2020) 245(16):1428–36. doi: 10.1177/1535370220945987

65. Yang C, Jia X, Zhou J, Sun Q, Ma Z. The MiR-17-92 gene cluster is a blood-based marker for cancer detection in non-Small-Cell lung cancer. *Am J Med Sci* (2020) 360 (3):248–60. doi: 10.1016/j.amjms.2020.05.004

66. Zhang X, Hu Y, Gong C, Zhang C. Overexpression of miR-518b in non-small cell lung cancer serves as a biomarker and facilitates tumor cell proliferation, migration and invasion. *Oncol Lett* (2020) 20(2):1213–20. doi: 10.3892/ol.2020.11667

67. Zhang Y, Xu H. Serum exosomal miR-378 upregulation is associated with poor prognosis in non–small-cell lung cancer patients. *J Clin Lab Anal* (2020) 34(6):1–7. doi: 10.1002/jcla.23237

68. Liu J, Han Y, Liu X, Wei S. Serum miR-185 is a diagnostic and prognostic biomarker for non-small cell lung cancer. *Technol Cancer Res Treat* (2020) 19:1–7. doi: 10.1177/1533033820973276

69. Xue WX, Zhang MY, Rui Li X, Yin YH, Qu YQ. Serum miR-1228-3p and miR-181a-5p as Noninvasive biomarkers for non-small cell lung cancer diagnosis and prognosis. *BioMed Res Int* (2020) 2020:9601876. doi: 10.1155/2020/9601876

70. Kumar S, Sharawat SK, Ali A, Gaur V, Malik PS, Pandey M. Differential expression of circulating serum miR-1249-3p, miR-3195, and miR-3692-3p in non-small cell lung cancer. *Hum Cell* (2020) 33(3):839–49. doi: 10.1007/s13577-020-00351-9

71. Wang X, Jiang X, Li J, Wang J, Binang H, Shi S, et al. Serum exosomal miR-1269a serves as a diagnostic marker and plays an oncogenic role in non-small cell lung cancer. *Thorac Cancer* (2020) 11(12):3436–47. doi: 10.1111/1759-7714.13644

72. Xie H, Chen J, Lv X, Zhang L, Wu J, Ge X, et al. Clinical value of serum and exhaled breath condensate miR-186 and IL-1β levels in non-small cell lung cancer. *Technol Cancer Res Treat* (2020) 19:1–8. doi: 10.1177/1533033820947490

73. Lv J, An J, Zhang YD, Li ZX, Zhao GL, Gao J, et al. A three serum miRNA panel as diagnostic biomarkers of radiotherapy-related metastasis in non-small cell lung cancer. *Oncol Lett* (2020) 20(5):1–11. doi: 10.3892/ol.2020.12099

74. Yang G, Wang T, Qu X, Chen S, Han Z, Chen S, et al. Exosomal miR-21/Let-7a ratio distinguishes non-small cell lung cancer from benign pulmonary diseases. *Asia Pac J Clin Oncol* (2020) 16(4):280–6. doi: 10.1111/ajco.13343

75. Xu S, Zheng L, Kang L, Xu H, Gao L. microRNA-let-7e in serum-derived exosomes inhibits the metastasis of non-small-cell lung cancer in a SUV39H2/LSD1/CDH1-dependent manner. *Cancer Gene Ther* (2020) 28(3-4):250–64. doi: 10.1038/s41417-021-00402-9

76. Wu Q, Yu L, Lin X, Zheng Q, Zhang S, Chen D, et al. Combination of serum miRNAs with serum exosomal miRNAs in early diagnosis for non-small-cell lung cancer. *Cancer Manage Res* (2020) 12:485–95. doi: 10.2147/CMAR.S232383

77. Ding H, Wen W, Ding Q, Zhao X. Diagnostic valuation of serum miR-184 and miR-191 in patients with non-Small-Cell lung cancer. *Cancer Control* (2020) 27(1):1–9. doi: 10.1177/1073274820964783

78. Ying L, Du L, Zou R, Shi L, Zhang N, Jin J, et al. Development of a serum miRNA panel for detection of early stage non-small cell lung cancer. *Proc Natl Acad Sci USA* (2020) 117(40):25036–42. doi: 10.1073/pnas.2006212117

79. Shao L, Lu X, Zhou Y, Wang Y, Wang X, Zhuang Z, et al. Altered miR-93-5p/miR-18a expression in serum for diagnosing non-small cell lung cancer. *Am J Transl Res* (2021) 13(5):5073–9.

80. Luo R, Liu H, Chen J. Reduced circulating exosomal miR-382 predicts unfavorable outcome in non-small cell lung cancer. *Int J Clin Exp Pathol* (2021) 14 (4):469–74.

81. Wan N, Zheng J. MicroRNA-891a-5p is a novel biomarker for non-small cell lung cancer and targets HOXA5 to regulate tumor cell biological function. *Oncol Lett* (2021) 22:1. doi: 10.3892/ol.2021.12768

82. Kryczka J, Migdalska-Sęk M, Kordiak J, Kiszałkiewicz JM, Pastuszak-Lewandoska D, Antczak A, et al. Serum extracellular vesicle-derived mirnas in patients with non-small cell lung cancer–search for non-invasive diagnostic biomarkers. *Diagnostics* (2021) 11:3. doi: 10.3390/diagnostics11030425

83. Chen R, Qian Z, Xu X, Zhang C, Niu Y, Wang Z, et al. Exosomes-transmitted miR-7 reverses gefitinib resistance by targeting YAP in non-small-cell lung cancer. *Pharmacol Res* (2021) 165(January):105442. doi: 10.1016/j.phrs.2021.105442

84. Liu F, Li T, Hu P, Dai L. Upregulation of serum miR-629 predicts poor prognosis for non- small-cell lung cancer. *Dis Markers* (2021) 2021:8819934. doi: 10.1155/2021/8819934

85. Wang T, Li W, Li H, Li W. Dysregulation of exosomal miR-192 and miR-194 expression in lung adenocarcinoma patients. *Saudi J Biol Sci* (2021) 28(3):1561–8. doi: 10.1016/j.sjbs.2021.01.013

86. Li X, Chen C, Wang Z, Liu J, Sun W, Shen K, et al. Elevated exosome-derived miRNAs predict osimertinib resistance in non-small cell lung cancer. *Cancer Cell Int* (2021) 21(1):1–14. doi: 10.1186/s12935-021-02075-8

87. Xia J, Luo M, Dai L, Wang L, Wang L, Zhu J. Serum exosomal microRNAs as predictive markers for EGFR mutations in non–small-cell lung cancer. *J Clin Lab Anal* (2021) 35(5):1–12. doi: 10.1002/jcla.23743

88. Zhao YL, Zhang JX, Yang JJ, Wei YB, Peng JF, Fu CJ, et al. MiR-205-5p promotes lung cancer progression and is valuable for the diagnosis of lung cancer. *Thorac Cancer* (2022) 13(6):832–43. doi: 10.1111/1759-7714.14331

89. Liu B, Zhang R, Zhu Y, Hao R. Exosome-derived microRNA-433 inhibits tumorigenesis through incremental infiltration of CD4 and CD8 cells in non-small cell lung cancer. *Oncol Lett* (2021) 22(2):1–12. doi: 10.3892/ol.2021.12868

90. Wu J, Han X, Yang X, Li Y, Liang Y, Sun G, et al. MiR-138-5p suppresses the progression of lung cancer by targeting SNIP1. *Thorac Cancer* (2022) 2023:1–12. doi: 10.1111/1759-7714.14791

91. Zeng X, Liu D, Peng G, Liu J, Yang H. MiroRNA-31-3p promotes the invasion and metastasis of non-Small-Cell lung cancer cells by targeting forkhead box 1 (FOXO1). *Comput Math Methods Med* (2022) 2022. doi: 10.1155/2022/4597087

92. Jiang X, Yuan Y, Tang L, Wang J, Zhang D, Cho WC, et al. Identification and validation prognostic impact of MiRNA-30a-5p in lung adenocarcinoma. *Front Oncol* (2022) 12:1–16. doi: 10.3389/fonc.2022.831997

93. Khandelwal A, Seam RK, Gupta M, Rana MK, Prakash H, Vasquez KM, et al. Circulating microRNA-590-5p functions as a liquid biopsy marker in non-small cell lung cancer. *Cancer Sci* (2020) 111(3):826–39. doi: 10.1111/cas.14199

94. Zhao X, Li M, Dai X, Yang Y, Peng Y, Xu C, et al. Downregulation of exosomal miR-1273a increases cisplatin resistance of non-small cell lung cancer by upregulating the expression of syndecan binding protein. *Oncol Rep* (2020) 44(5):2165–73. doi: 10.3892/or.2020.7753

95. Peng XX, Yu R, Wu X, Wu SY, Pi C, Chen ZH, et al. Correlation of plasma exosomal microRNAs with the efficacy of immunotherapy in EGFR/ALK wild-type advanced non-small cell lung cancer. *J Immunother Cancer* (2020) 8(1):1–7. doi: 10.1136/jitc-2019-000376

96. Papadaki C, Monastirioti A, Rounis K, Makrakis D, Kalbakis K, Nikolaou C, et al. Circulating microRNAs regulating DNA damage response and responsiveness to cisplatin in the prognosis of patients with non-small cell lung cancer treated with first-line platinum chemotherapy. *Cancers (Basel)* (2020) 12:5. doi: 10.3390/cancers12051282

97. Dong X, Chang M, Song X, Ding S, Xie L, Song X. Plasma miR-1247-5p, miR-301b-3p and miR-105-5p as potential biomarkers for early diagnosis of non-small cell lung cancer. *Thorac Cancer* (2021) 12(4):539–48. doi: 10.1111/1759-7714.13800

98. Leonetti A, Capula M, Minari R, Mazzaschi G, Gregori A, El Hassouni B, et al. Dynamic evaluation of circulating mirna profile in egfr- mutated nsclc patients treated with egfr-tkis. *Cells* (2021) 10(6). doi: 10.3390/cells10061520

99. Zheng Q, Ding H, Wang L, Yan Y, Wan Y, Yi Y, et al. Circulating exosomal miR-96 as a novel biomarker for radioresistant non-Small-Cell lung cancer. *J Oncol* (2021) 2021. doi: 10.1155/2021/5893981

100. Wu Y, Yin Q, Zhou YL, He L, Zou ZQ, Dai XY, et al. Evaluation of microRNAs as potential biomarkers in circulating HPV-DNA-positive non-small cell lung cancer patients. *Cancer Biol Ther* (2021) 22(2):136–48. doi: 10.1080/15384047.2021.1872155

101. Zuo M, Yao L, Wen L, Shen J, Zhang N, Bai T, et al. The expression of miRNA-216b is negatively correlated with 18F-FDG uptake in non-small cell lung cancer. *World J Surg Oncol* (2021) 19(1):1–10. doi: 10.1186/s12957-021-02376-2

102. Khandelwal A, Sharma U, Barwal TS, Seam RK, Gupta M, Rana M, et al. Circulating miR-320a acts as a tumor suppressor and prognostic factor in non-small cell lung cancer. *Front Oncol* (2021) 11. doi: 10.3389/fonc.2021.645475

103. Jiang HG, Dai CH, Xu YP, Jiang Q, Xia XB, Shu Y, et al. Four plasma miRNAs act as biomarkers for diagnosis and prognosis of non-small cell lung cancer. *Oncol Lett* (2021) 22(5):1–12. doi: 10.3892/ol.2021.13053

104. Yang XR, Pi C, Yu R, Fan XJ, Peng XX, Zhang XC, et al. Correlation of exosomal microRNA clusters with bone metastasis in non-small cell lung cancer. *Clin Exp Metastasis* (2021) 38(1):109–17. doi: 10.1007/s10585-020-10062-y

105. Monastirioti A, Papadaki C, Rounis K, Kalapanida D, Mavroudis D, Agelaki S. A prognostic role for circulating micrornas involved in macrophage polarization in advanced non-small cell lung cancer. *Cells* (2021) 10:8. doi: 10.3390/cells10081988

106. Kim DH, Park H, Choi YJ, Kang MH, Kim TK, Pack CG, et al. Exosomal miR-1260b derived from non-small cell lung cancer promotes tumor metastasis through the inhibition of HIPK2. *Cell Death Dis* (2021) 12(8). doi: 10.1038/s41419-021-04024-9

107. Monastirioti A, Papadaki C, Kalapanida D, Rounis K, Michaelidou K, Papadaki MA, et al. Plasma-based microRNA expression analysis in advanced stage NSCLC patients treated with nivolumab. *Cancers (Basel)* (2022) 14(19):1–18. doi: 10.3390/cancers14194739

108. Jiang YF, Wei SN, Geng N, Qin WW, He X, Wang XH, et al. Evaluation of circulating small extracellular vesicle-derived miRNAs as diagnostic biomarkers for differentiating between different pathological types of early lung cancer. *Sci Rep* (2022) 12(1):1–14. doi: 10.1038/s41598-022-22194-0

109. Wu Y, Jing H, Zhang J. MicroRNA-340 and MicroRNA-450b-5p: plasma biomarkers for detection of non-Small-Cell lung cancer. *J Environ Public Health* (2022) 2022:8024700. doi: 10.1155/2022/8024700

110. Vadla GP, Daghat B, Patterson N, Ahmad V, Perez G, Garcia A, et al. Combining plasma extracellular vesicle let-7b-5p, miR-184 and circulating miR-22-3p levels for NSCLC diagnosis and drug resistance prediction. *Sci Rep* (2022) 12(1):1–12. doi: 10.1038/s41598-022-10598-x

111. Haranguş A, Lajos R, Budisan L, Zanoaga O, Ciocan C, Bica C, et al. Identification of potential microRNA panels for Male non-small cell lung cancer identification using microarray datasets and bioinformatics methods. *J Pers Med* (2022) 12:12. doi: 10.3390/jpm12122056

112. Marconi S, Croce M, Chiorino G, Rossi G, Guana F, Profumo A, et al. A circulating risk score, based on combined expression of exo-miR-130a-3p and fibrinopeptide a, as predictive biomarker of relapse in resectable non-small cell lung cancer patients. *Cancers (Basel)* (2022) 14:14. doi: 10.3390/cancers14143412

113. Pantano F, Zalfa F, Iuliani M, Simonetti S, Manca P, Napolitano A, et al. Large–Scale profiling of extracellular vesicles identified miR–625–5p as a novel biomarker of immunotherapy response in advanced non–Small–Cell lung cancer patients. *Cancers (Basel)* (2022) 14:10. doi: 10.3390/cancers14102435

114. Su Y, Guarnera MA, Bin Fang H, Jiang F. Small non-coding RNA biomarkers in sputum for lung cancer diagnosis. *Mol Cancer* (2016) 15(1):1–4. doi: 10.1186/s12943-016-0520-8

115. Su J, Leng Q, Lin Y, Ma J, Jiang F, Lee CJ, et al. Integrating circulating immunological and sputum biomarkers for the early detection of lung cancer. *biomark Cancer* (2018) 10:1179299X1875929. doi: 10.1177/1179299X18759297

116. Bagheri A, Khorshid HRK, Tavallaie M, Mowla SJ, Sherafatian M, Rashidi M, et al. A panel of noncoding RNAs in non–small-cell lung cancer. *J Cell Biochem* (2019) 120(5):8280–90. doi: 10.1002/jcb.28111

117. Liao J, Shen J, Leng Q, Qin M, Zhan M, Jiang F. MicroRNA-based biomarkers for diagnosis of non-small cell lung cancer (NSCLC). *Thorac Cancer* (2020) 11(3):762–8. doi: 10.1111/1759-7714.13337

118. Lin Y, Holden V, Dhilipkannah P, Deepak J, Todd NW, Jiang F. A non-coding RNA landscape of bronchial epitheliums of lung cancer patients. *Biomedicines* (2020) 8 (4):1–15. doi: 10.3390/biomedicines8040088

119. Li N, Dhilipkannah P, Jiang F. High-throughput detection of multiple miRNAs and methylated DNA by droplet digital PCR. *J Pers Med* (2021) 11(5). doi: 10.3390/jpm11050359

120. Xing L, Su J, Guarnera MA, Zhang H, Cai L, Zhou R, et al. Sputum microRNA biomarkers for identifying lung cancer in indeterminate solitary pulmonary nodules. *Clin Cancer Res* (2015) 21(2):484–9. doi: 10.1158/1078-0432.CCR-14-1873

121. Volinia S, Galasso M, Costinean S, Tagliavini L, Gamberoni G, Drusco A, et al. Reprogramming of miRNA networks in cancer and leukemia. *Genome Res* (2010) 20 (5):589–99. doi: 10.1101/gr.098046.109

122. Li Y-Y, Shao J-P, Zhang S-P, Xing G-Q, Liu H-J. E-mail miR-519d-3p inhibits cell proliferation and invasion of gastric cancer by downregulating b-cell lymphoma 6. *Cytogenet Genome Res* (2018) 154:12–9. doi: 10.1159/000487372

123. Ye X, Lv H. MicroRNA-519d-3p inhibits cell proliferation and migration by targeting TROAP in colorectal cancer. *Biomed Pharmacother* (2018) 105(621):879–86. doi: 10.1016/j.biopha.2018.04.114

124. Su J, Liao J, Gao L, Shen J, Guarnera MA, Zhan M, et al. Analysis of small nucleolar RNAs in sputum for lung cancer diagnosis. *Oncotarget* (2016) 7(5):5131–42. doi: 10.18632/oncotarget.4219

125. Ye Z, Shen N, Weng Y, Li K, Hu L, Liao H, et al. Low miR-145 silenced by DNA methylation promotes NSCLC cell proliferation, migration and invasion by targeting mucin 1. *Cancer Biol Ther* (2015) 16(7):1071–9. doi: 10.1080/15384047.2015.1046024

126. Chen Z, Zeng H, Guo Y, Liu P, Pan H, Deng A, et al. MiRNA-145 inhibits non-small cell lung cancer cell proliferation by targeting c-myc. *J Exp Clin Cancer Res* (2010) 29(1):151. doi: 10.1186/1756-9966-29-151

127. Zhang Y, Zhang Y, Yin Y, Li S. Detection of circulating exosomal miR-17-5p serves as a novel non-invasive diagnostic marker for non-small cell lung cancer patients. *Pathol Res Pract* (2019) 215(8):152466. doi: 10.1016/j.prp.2019.152466

128. Ventura A, Young AG, Winslow MM, Lintault L, Meissner A, Erkeland SJ, et al. Targeted deletion reveals essential and overlapping functions of the miR-1792 family of miRNA clusters. *Cell* (2008) 132(5):875–86. doi: 10.1016/j.cell.2008.02.019

129. Zhou P, Ma L, Zhou J, Jiang M, Rao E, Zhao Y, et al. miR-17-92 plays an oncogenic role and conveys chemo-resistance to cisplatin in human prostate cancer cells. *Int J Oncol* (2016) 48(4):1737–48. doi: 10.3892/ijo.2016.3392

130. Liang C, Zhang X, Wang HM, Liu XM, Zhang XJ, Zheng B, et al. MicroRNA-18a-5p functions as an oncogene by directly targeting IRF2 in lung cancer. *Cell Death Dis* (2017) 8(5):1–10. doi: 10.1038/cddis.2017.145

131. Sekar D, Panagal M, Manimaran D, Ahmad N, Rao TN, Alomar SY. miR-185 and its anti-miR as a biomarker and therapeutic target for oral cancer. *Oral Oncol* (2022) 129:105873. doi: 10.1016/j.oraloncology.2022.105873

132. Li Y, Zhong W, Zhu M, Li M, Yang Z. MiR-185 inhibits prostate cancer angiogenesis induced by the nodal/ALK4 pathway. *BMC Urol* (2020) 20(1):1–11. doi: 10.1186/s12894-020-00617-2

133. Zhou CW, Zhao WJ, Zhu YG, Zhao XD. MiR-185 inhibits tumor growth and enhances chemo-resistance via targeting SRY-related high mobility group box transcription factor 13 in non-small-cell carcinoma. *Am J Transl Res* (2018) 10 (8):2600–9. doi: 1007/s00432-020-03127-8

134. Zhao L, Zhang Y, Liu J, Yin W, Jin D, Wang D, et al. MiR-185 inhibits the proliferation and invasion of non-small cell lung cancer by targeting KLF7. *Oncol Res* (2019) 27(9):1015–23. doi: 10.3727/096504018X15247341491655

135. Ferrari E, Gandellini P. Unveiling the ups and downs of miR-205 in physiology and cancer: transcriptional and post-transcriptional mechanisms. *Cell Death Dis* (2020) 11(11). doi: 10.1038/s41419-020-03192-4

136. Zhang R, Chen L, Cui YD, Li G. The association between human papillomavirus infection and smoking, age, gender in lung cancer patients: a meta-analysis. *Iran J Public Health* (2019) 48(1):1–8. doi: 10.18502/ijph.v48i1.777

137. Burd EM. Human papillomavirus and cervical cancer. *Clin Microbiol Rev* (2003) 16(1):1–17. doi: 10.1128/CMR.16.1.1-17.2003

138. Coppock JD, Lee JH. mTOR, metabolism, and the immune response in HPV-positive head and neck squamous cell cancer. *World J Otorhinolaryngol - Head Neck Surg* (2016) 2(2):76–83. doi: 10.1016/j.wjorl.2016.05.010

139. Wang H, Xu J, Ding L. MicroRNA-21 was a promising biomarker for lung carcinoma diagnosis: an update meta-analysis. *Thorac Cancer* (2022) 13(3):316–21. doi: 10.1111/1759-7714.14242

140. Belinsky SA, Liechty KC, Gentry FD, Wolf HJ, Rogers J, Vu K. Promoter hypermethylation of multiple genes in sputum precedes lung cancer incidence in a high-risk cohort. *Cancer Res* (2006) 66(6):3338–44. doi: 10.1158/0008-5472.CAN-05-3408

141. Thunnissen FBJM. Sputum examination for early detection of lung cancer. *J Clin Pathol* (2003) 56(11):805–10. doi: 10.1136/jcp.56.11.805

142. Nicholas B, Djukanović R. Induced sputum: a window to lung pathology. *Biochem Soc Trans* (2009) 37(4):868–72. doi: 10.1042/BST0370868

143. D'amato M, Iadarola P, Viglio S. Proteomic analysis of human sputum for the diagnosis of lung disorders: where are we today? *Int J Mol Sci* (2022) 23(10). doi: 10.3390/ijms23105692

144. Zhang X, Wang Q, Zhang S, Roever L. MicroRNAs in sputum specimen as noninvasive biomarkers for the diagnosis of nonsmall cell lung cancer: an updated meta-analysis. *Med (United States)* (2019) 98(6):1–9. doi: 10.1097/MD.0000000000014337

145. He RQ, Cen WL, Cen JM, Cen WN, Li JY, Li MW, et al. Clinical significance of miR-210 and its prospective signaling pathways in non-small cell lung cancer: evidence from gene expression omnibus and the cancer genome atlas data mining with 2763 samples and validation via real-time quantitative PCR. *Cell Physiol Biochem* (2018) 46 (3):925–52. doi: 10.1159/000488823

146. Shen J, Liao J, Guarnera MA, Fang H, Cai L, Stass SA, et al. Analysis of microRNAs in sputum to improve computed tomography for lung cancer diagnosis. *J Thorac Oncol* (2014) 9(1):33–40. doi: 10.1097/JTO.0000000000000025

147. Patterson GA. Lung cancer staging. *Chest* (1991) 100(2):520–2. doi: 10.1378/chest.100.2.520

148. Detterbeck FC, Boffa DJ, Tanoue LT. The new lung cancer staging system. *Chest* (2009) 136(1):260–71. doi: 10.1378/chest.08-0978

149. Salama JK, Vokes EE. New radiotherapy and chemoradiotherapy approaches for non-small-cell lung cancer. *J Clin Oncol* (2013) 31(8):1029–38. doi: 10.1200/JCO.2012.44.5064

150. Remon J, Hendriks LEL. Targeted therapies for unresectable stage III non-small cell lung cancer. *Mediastinum* (2021) 5(2):0–2. doi: 10.21037/med-21-8

151. Kris MG, Natale RB, Herbst RS, Lynch TJ Jr, Prager D, Belani CP, et al. Efficacy of gefitinib , an inhibitor of tyrosine kinase , in symptomatic patients with non-small cell lung cancer. *J Am Med Assoc* (2003) 290(16):2149–58. doi: 10.3390/cancers14194739

152. Remon J, Steuer CE, Ramalingam SS, Felip E. Osimertinib and other third-generation EGFR TKI in EGFR-mutant NSCLC patients. *Ann Oncol* (2018) 29 (February):i20–7. doi: 10.1093/annonc/mdx704

153. Rossi A, Di Maio M. Platinum-based chemotherapy in advanced non-small-cell lung cancer: optimal number of treatment cycles. *Expert Rev Anticancer Ther* (2016) 16(6):653–60. doi: 10.1586/14737140.2016.1170596

154. D'Addario G, Pintilie M, Leighl NB, Feld R, Cerny T, Shepherd FA. Platinum-based versus non-platinum-based chemotherapy in advanced non-small-cell lung cancer: a meta-analysis of the published literature. *J Clin Oncol* (2005) 23(13):2926–36. doi: 10.1200/JCO.2005.03.045

155. Wan P, Zou F, Zhang X, Li H, Dulak A, Tomko RJ Jr, et al. miR21 negatively regulates CDC25A and cell cylce progression in colpon cancer cells. *Cancer Res* (2009) 69(20):8157–65. doi: 10.1158/0008-5472.CAN-09-1996

156. Yang Z, Fang S, Di Y, Ying W, Tan Y, Gu W. Modulation of NF-κB/miR-21/PTEN pathway sensitizes non-small cell lung cancer to cisplatin. *PloS One* (2015) 10 (3):1–13. doi: 10.1371/journal.pone.0121547

157. Donzelli S, Fontemaggi G, Fazi F, Di Agostino S, Padula F, Biagioni F, et al. MicroRNA-128-2 targets the transcriptional repressor E2F5 enhancing mutant p53 gain of function. *Cell Death Differ* (2012) 19(6):1038–48. doi: 10.1038/cdd.2011.190

158. Krol J, Loedige I, Filipowicz W. The widespread regulation of microRNA biogenesis, function and decay. *Nat Rev Genet* (2010) 11(9):597–610. doi: 10.1038/nrg2843

159. Atkinson AJ, Colburn WA, DeGruttola VG, DeMets DL, Downing GJ, Hoth DF, et al. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* (2001) 69(3):89–95. doi: 10.1067/mcp.2001.113989

160. Schaffner F, Merlin J-L, von Bubnoff N. Circulating miRNAs as Biomarker in Cancer. *Tumor liquid biopsies* (Springer:Circulating miRNAs as Biomarker in Cancer) (2019) 215:277–98.. doi: 10.1007/978-3-030-26439-0

161. Wang K, Yuan Y, Cho J-H, Mcclarty S, Baxter D. Comparing the MicroRNA spectrum between serum and plasma. *PloS One* (2012) 7(7):41561. doi: 10.1371/journal.pone.0041561

162. Huggett JF, Novak T, Garson JA, Green C, Morris-Jones SD, Miller RF, et al. Differential susceptibility of PCR reactions to inhibitors: an important and unrecognised phenomenon. *BMC Res Notes* (2008) 1:1–9. doi: 10.1186/1756-0500-1-70

163. Tiberio P, Callari M, Angeloni V, Daidone MG, Appierto V. Challenges in using circulating miRNAs as cancer biomarkers. *BioMed Res Int* (2015) 2015:731479. doi: 10.1155/2015/731479

164. Jarry J, Schadendorf D, Greenwood C, Spatz A, van Kempen LC. The validity of circulating microRNAs in oncology: five years of challenges and contradictions. *Mol Oncol* (2014) 8(4):819–29. doi: 10.1016/j.molonc.2014.02.009

165. Jones SR, Carley S, Harrison M. An introduction to power and sample size estimation. *Emerg Med J* (2003) 20:453–8. doi: 10.1136/emj.20.5.453

166. Ralla B, Stephan C, Meller S, Dietrich D, Kristiansen G, Jung K. Nucleic acid-based biomarkers in body fluids of patients with urologic malignancies. *Crit Rev Clin Lab Sci* (2014) 51(4):200–31. doi: 10.3109/10408363.2014.914888

167. Mestdagh P, Hartmann N, Baeriswyl L, Andreasen D, Bernard N, Chen C, et al. Evaluation of quantitative mirnA expression platforms in the micrornA quality control (mirQC) study. *Nat Methods* (2014) 11(8):809–15. doi: 10.1038/nmeth.3014

168. Tyagi A, Wu SY, Sharma S, Wu K, Zhao D, Deshpande R, et al. Exosomal miR-4466 from nicotine-activated neutrophils promotes tumor cell stemness and metabolism in lung cancer metastasis. *Oncogene* (2022) 41(22):3079–92. doi: 10.1038/s41388-022-02322-w

169. Condrat CE, Thompson DC, Barbu MG, Bugnar OL, Boboc A, Cretoiu D, et al. miRNAs as biomarkers in disease: latest findings regarding their role in diagnosis and prognosis. *Cells* (2020) 9(2):1–33. doi: 10.3390/cells9020276

170. Preethi KA, Selvakumar SC, Ross K, Jayaraman S, Tusubira D, Sekar D. Liquid biopsy: exosomal microRNAs as novel diagnostic and prognostic biomarkers in cancer. *Mol Cancer* (2022) 21(1):1–15. doi: 10.1186/s12943-022-01525-9

171. Takizawa S, Matsuzaki J, Ochiya T. Circulating microRNAs: challenges with their use as liquid biopsy biomarkers. *Cancer Biomarkers* (2022) 35(1):1–9. doi: 10.3233/CBM-210223

172. Wang H, Peng R, Wang J, Qin Z, Xue L. Circulating microRNAs as potential cancer biomarkers: the advantage and disadvantage. *Clin Epigenet* (2018) 10(1):1–10. doi: 10.1186/s13148-018-0492-1

frontiers | Frontiers in Oncology

TYPE Systematic Review
PUBLISHED 26 July 2023
DOI 10.3389/fonc.2023.1201713

Check for updates

OPEN ACCESS

EDITED BY
Hongda Liu,
Nanjing Medical University, China

REVIEWED BY
Elena Gershtein,
Russian Cancer Research Center NN Blokhin, Russia
Yuting Ke,
Massachusetts Institute of Technology, United States

*CORRESPONDENCE
Elvan Wiyarta
elvan.wiyarta@ui.ac.id

RECEIVED 07 April 2023
ACCEPTED 12 July 2023
PUBLISHED 26 July 2023

CITATION
Wiyarta E, Nugraha DA, Ramadani MI, Gustya GF, Ammar MF, Edwar HD, Kheirizzad N, Mukhlisah MN, Burhan E and Syahruddin E (2023) Clinical utility and diagnostic value of tumor-educated platelets in lung cancer: a systematic review and meta-analysis.
Front. Oncol. 13:1201713.
doi: 10.3389/fonc.2023.1201713

COPYRIGHT
© 2023 Wiyarta, Nugraha, Ramadani, Gustya, Ammar, Edwar, Kheirizzad, Mukhlisah, Burhan and Syahruddin. 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 utility and diagnostic value of tumor-educated platelets in lung cancer: a systematic review and meta-analysis

Elvan Wiyarta[1*], Darrin Ananda Nugraha[1], Muhammad Indera Ramadani[1], Gita Fajri Gustya[1], Muhammad Farrasy Ammar[1], Hana Dzakira Edwar[1], Nildza Kheirizzad[1], Mutiah Nurul Mukhlisah[1], Erlina Burhan[2] and Elisna Syahruddin[3]

[1]Respiratory and Tuberculosis Research and Training Center (SATURATE), Faculty of Medicine, Persahabatan National Hospital, Universitas Indonesia, Jakarta, Indonesia, [2]Infection Division, Department of Pulmonology, Faculty of Medicine, Universitas Indonesia, Persahabatan National Hospital, Jakarta, Indonesia, [3]Oncology Division, Department of Pulmonology, Faculty of Medicine, Universitas Indonesia, Persahabatan National Hospital, Jakarta, Indonesia

Background: The review addresses the knowledge gap concerning the diagnostic value and clinical utility of tumor-educated platelets (TEPs) in adult patients with lung cancer.

Methods: We searched twelve databases: PubMed, CENTRAL, EMBASE, CINAHL, MEDLINE, Scopus, ProQuest, MedRxiv, BioRxiv, SSRN, Clinicaltrials.gov, and CNKI up to 24 March 2023, to include any diagnostic study regarding TEPs and LC. TEPs diagnostic value was evaluated from pooled sensitivity and specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), diagnostic odds ratio (DOR), and the area under the curve (AUC). QUADAS 2 was used to assess the risk of bias. Heterogeneity analysis was assessed using the receiver operating characteristic (ROC) plane, Galbraith plot, bivariate boxplot, sensitivity analysis, and meta-regression. TEPs clinical utility was evaluated from Fagan's nomogram.

Results: 44 reports from 10 studies, including 7,858 events and 6,632 controls, were analyzed. The pooled sensitivity, specificity, PLR, NLR, and DOR were 0.80 (95% CI 0.79–0.80), 0.69 (95% CI 0.69–0.70), 2.92 (95% CI 2.50–3.41), 0.26 (95% CI 0.21–0.32), and 12.1 (95% CI 8.61–16.76), respectively. In addition, the AUC of the Summary ROC curve was 0.85 (95% CI: 0.81-0.88). The overall risk of bias was low. Heterogeneity may result from cancer stage, cancer control, measuring equipment, and RNA types across studies. There was no apparent publication bias (p=0.29) with significant positive (79%) and negative (22%) post-test probability, according to Deeks funnel plot asymmetry test and Fagan's nomogram.

Conclusion: TEPs could be a moderately effective candidate biomarker for LC diagnosis.

KEYWORDS

diagnosis, review, liquid biopsy, RNA, biomarker

Introduction

Lung cancer (LC) remains the leading cause of cancer-related deaths worldwide, representing approximately 18.4% of worldwide cancer deaths (1). LC is classified into two main types: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC) (2). NSCLC accounts for about 83% of LC cases, while SCLC accounts for approximately 13% (2). LC must be diagnosed early because it is an invasive and rapidly progressive disease (3, 4). Open surgical tissue biopsy is the current gold standard for diagnosing LC, but this procedure is invasive and associated with high failure rates for mutation evaluation (3, 5). Furthermore, obtaining a tissue biopsy from a single tissue location offers a restricted view of the disease, which may not capture the complexity and diversity of the Stumor (6–8).

Liquid biopsy has emerged as a valuable method for diagnosing LC because it offers the advantages of easier accessibility and more excellent coverage of cancer heterogeneity (8, 9). Blood samples for liquid biopsy can be analyzed for circulating tumor cells (CTCs), cell-free DNA (cfDNA), circulating tumor DNA (ctDNA), and tumor-educated platelets (TEPs) (10). Much research has focused on CTCs, ctDNA, and cfDNA, but these components have limitations. CTCs are challenging to detect due to their rarity and the need for an adequate number of tumor cells and multiple blood samples (11). ctDNA is less stable and has a short half-life (12). Lastly, quantitative cfDNA analysis has unsatisfactory sensitivity (13).

TEPs are a form of liquid biopsy that is becoming an increasingly popular research topic. TEPs are a type of platelet that has been exposed to the cancer microenvironment and subsequently modified to promote cancer growth and progression (14–16). When tumor cells proliferate, they release signals that activate platelets, causing them to undergo morphological and functional changes (14–16). Cancer sequestering extracellular vesicle (EV)-derived ribonucleic acid (RNA) and which alters the platelet RNA profile (17). Directly or indirectly, cancer cells can affect the RNA content of platelets, resulting in the "education" of platelets mediated by the tumor (14). As a novel source of potential biomarkers, alterations in the platelet RNA profile induced by tumors have been described (14). Individual platelet RNA biomarkers and complex RNA signatures can be used for early cancer detection and treatment monitoring (14, 18).

However, numerous TEPs studies, particularly those on LC, remain equivocal and inconclusive (19). Diagnostic studies involving TEPs with LC are difficult to conclude due to the diversity of methods and types of RNA employed, the small sample size, and the limitations of the research design. A comprehensive analysis of this beneficial biomarker's diagnostic value and clinical utility is crucial. This first comprehensive systematic review and meta-analysis on the diagnostic value and clinical utility of TEPs in LC was conducted to address this knowledge gap.

Methods

Search strategy

This systematic review and meta-analysis followed the Preferred Reporting Items for Systematic Reviews and Meta-analyses guidelines (PRISMA) (20). We systematically searched the literature in twelve databases: PubMed, Cochrane Controlled Register of Trials (CENTRAL), Ovid EMBASE, EBSCO Cumulated Index to Nursing and Allied Health Literature (CINAHL), Ovid MEDLINE, Scopus, ProQuest, MedRxiv, BioRxiv, Social Science Research Network (SSRN), Clinicaltrials.gov, and China National Knowledge Infrastructure (CNKI) up to 24 March 2023. The literature on TEPs for the diagnosis of LC was retrieved using the following general search strategies: (tumor-educated platelet) AND (LC) AND (diagnosis OR specificity OR sensitivity OR receiver operating characteristics OR ROC). The general searching strategy is used by first adjusting it to the format in each database, which can be seen in Supplementary Material S1, along with the Participant, Index test, Comparison, and Outcome (PICO) for this study (Supplementary Material S2). We also searched the reference lists of relevant research, systematic reviews, and meta-analyses to identify missed articles during the initial search. Two independent reviewers (DN and MR) independently examined all identified studies. All disagreements are resolved through discussion with the third reviewer (EW).

Inclusion and exclusion criteria

Inclusion criteria were: (1) diagnostic study design, (2) studies that include TEPs assessment for LC, (3) human-based studies, and (4) absolute numbers of true-positive, false-positive, or false-negative, or true-negative could be calculated from the study. On the other hand, the exclusion criteria were: (1) non-human subject research, (2) case report/series, (3) commentary/viewpoint, (4) narrative review, and (5) irretrievable full-text article.

Data extraction

Two authors (D.A.N. and M.I.R.) independently extracted data from the full texts of all included studies. The following data were extracted: first author, year of publication, publication location, ethnicity, sample size, age, gender, cancer type, cancer stage, laboratory methods, primer, gene symbol, RNA type, sensitivity, specificity, true positive, false positive, false negative, and true positive.

Quality assessment

pt?>Two independent evaluators (D.A.N. and M.I.R.) assessed the quality of these studies using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) framework, which consists of four domains: patient selection, index test, reference standard, and flow and timing. Each field is used to assess bias, and the first three domains were used to evaluate the applicability of these studies to clinical practice. The likelihood of prejudice and mistrust was categorized as "low," "high," or "unclear." All researchers engaged in discussions to resolve the differences.

Statistical analysis

Meta-analyses were performed using Review Manager 5.3 (Cochrane Collaboration, Oxford, England), STATA 12.0 (Stata Corp LP, TX, USA), and Meta-DiSc 1.4 (Romany Cajal Hospital, Madrid, Spain). The sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), diagnostic odds ratio (DOR), and corresponding 95% confidence intervals (CI) were used to ascertain the diagnostic value of TEPs in LC in a meta-analysis of diagnostic accuracy. The area under the curve (AUC) of the summary receiver operating characteristic curve (SROC) was determined to quantify the diagnostic accuracy. The SROC curve method is a meta-analysis of multiple distinct detection index investigations. By fitting the SROC curve, diagnostic accuracy is comprehensively evaluated based on the weight of their odds ratio.

The analyses for heterogeneity were conducted using the Q test and I2 statistics. P values less than 0.05 were considered statistically significant. A random effects model was employed when I2 > 50% and p-value <0.05 indicated considerable heterogeneity between the included studies. Otherwise, if there was no apparent heterogeneity, the fixed effects model was applied to evaluate the aggregated results. The heterogeneity caused by the threshold effect was assessed using the ROC plane. Galbraith plot and bivariate boxplots were utilized to determine the heterogeneity level. Subgroup analysis and meta-regression were used to evaluate the source of heterogeneity. Individual subgroup results were examined.

Using sensitivity analysis, the reliability of the results was determined. Our results were validated using a variety of statistical tests, including goodness-of-fit, bivariate normality, influence analysis, and outlier detection. The clinical value of TEPs as a diagnostic method was determined using Fagan's nomogram. Deeks funnel plot asymmetry test was used to examine potential publication bias. A p-value greater than 0.1 indicates the absence of publication bias.

Results

Literature searching and study screening

We searched 3837 articles in total from PubMed(723), CENTRAL (148), EMBASE(831), CINAHL(222), MEDLINE(353), Scopus(954), ProQuest(259), MedRxiv(98), BioRxiv(218)), SSRN(4), Clinicaltrials.gov(7), and CNKI(20). Of these, 3848 records were excluded after reading the title and abstract due to the following: duplication, not describing TEPs with LC, not a human model, or studies based on databases. Subsequently, 38 articles remained. After reading the entire text, 28 articles were excluded because of insufficient data, review/letter/editorial, and inappropriate design. Apart from going through the database, a literature search was also carried out through a citation search for the included papers. Citation search retrieves and assesses a study that is then included. Thus, ten studies were included (21–30). Of these ten studies, some had presented more than one report, so the total number of reports included in this review was 44. A flowchart of the whole selection process is shown in Figure 1.

Study characteristics and quality assessment

This review analyzed 14,490 patients, 7,858 events, and 6,632 controls. Of all the studies included, most were conducted in an Asian population (22, 24–30), dominated by men aged around 40 and involving various LC stages. All information regarding the demographic and clinical characteristics of the included studies can be seen in Supplementary Material S3. This review also summarizes the laboratory methods and primers used in each study, including platelet isolation, RNA quality assessment, RNA detection, RNA extraction, reverse transcription, PCR, and primers. These results can be seen in Supplementary Material S4, S5. The quality assessment is shown in Supplementary Material S6. In this review, the domain of patient selection, index test, and flow and timing has a high risk of bias. On the other hand, the domain of patient selection, index tests, and reference standards has a low applicability concern. The outcome summary of the included studies in the meta-analysis is shown in Table 1.

Diagnosis value of TEPs for LC

Forty-four reports from 10 eligible diagnostic studies (21–30) were meta-analyzed and illustrated in Figure 2. These plots indicate significant heterogeneity, with I2 values of 96.5%, 93.8%, 92.6%, 93.1%, and 90.6%, respectively, for sensitivity, specificity, PLR, NLR, and DOR; thus, we employed a random effects model in this meta-analysis. The pooled sensitivity, specificity, PLR, NLR, and DOR were 0.80 (95% CI 0.79–0.80), 0.69 (95% CI 0.69–0.70), 2.92 (95% CI 2.50–3.41), 0.26 (95% CI 0.21–0.32), and 12.1 (95% CI 8.61–16.76), respectively. In addition, the AUC of the SROC curve, which indicates diagnostic accuracy, was 0.85 (95% CI: 0.81-0.88), indicating that the diagnostic value of TEPs was moderate (Figure 3).

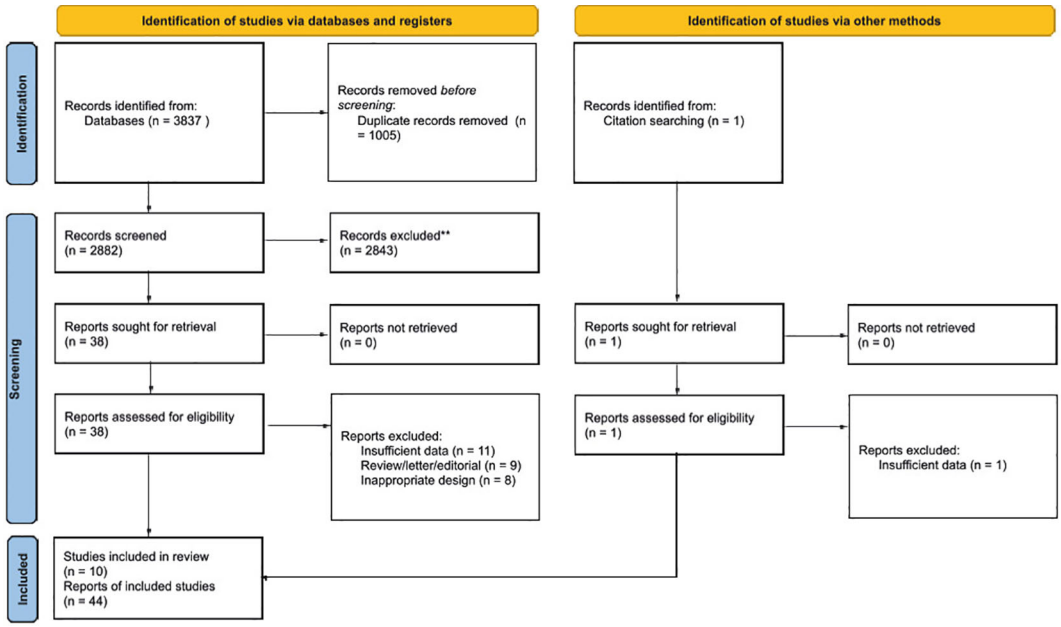


FIGURE 1
Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 flow chart of the study selection process.

TABLE 1 Outcome Summary of Studies Included in Meta-Analysis.

Author, year	Ethnicity	Gene Symbol	RNA Type	Event	Control	Total	Cancer Type	Control Type	Sensitivity	Specificity
Best, 2017 (21)	Caucasian	thromboSeq	vary	85	33	118	Early NSCLC	HS	72.10%	88.90%
				381	76	457	Late NSCLC	HS	88.10%	84.90%
Luo, 2018 (22)	Chinese	MAGI2-AS3	lncRNA	68	60	128	AD	HS	77.90%	81.70%
				33	60	93	SCC	HS	78.70%	88.30%
		ZFAS1	lncRNA	68	60	128	AD	HS	75.00%	66.70%
				33	60	93	SCC	HS	93.90%	43.30%
Sheng, 2018 (23)	Caucasian	48-genes biomarker panel	vary	402	231	633	NSCLC	HS	92.50%	82.70%
Xue, 2018 (24)	Chinese	ACIN1	mRNA	156	58	214	LC	HS	82.70%	44.80%
Liu, 2019 (25)	Chinese	MAX+MTURN +HLA-B	mRNA	127	82	209	LC	HS	60.6%	81.7%
				33	82	115	Early LC	HS	72.70%	85.40%
Xing, 2019 (26)	Chinese	ITGA2B	mRNA	131	45	176	NSCLC	HS+BPN	91.20%	56.50%
				102	21	123	NSCLC	BPN	91.20%	59.40%
				84	42	126	NSCLC stage I	HS+BPN	87.80%	56.50%
				49	24	73	NSCLC stage I	BPN	87.80%	59.40%
		SELP	mRNA	125	51	176	NSCLC	HS+BPN	96.70%	43.10%
				102	21	123	NSCLC	BPN	96.70%	43.80%
				75	51	126		HS+BPN	93.10%	43.10%

(Continued)

TABLE 1 Continued

Author, year	Ethnicity	Gene Symbol	RNA Type	Event	Control	Total	Cancer Type	Control Type	Sensitivity	Specificity
							NSCLC stage I			
				52	21	73	NSCLC stage I	BPN	93.10%	43.80%
Dong, 2020 (27)	Chinese	RNU1	snRNA	382	361	743	NSCLC	HS	74.60%	66.50%
				80	361	441	Early NSCLC	HS	90.00%	38.50%
		RNU2	snRNA	382	361	743	NSCLC	HS	81.40%	74.20%
				80	360	440	Early NSCLC	HS	78.80%	74.20%
		RNU5	snRNA	534	209	743	NSCLC	HS	90.10%	63.70%
				80	361	441	Early NSCLC	HS	86.30%	60.70%
		Combination	snRNA	382	361	743	NSCLC	HS	85.90%	70.10%
				80	361	441	Early NSCLC	HS	93.80%	60.70%
Yao, 2020 (28)	Chinese	CD274	circRNA	562	6	568	NSCLC	HS	90.91%	83.91%
		ITGA2B	circRNA	561	7	568	NSCLC	HS	98.93%	79.31%
		TIMP1	circRNA	509	59	568	NSCLC	HS	89.61%	78.16%
		FLNA	circRNA	516	7	523	NSCLC	HS	90.91%	83.91%
Dong, 2021 (29)	Chinese	SNORD55	snoRNA	290	189	479	NSCLC	HS	79.30%	68.30%
				91	189	280	Early NSCLC	HS	91.20%	49.70%
				204	189	393	AD	HS	77.90%	68.30%
				68	189	257	Early AD	HS	89.70%	49.70%
				76	189	265	SCC	HS	72.40%	77.70%
				19	189	208	Early SCC	HS	68.40%	93.10%
Li, 2021 (30)	Chinese	linc-GTF2H2-1	lncRNA	167	202	369	LC	HS	68.30%	81.70%
				47	202	249	Early LC	HS	69.60%	81.70%
		RP3-466P17.2	lncRNA	167	192	359	LC	HS	77.80%	67.80%
				47	202	249	Early LC	HS	87.20%	55.90%
		lnc-ST8SIA4-12	lncRNA	167	202	369	LC	HS	73.70%	69.80%
				47	202	249	Early LC	HS	83.00%	68.30%
		Combination	lncRNA	167	202	369	LC	HS	82.60%	87.10%
				47	202	249	Early LC	HS	93.60%	69.80%

AD, adenocarcinoma; BPN, benign pulmonary nodules; circRNA, circular ribonucleic acid; HS, healthy subject; LC, lung cancer; lncRNA, long non-coding ribonucleic acid; mRNA, messenger ribonucleic acid; NR, not reported; NSCLC, non-small cell lung cancer; RNA, ribonucleic acid; SCC, squamous cell carcinoma; snoRNA, small nucleolar ribonucleic acid; snRNA, small nuclear ribonucleic acid.

Heterogeneity and sensitivity analysis

As illustrated in Figure 2, apparent heterogeneity was found in the pooled sensitivity (I2 = 96.5%, P < 0.001), specificity (I2 = 93.8%, P < 0.001), PLR (I2 = 92.6%, P < 0.001), NLR (I2 = 93.1%, P < 0.001), and DOR (I2 = 90.6%, P < 0.001). To find the potential source of this heterogeneity, we carried out the ROC plane, Galbraith plot, bivariate boxplot, sensitivity analysis, subgroup analysis, and meta-regression analysis. No typical shoulder arm was observed in the ROC plane (Figure 4A), indicating that the threshold effect produced no significant heterogeneity. Twenty-two of the 44 reports in the Galbraith plot (Figure 4B) and 7 of the 44 reports in the bivariate boxplot (Figure 4C)

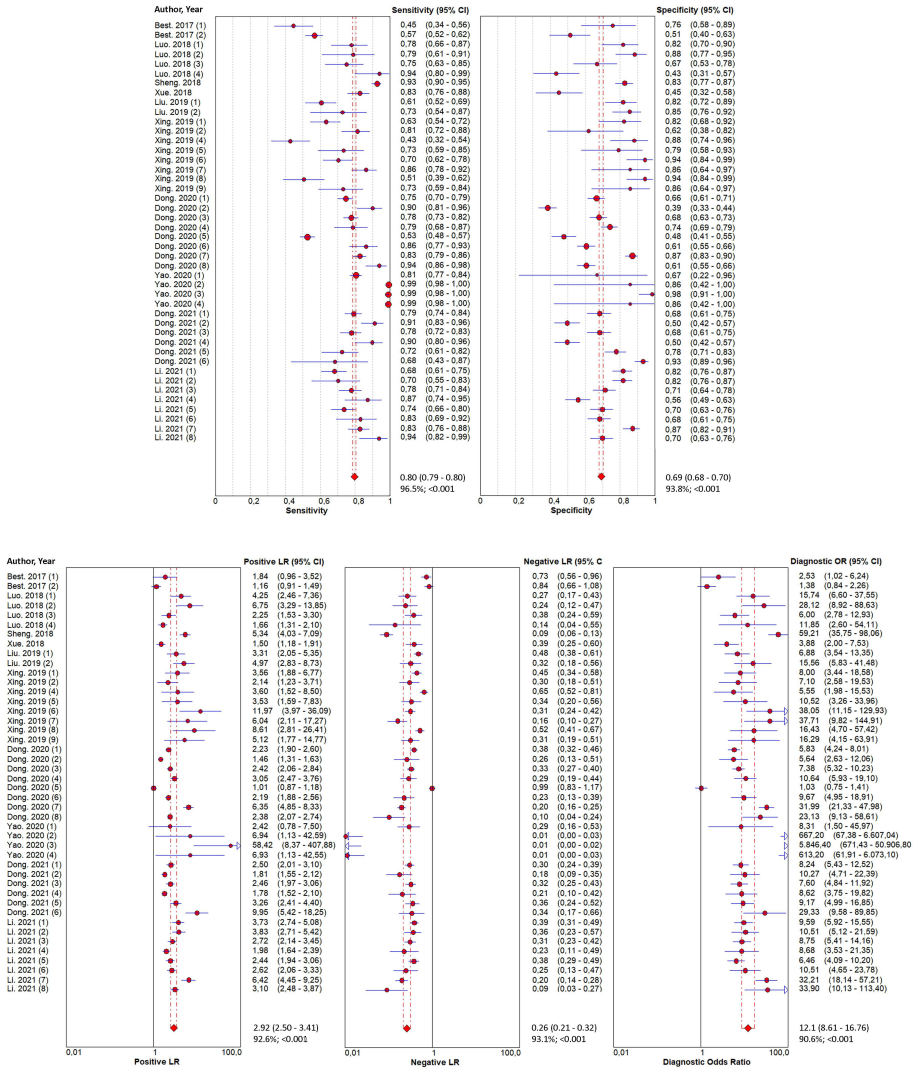


FIGURE 2
Forest plots of the diagnostic value for tumor-educated platelets in detecting lung cancer. Diagnostic value was analyzed using sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, and diagnostic odds ratio.

fell outside the 95% CI. These results show that some of these reports may have influenced this review's heterogeneity. To confirm this result, we conducted a sensitivity analysis to determine the stability of our results using goodness-of-fit, bivariate normality, influence analysis, and outlier detection (Supplementary Material S7). The sensitivity analysis results show six reports from three studies (26–28) that mainly affect heterogeneity. After excluding these six reports, the I2 for heterogeneity decreased concerning sensitivity (96.5% to 89.9%) and specificity (93.8% to 91.6%). The overall results showed only minimal changes. To analyze the potential source of heterogeneity, we carried out further subgroup analysis (Table 2) and meta-regression analysis (Figure 4D). All studies were grouped according to ethnicity (Caucasian and Chinese), RNA type (mRNA, lncRNA, snRNA, circRNA, and snoRNA), control type (HS, BPN, and HS+BPN), cancer type (NSCLC, AD, SCC, and LC), sample size (<100 and ≥100), and cancer stage (early and unspecified). We found that the cancer stage and control type may have accounted for part of the heterogeneity, with $p < 0.01$ and $p < 0.01$ for sensitivity and specificity in the cancer stage and $p < 0.01$ for specificity in the control type. Furthermore, the subgroup analysis based on the control type in the BPN group showed a decrease in I2 to 45.5% and 31.9%, respectively, for sensitivity and specificity.

Clinical value of TEPs for LC and publication bias

According to Fagan's nomogram (Figure 5A), the positive post-test probability of diagnosing LC would rise to 79%, while the negative post-test probability would drop to 22%, with a pre-test probability of 54%. The pre-test probability is calculated based on the prevalence of the event group in the total included study (Table 1). The P value of Deeks funnel plot asymmetry test (Figure 5B) was 0.29, showing no publication bias across the studies.

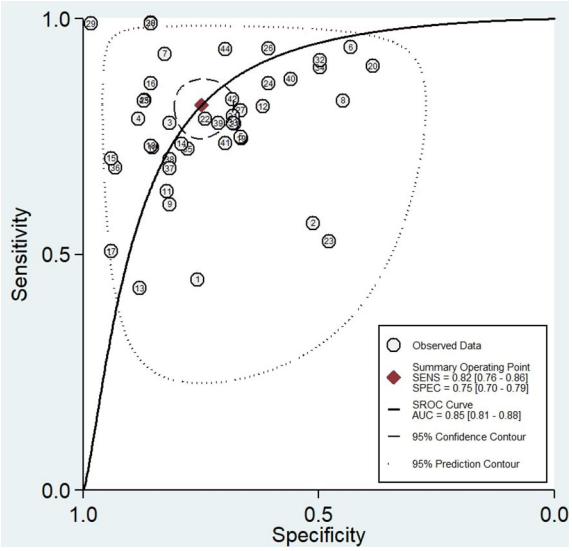


FIGURE 3
Summary Receiver Operating Characteristic (SROC) curve of tumor-educated platelets in detecting lung cancer.

Discussion

TEPs broadened the spectrum of liquid biopsy applications and may enable blood-based cancer diagnostics, especially in LC (31). Although several other biomarkers, such as carcinoembryonic antigen (CEA) and Cytokeratin-19 fragment (CYFRA21-1), are commonly used to diagnose LC (32–36), they have low diagnostic effectiveness for the initial stages of lung cancer (34, 37–44). TEPs offer the advantages of faster results, a less invasive nature, and a more convenient technique for diagnosing LC (22). The advantages of TEPs, compared to other liquid biopsy techniques, are related to their abundance, high stability in blood, and ease of isolation (45). Despite TEPs being the relatively more straightforward liquid biopsy methods, the applications of TEPs profiles are still in early development, which may require a lengthy process from biomarker discovery, and design verification, to approval (46, 47). This shortcoming and the high cost of low-input deep sequencing (48), hinder the clinical application of TEPs for detecting LC (46).

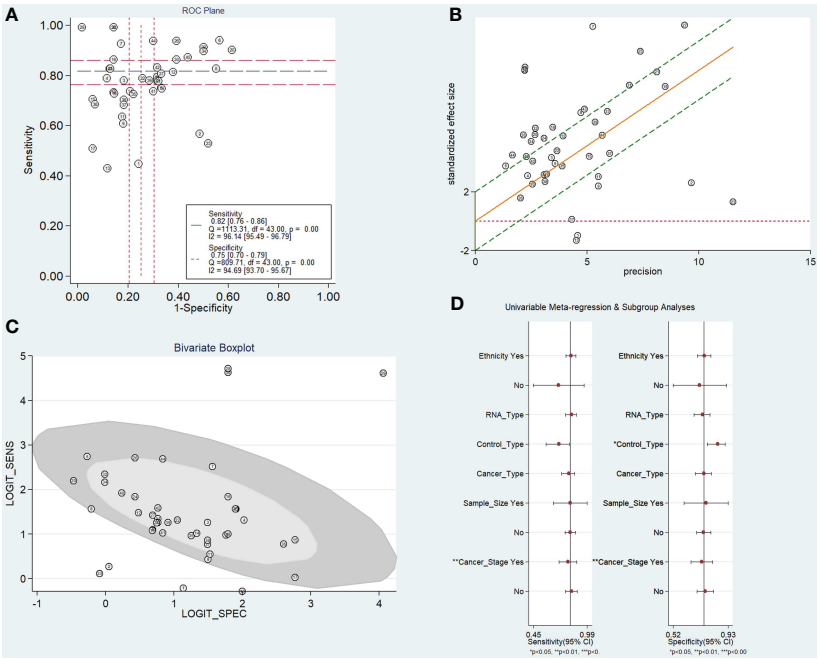


FIGURE 4
Heterogeneity analysis of diagnostic tests. **(A)** Receiver Operating Characteristic plane. **(B)** Galbraith plot. **(C)** Bivariate boxplot. **(D)** Subgroup and meta-regression analysis for heterogeneity.

TABLE 2 Subgroup analysis of the diagnostic efficacy of tumor-educated platelet in lung cancer.

Parameter	No. Report	No. Patient	Sensitivity (95% CI)	Heterogeneity (I^2; P-value)	Specificity (95% CI)	Heterogeneity (I^2; P-value)
Ethnicity						
Caucasian	3 (21, 23)	1090	0.721 (0.690-0.751)	98.9%; <0.001	0.750 (0.700-0.795)	92.8%; <0.001
Chinese	41 [22, 24–30]	13151	0.805 (0.796-0.814)	96.0%; <0.001	0.688 (0.677-0.700)	93.9%; <0.001
RNA Type						
mRNA	11 (24–26)	1534	0.695 (0.666-0.723)	87.8%; <0.001	0.805 (0.768-0.839)	83.2%; <0.001
lncRNA	12 (22, 30)	2904	0.778 (0.752-0.803)	66.3%; 0.001	0.728 (0.707-0.748)	89.8%; <0.001
snRNA	8 (27)	4735	0.730 (0.709-0.749)	95.9%; <0.001	0.637 (0.619-0.655)	97.1%; <0.001
circRNA	4 (28)	2227	0.942 (0.932-0.952)	98.7%; <0.001	0.937 (0.858-0.979)	62.6%; 0.046
snoRNA	6 (29)	1882	0.803 (0.773-0.831)	71.9%; 0.003	0.678 (0.650-0.705)	96.3%; <0.001
Control Type						
HS	36 (21–25, 27–30)	13494	0.807 (0.798-0.816)	96.8%; <0.001	0.684 (0.672-0.695)	94.5%; <0.001
BPN	4 (26)	392	0.803 (0.754-0.846)	**45.5%; 0.138**	0.782 (0.680-0.863)	**31.9%; 0.221**
HS+BPN	4 (26)	604	0.590 (0.541- 0.638)	84.2%; <0.001	0.899 (0.847-0.938)	38.8%; 0.179
Cancer Type						
NSCLC	25 (21, 23, 26–29)	9925	0.803 (0.793-0.813)	97.9%; <0.001	0.665 (0.650-0.680)	94.4%; <0.001
AD	4 (22, 29)	906	0.794 (0.752-0.832)	**52.3%; 0.098**	0.627 (0.582-0.669)	88.6%; <0.001
SCC	4 (22, 29)	659	0.776 (0.704-0.838)	65.2%; 0.035	0.807 (0.770-0.841)	95.6%; <0.001
LC	11 (24, 25, 30)	3000	0.761 (0.736-0.785)	78.1%; <0.001	0.732 (0.712-0.753)	90.3%; <0.001
Sample Size						
<100	4 (22, 26)	332	0.784 (0.714-0.844)	60.7%; 0.054	0.703 (0.627-0.772)	91.1%; <0.001
≥100	40 (21–30)	14158	0.796 (0.787-0.805)	96.8%; <0.001	0.691 (0.680-0.702)	94.1%; <0.001
Cancer Stage						
Early	17 (21, 25–27, 29, 30)	4135	0.755 (0.728-0.780)	90.9%; <0.001	0.645 (0.628-0.662)	95.4%; <0.001
Unspecified	27 (21–30)	10355	0.802 (0.793-0.812)	97.5%; <0.001	0.731 (0.716-0.746)	91.1%; <0.001

AD, adenocarcinoma; BPN, benign pulmonary nodule; CI, confidence interval; circRNA, circular ribonucleic acid; HS, healthy subject; LC, lung cancer; lncRNA, long non-coding ribonucleic acid; mRNA, messenger ribonucleic acid; NSCLC, non-small cell lung cancer; RNA, ribonucleic acid; SCC, squamous cell carcinoma; snoRNA, small nucleolar ribonucleic acid; snRNA, small nuclear ribonucleic acid.

Nonetheless, TEPs have a potential role as a biomarker for diagnosing LC, which is comprehensively discussed in this review.

As seen in Table 1, most of the studies included have substantially diverse results, which lead to inconclusive diagnostic value and clinical utility of TEPs. Each study analyzed in this review used a distinct control group and examined varying stages of lung cancer, all while employing different types of RNA to diagnose the disease. Most of the research included in this review enrolled HS as the control group (21–25, 27–30), while one study enrolled patients with BPN (26). Several studies have also used populations of different stages of cancer, starting from the early stage to the late stage, with non-uniform proportions (21–30). It is also evident that all studies employ diverse RNA types, including mRNA, rRNA, miRNA, snRNA, snoRNA, asRNA, tRNA, circRNA, and lncRNA (28, 48). In patients with cancerous tumors, a particular type of RNA may interact differently than other RNA (29). The results of RNA expression measurement from the same RNA family can show different results (25, 29, 30). Furthermore, it is believed that multiple pathways are involved in the formation of TEPs, including direct communication between tumor cells and platelets, the transmission of information from tumor cells to platelets through extracellular vesicles, and the influence of tumor cells on megakaryocytes (29). Due to the inconclusive results of the various included studies, which were summarized qualitatively, the assessment of the diagnostic value and clinical utility of TEPs was continued with meta-analysis, a more objective assessment.

In the present meta-analysis, screening was performed on 44 reports from 10 eligible studies. The pooled sensitivity, specificity, PLR, NLR, DOR, and SROC AUC results indicated that TEPs have a moderate diagnostic accuracy for LC. The sensitivity analysis confirmed the consistency of the findings, and the Deeks funnel plot asymmetry test demonstrated no apparent publication bias. The

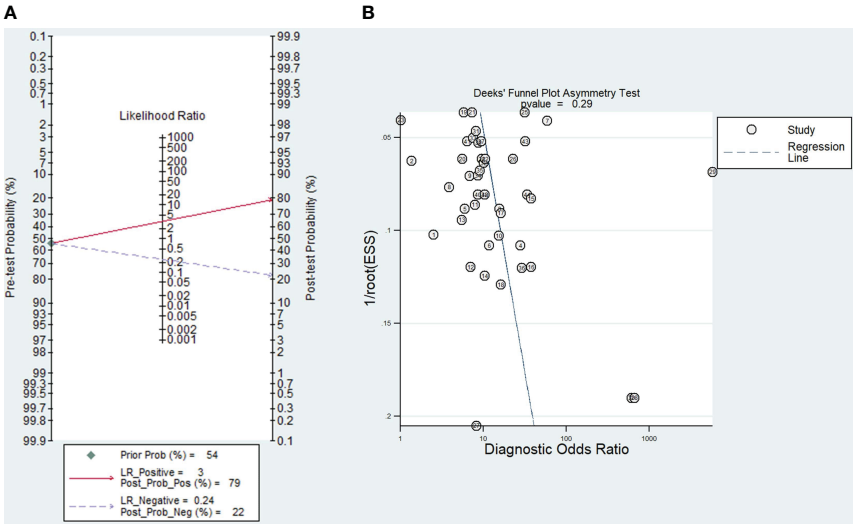


FIGURE 5
(A) Fagan's nomogram exploring the tumor-educated platelets' clinical utility in lung cancer with the corresponding **(B)** Deek's funnel plot.

Fagan diagram also illustrates its clinical application advantages, primarily attributable to its moderately high positive and negative predictive values.

However, the pooled results must be interpreted cautiously because this review has apparent heterogeneity. Heterogeneity analysis was carried out to find the cause of the emergence of heterogeneity. The ROC plane was utilized to determine whether heterogeneity resulted from the threshold effect. The results demonstrated an atypical shoulder arm, indicating that TEPs have no threshold effect. Six reports from three studies contributed to the significant heterogeneity in TEPs analyses, as indicated by the Galbraith plot, bivariate boxplot, and sensitivity analysis. After excluding these reports, the I2 for heterogeneity decreased. In addition, meta-regression analysis of TEPs revealed that cancer stage and control type may have contributed to the high degree of heterogeneity. Other factors, disparities in measuring equipment and the use of various types of RNA across studies, may also contribute to this heterogeneity.

Heterogeneity due to cancer stages may be associated with the cancer stage and could be linked to the characteristics of platelet RNA in the advancement of tumors (49). For instance, numerous mRNA molecules exhibited different expressions in the platelets of individuals with localized and metastatic cancer (49). Additionally, gene expression was upregulated from the early to the later stage of cancer (49). We also found that the control type attributed to our study's overall heterogeneity. Some studies used patients with BPN as the control arm, which could lead to heterogeneity because BPN is a constellation of diseases that may result from numerous inflammatory conditions (e.g., tuberculosis, pneumonia, pulmonary abscess), each having different baseline conditions (26).

Our subgroup analysis categorized reports based on different types of RNA used. Each type of RNA has its specific function in gene expression that occurs during tumor cell cycle (31, 50, 51). Also, previous studies reported that tumor-specific signaling in patients with cancer led to distinct RNA processing compared to healthy donors, resulting in numerous variations of genetics between each type of RNA. This finding may result in substantial heterogeneity in our study results (31, 50, 51). We also highlighted the potential effect of different measuring equipment across studies. The results between different measuring equipment may not be comparable due to different data measurements.

According to Fagan's nomogram, our findings further demonstrate the solid clinical value of TEPs, as evidenced by a 25% increase and a 32% decrease in post-test probability values. This finding suggests that TEPs possess a robust diagnostic capability for LC detection. Given its less invasive nature (51), another potential of TEPs in monitoring and prognosticating NSCLC warrants further investigation. TEPs offer superior clinical applications compared to other liquid biopsy approaches, as platelet isolation is an economical, straightforward process routinely performed for many years (52).

Nonetheless, several limitations of this meta-analysis should be highlighted, both from the evidence included and the review process conducted. First, there was substantial heterogeneity among the included studies and in the subgroup analysis. Second, most of the studies included in this meta-analysis analyzed data from Asian cohorts, and there is a dearth of information regarding TEPs of other ethnicities. Lastly, TEPs are a recently discovered tumor biomarker, and only a limited number of studies (10 studies) can be included in the meta-analysis. This results in incorporating its reports (44 reports in total), which reduces the robustness of certain aggregated analysis results.

This review has implications for clinical practice, future research, and policy. This review demonstrates that TEPs can be a reliable diagnostic instrument for LC in clinical practice. To anticipate the rapid development of science, policies regarding the withdrawal of TEPs in patients with LC can be initiated as early as feasible. In addition, to resolve heterogeneity in this research, future studies on TEPs in LC must have a narrower focus based on factors that influence heterogeneity, e.g., cancer stage, cancer control, methods used, and

RNA type. Future studies must conduct TEPs research on specific LC populations, focusing, for instance, on diagnosing early-stage LC compared to HS patients using standardized methodologies for a particular RNA type. It is anticipated that future studies will be able to provide a more uniform picture of the population for implementing TEPs. In addition, in order to accommodate the external validity and generalizability of TEP in lung cancer, further studies are needed to analyze individual datasets, especially with diagnostic study design.

Conclusion

TEPs could be a moderately effective candidate biomarker for LC diagnosis. This review establishes an essential standard for using TEPs as biomarkers in the early detection of LC. Due to potential limitations, additional research is necessary to corroborate the diagnostic value and clinical utility of TEPs in LC.

Data availability statement

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

Author contributions

Conceptualization, EW, GG, and MA. Methodology, EW, DN, MR. Software, GG and MA. Validation, HE, NK, and MM. Formal analysis, EW, DN, and MR. Investigation, GG and MA. Resources, EW. Data curation, EW. Writing—original draft preparation, EW and NK. Writing—review and editing, all authors. Visualization, EW. Supervision, EB and ES. Project administration, EW. Funding acquisition, EB and ES. 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/fonc.2023.1201713/full#supplementary-material

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* (2018) 68(6):394–424. doi: 10.3322/caac.21492

2. Miller KD, Siegel RL, Lin CC, Mariotto AB, Kramer JL, Rowland JH, et al. Cancer treatment and survivorship statistics, 2016. *CA Cancer J Clin* (2016) 66(4):271–89. doi: 10.3322/caac.21349

3. Nooreldeen R, Bach H. Current and Future Development in Lung Cancer Diagnosis. *Int J Mol Sci* (2021) 22(16):8661. doi: 10.3390/ijms22168661

4. Health Commission Of The People's Republic Of China N. National guidelines for diagnosis and treatment of lung cancer 2022 in China (English version). *Chin J Cancer Res* (2022) 34(3):176–206. doi: 10.21147/j.issn.1000-9604.2022.03.03

5. Vanderlaan PA, Yamaguchi N, Folch E, Boucher DH, Kent MS, Gangadharan SP, et al. Success and failure rates of tumor genotyping techniques in routine pathological samples with non-small-cell lung cancer. *Lung Canc.* (2014) 84(1):39–44. doi: 10.1016/j.lungcan.2014.01.013

6. Gilson P, Merlin JL, Harlé A. Deciphering Tumour Heterogeneity: From Tissue to Liquid Biopsy. *Cancers (Basel).* (2022) 14(6):1384. doi: 10.3390/cancers14061384

7. Jamal-Hanjani M, Quezada SA, Larkin J, Swanton C. Translational implications of tumor heterogeneity. *Clin Cancer Res* (2015) 21(6):1258–66. doi: 10.1158/1078-0432.CCR-14-1429

8. Liu L, Lin F, Ma X, Chen Z, Yu J. Tumor-educated platelet as liquid biopsy in lung cancer patients. *Crit Rev Oncol Hematol* (2020) 146:102863. doi: 10.1016/j.critrevonc.2020.102863

9. Michela B. Liquid Biopsy: A Family of Possible Diagnostic Tools. *Diagnost. (Basel)* (2021) 11(8):1391. doi: 10.3390/diagnostics11081391

10. Eslami SZ, Cortes-Hernandez LE, Cayrefourcq L, Alix-Panabieres C. The Different Facets of Liquid Biopsy: A Kaleidoscopic View. *Cold Spring Harb Perspect Med* (2020) 10(6):37333. doi: 10.1101/cshperspect.a037333

11. Tamminga M, Groen HH, Hiltermann TJ. Investigating CTCs in NSCLC-a reaction to the study of Jia-Wei Wan: a preliminary study on the relationship between circulating tumor cells count and clinical features in patients with non-small cell lung cancer. *J Thorac Dis* (2016) 8(6):1032–6. doi: 10.21037/jtd.2016.04.17

12. Mamdani H, Ahmed S, Armstrong S, Mok T, Jalal SI. Blood-based tumor biomarkers in lung cancer for detection and treatment. *Transl Lung Cancer Res* (2017) 6(6):648–60. doi: 10.21037/tlcr.2017.09.03

13. Zhou Q, Li W, Leng B, Zheng W, He Z, Zuo M, et al. Circulating Cell Free DNA as the Diagnostic Marker for Ovarian Cancer: A Systematic Review and Meta-Analysis. *PloS One* (2016) 11(6):e0155495. doi: 10.1371/journal.pone.0155495

14. D'Ambrosi S, Nilsson RJ, Wurdinger T. Platelets and tumor-associated RNA transfer. *Blood* (2021) 137(23):3181–91. doi: 10.1182/blood.2019003978

15. McAllister SS, Weinberg RA. The tumour-induced systemic environment as a critical regulator of cancer progression and metastasis. *Nat Cell Biol* (2014) 16(8):717–27. doi: 10.1038/ncb3015

16. Xu XR, Zhang D, Oswald BE, Carrim N, Wang X, Hou Y, et al. Platelets are versatile cells: New discoveries in hemostasis, thrombosis, immune responses, tumor metastasis and beyond. *Crit Rev Clin Lab Sci* (2016) 53(6):409–30. doi: 10.1080/10408363.2016.1200008

17. Best MG, Wesseling P, Wurdinger T. Tumor-Educated Platelets as a Noninvasive Biomarker Source for Cancer Detection and Progression Monitoring. *Cancer Res* (2018) 78(13):3407–12. doi: 10.1158/0008-5472.CAN-18-0887

18. Yu L, Guo Y, Chang Z, Zhang D, Zhang S, Pei H, et al. Bidirectional Interaction Between Cancer Cells and Platelets Provides Potential Strategies for Cancer Therapies. *Front Oncol* (2021) 11:764119. doi: 10.3389/fonc.2021.764119

19. Varkey J, Nicolaides T. Tumor-Educated Platelets: A Review of Current and Potential Applications in Solid Tumors. *Cureus* (2021) 13(11):e19189. doi: 10.7759/cureus.19189

20. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* (2021) 372:n71. doi: 10.1136/bmj.n71

21. Best MG, Sol N, In 't Veld SGJG, Vancura A, Muller M, Niemeijer ALN, et al. Swarm Intelligence-Enhanced Detection of Non-Small-Cell Lung Cancer Using Tumor-Educated Platelets. *Cancer Cell* (2017) 32(2):238–52.e9. doi: 10.1016/j.ccell.2017.07.004

22. Luo CL, Xu ZG, Chen H, Ji J, Wang YH, Hu W, et al. LncRNAs and EGFRvIII sequestered in TEPs enable blood-based NSCLC diagnosis. *Cancer Manage Res* (2018) 10:1449–59. doi: 10.2147/CMAR.S164227

23. Sheng M, Dong Z, Xie Y. Identification of tumor-educated platelet biomarkers of non-small-cell lung cancer. *OncoTargets Ther* (2018) 11:8143–51. doi: 10.2147/OTT.S177384

24. Xue L, Xie L, Song X, Song X. Expression and significance of ACIN1 mRNA in platelets of lung cancer. *Chin J Lung Canc.* (2018) 21(9):677–81. doi: 10.3779/j.issn.1009-3419.2018.09.05

25. Liu L, Song X, Li X, Xue L, Ding S, Niu L, et al. A three-platelet mRNA set: MAX, MTURN and HLA-B as biomarker for lung cancer. *J Cancer Res Clin Oncol* (2019) 145 (11):2713–23. doi: 10.1007/s00432-019-03032-9

26. Xing S, Zeng T, Xue N, He Y, Yan-zhen L, Hui-lan L, et al. Development and Validation of Tumor-educated Blood Platelets Integrin Alpha 2b (ITGA2B) RNA for Diagnosis and Prognosis of Non-small-cell Lung Cancer through RNA-seq. *Int J Biol Sci* (2019) 15(9):1977–92. doi: 10.7150/ijbs.36284

27. Dong X, Ding S, Yu M, Niu L, Xue L, Zhao Y, et al. Small Nuclear RNAs (U1, U2, U5) in Tumor-Educated Platelets Are Downregulated and Act as Promising Biomarkers in Lung Cancer. *Front Oncol* (2020) 10. doi: 10.3389/fonc.2020.01627

28. Yao B, Yan X, Jing He, Jiao Z, Zhuang Yi, Qu S, et al. *A panel of Platelet circRNAs serve as biomarker for lung cancer detection*. SSRN Preprint (2020).

29. Dong X, Song X, Ding S, Yu M, Shang X, Wang K, et al. Tumor-educated platelet SNORD55 as a potential biomarker for the early diagnosis of non-small cell lung cancer. *Thorac Canc.* (2021) 12(5):659–66. doi: 10.1111/1759-7714.13823

30. Li X, Liu L, Song X, Wang K, Niu L, Xie L, et al. TEP linc-GTF2H2-1, RP3-466P17.2, and lnc-ST8SIA4-12 as novel biomarkers for lung cancer diagnosis and progression prediction. *J Cancer Res Clin Oncol* (2021) 147(6):1609–22. doi: 10.1007/s00432-020-03502-5

31. Calverley DC, Phang TL, Choudhury QG, Gao B, Oton AB, Weyant MJ, et al. Significant downregulation of platelet gene expression in metastatic lung cancer. *Clin Transl Sci* (2010) 3(5):227–32. doi: 10.1111/j.1752-8062.2010.00226.x

32. Xie Y, Zhang Y, Du L, Jiang X, Yan S, Duan W, et al. Circulating long noncoding RNA act as potential novel biomarkers for diagnosis and prognosis of non-small cell lung cancer. *Mol Oncol* (2018) 12(5):648–58. doi: 10.1002/1878-0261.12188

33. Ren S, Zhang S, Jiang T, He Y, Ma Z, Cai H, et al. Early detection of lung cancer by using an autoantibody panel in Chinese population. *Oncoimmunology* (2018) 7(2): e1384108. doi: 10.1080/2162402X.2017.1384108

34. Molina R, Marrades RM, Auge JM, Escudero JM, Vinolas N, Reguart N, et al. Assessment of a Combined Panel of Six Serum Tumor Markers for Lung Cancer. *Am J Respir Crit Care Med* (2016) 193(4):427–37. doi: 10.1164/rccm.201404-0603OC

35. Zamcheck N, Pusztaszeri GCEA. AFP and other potential tumor markers. *CA Cancer J Clin* (1975) 25(4):204–14. doi: 10.3322/canjclin.25.4.204

36. Ferrigno D, Buccheri G, Biggi A. Serum tumour markers in lung cancer: history, biology and clinical applications. *Eur Respir J* (1994) 7(1):186–97. doi: 10.1183/09031936.94.07010186

37. Molina R, Filella X, Auge JM, Fuentes R, Bover I, Rifa J, et al. Tumor markers (CEA, CA 125, CYFRA 21-1, SCC and NSE) in patients with non-small cell lung cancer as an aid in histological diagnosis and prognosis. Comparison with the main clinical and pathological prognostic factors. *Tumour Biol* (2003) 24(4):209–18. doi: 10.1159/000074432

38. Lai Y, Wang X, Zeng T, Xing S, Dai S, Wang J, et al. Serum VEGF levels in the early diagnosis and severity assessment of non-small cell lung cancer. *J Canc.* (2018) 9 (9):1538–47. doi: 10.7150/jca.23973

39. Ludwig JA, Weinstein JN. Biomarkers in cancer staging, prognosis and treatment selection. *Nat Rev Canc.* (2005) 5(11):845–56. doi: 10.1038/nrc1739

40. Sajid KM, Parveen R, Durr e S, Chaouachi K, Naeem A, Mahmood R, et al. Carcinoembryonic antigen (CEA) levels in hookah smokers, cigarette smokers and non-smokers. *J Pak Med Assoc* (2007) 57(12):595–9.

41. Correa-Gallego C, Warshaw AL, Fernandez-del Castillo C. Fluid CEA in IPMNs: A useful test or the flip of a coin? *Am J Gastroenterol* (2009) 104(3):796–7. doi: 10.1038/ajg.2008.158

42. Sreenarasimhaiah J, Lara LF, Jazrawi SF, Barnett CC, Tang SJ. A comparative analysis of pancreas cyst fluid CEA and histology with DNA mutational analysis in the detection of mucin producing or malignant cysts. *JOP* (2009) 10(2):163–8.

43. Li X, Asmitananda T, Gao L, Gai D, Song Z, Zhang Y, et al. Biomarkers in the lung cancer diagnosis: a clinical perspective. *Neoplasma* (2012) 59(5):500–7. doi: 10.4149/neo_2012_064

44. Plebani M, Basso D, Navaglia F, De Paoli M, Tommasini A, Cipriani A. Clinical evaluation of seven tumour markers in lung cancer diagnosis: can any combination improve the results? *Br J Cancer* (1995) 72(1):170–3. doi: 10.1038/bjc.1995.296

45. Connal S, Cameron JM, Sala A, Brennan PM, Palmer DS, Palmer JD, et al. Liquid biopsies: the future of cancer early detection. *J Transl Med* (2023) 21(1):118. doi: 10.1186/s12967-023-03960-8

46. Chen M, Hou L, Hu L, Tan C, Wang X, Bao P, et al. Platelet detection as a new liquid biopsy tool for human cancers. *Front Oncol* (2022) 12:983724–. doi: 10.3389/fonc.2022.983724

47. Wang Y. Development of cancer diagnostics—from biomarkers to clinical tests. *Trans Cancer Res* (2015) 4(3):270–9. doi: 10.3978/j.issn.2218-676X.2015.06.02

48. Sun Y, Liu R, Xia X, Xing L, Jiang J, Bian W, et al. Large-Scale Profiling on lncRNAs in Human Platelets: Correlation with Platelet Reactivity. *Cells* (2022) 11 (14):2256. doi: 10.3390/cells11142256

49. Ge X, Yuan L, Cheng B, Dai K. Identification of seven tumor-educated platelets RNAs for cancer diagnosis. *J Clin Lab Anal* (2021) 35(6):e23791. doi: 10.1002/jcla.23791

50. Nilsson RJ, Balaj L, Hulleman E, van Rijn S, Pegtel DM, Walraven M, et al. Blood platelets contain tumor-derived RNA biomarkers. *Blood* (2011) 118(13):3680–3. doi: 10.1182/blood-2011-03-344408

51. Roweth HG, Battinelli EM. Lessons to learn from tumor-educated platelets. *Blood* (2021) 137(23):3174–80. doi: 10.1182/blood.2019003976

52. Cazenave J-P, Ohlmann P, Cassel D, Eckly A, Hechler B, Gachet C. Preparation of washed platelet suspensions from human and rodent blood. In: *Platelets and Megakaryocytes*, vol. 1. Functional Assays (2004). p. 13–28.

frontiers | Frontiers in Oncology

TYPE Case Report
PUBLISHED 16 August 2023
DOI 10.3389/fonc.2023.1182391

Check for updates

OPEN ACCESS

EDITED BY
Miao Liu,
Harvard Medical School, United States

REVIEWED BY
Weineng Feng,
First People's Hospital of Foshan, China
Jie Qian,
Shanghai Jiao Tong University, China
Francesco Pepe,
University of Naples Federico II, Italy
Edoardo Garbo,
University of Turin, Italy

*CORRESPONDENCE
Bastian Eul
bastian.eul@innere.med.uni-giessen.de

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

RECEIVED 08 March 2023
ACCEPTED 11 July 2023
PUBLISHED 16 August 2023

CITATION
Cekay M, Arndt PF, Dumitrascu R, Savai R, Braeuninger A, Gattenloehner S, Steiner D, Roller F, Tello K, Hattar K, Seeger W, Sibelius U, Grimminger F and Eul B (2023) Case Report: Durable therapy response to Osimertinib in rare EGFR Exon 18 mutated NSCLC.
Front. Oncol. 13:1182391.
doi: 10.3389/fonc.2023.1182391

COPYRIGHT
© 2023 Cekay, Arndt, Dumitrascu, Savai, Braeuninger, Gattenloehner, Steiner, Roller, Tello, Hattar, Seeger, Sibelius, Grimminger and Eul. 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.

Case Report: Durable therapy response to Osimertinib in rare EGFR Exon 18 mutated NSCLC

Michael Cekay[1†], Philipp F. Arndt[1,2†], Rio Dumitrascu[1], Rajkumar Savai[1,2,3,4], Andreas Braeuninger[5], Stefan Gattenloehner[5], Dagmar Steiner[6], Fritz Roller[7], Khodr Tello[1], Katja Hattar[1], Werner Seeger[1], Ulf Sibelius[1], Friedrich Grimminger[1] and Bastian Eul[1*]

[1]Department of Internal Medicine, Justus-Liebig-University Giessen, Universities of Giessen and Marburg Lung Center (UGMLC), Member of the German Center for Lung Research (DZL), Giessen, Germany, [2]Max Planck Institute for Heart and Lung Research, Member of the DZL, Member of CPI, Giessen, Germany, [3]Institute for Lung Health (ILH), Justus Liebig University, Giessen, Germany, [4]Frankfurt Cancer Institute (FCI), Goethe University, Frankfurt, Germany, [5]Department of Pathology, Justus-Liebig-University Giessen, Universities of Giessen and Marburg Lung Center (UGMLC), Giessen, Germany, [6]Department of Nuclear Medicine, Justus-Liebig-University Giessen, Universities of Giessen and Marburg Lung Center (UGMLC), Giessen, Germany, [7]Department of Radiology, Justus-Liebig-University Giessen, Universities of Giessen and Marburg Lung Center (UGMLC), Giessen, Germany

Up to 20% of all non-small cell lung cancer patients harbor tumor specific driver mutations that are effectively treated with tyrosine kinase inhibitors. However, for the rare EGFR deletion-insertion mutation of exon 18, there is very little evidence regarding the effectiveness of tyrosine kinase inhibitors. A particular challenge for clinicians in applying tyrosine kinase inhibitors is not only diagnosing a mutation but also interpreting rare mutations with unclear therapeutic significance. Thus, we present the case of a 65-year-old Caucasian male lung adenocarcinoma patient with an EGFR Exon 18 p.Glu709_Thr710delinsAsp mutation of uncertain therapeutic relevance. This patient initially received two cycles of standard platinum-based chemotherapy without any therapeutic response. After administration of Osimertinib as second line therapy, the patient showed a lasting partial remission for 12 months. Therapy related toxicities were limited to mild thrombocytopenia, which ceased after dose reduction of Osimertinib. To our knowledge, this is the first report of effective treatment of this particular mutation with Osimertinib. Hence, we would like to discuss Osimertinib as a viable treatment option in EGFR Exon 18 p.Glu709_Thr710delinsAsp mutated lung adenocarcinoma.

KEYWORDS

non-small cell lung cancer, EGFR mutation, tyrosine kinase inhibitors (TKIs), Osimertinib, EGFR Exon 18 insertion

Abbreviations: CT, Computed tomography; EBUS-TBNA, Ultrasonic-guided transbronchial needle aspiration; FDG-PET, fluorodeoxyglucose positron emission tomography-computed tomography; NSCLC, Non-small-cell lung cancer; TKIs, tyrosine kinase inhibitors; MRI, magnetic resonance imaging; Mbp, megabase pair; RECIST, Response Evaluation Criteria in Solid Tumors; NGS, next generation sequencing; PD-L1, programmed death-ligand 1; PPP6C, protein phosphatase 6 catalytic subunit.

1 Introduction

Lung cancer is a global health problem as it is the most common cause of cancer related deaths worldwide. The prognosis of lung cancer remains poor, as most patients initially present with distant metastases. During the past decades, therapeutic options for lung cancer have improved. Modern treatment of lung cancer relies on multimodal therapeutic concepts and includes radiation, surgery, chemotherapy, immunotherapy and targeted therapies with kinase inhibitors. Patients with driver mutations such as mutations in EGFR, BRAF, ALK, RET, KRAS Gly12Cys, ROS1 and NTRK1/2/3 fusions have most outstandingly benefited from the development of targeted therapies. Yet, these patients represent only a minority (~20%) of the entire lung cancer patient collective (1). A major challenge for oncologists during their day-to-day clinical routine is to determine whether a rare mutational pattern in a non-small cell lung cancer (NSCLC) patient might be responsive to an unapproved tyrosine kinase inhibitor (TKI) therapy (2–4). The scarce clinical evidence available does however show that TKIs such as Afatinib and Osimertinib indeed have clinical efficacy in rare EGFR mutations (5–8). Furthermore, for many mutations it is still unclear if they even have an activating character or if they are mere incidental findings.

2 Case presentation

We present the case of a 65-year-old Caucasian male who was diagnosed with stage IV NSCLC in September 2021. The patient initially presented with symptoms of progressive dyspnea, exercise intolerability and recurring thorax pain. As the patient had a known history of cardiovascular disease, cardiac magnetic resonance imaging (MRI) was performed, revealing an incidental nodule of the left posterior inferior lung lobe. He was then referred to our lung cancer center. The patient teaches law as a professor at a university and there was no known family history of cancer. However, the patient presented with 40 pack years of cigarette smoking. The patient did not show any further risk factors for lung cancer such as exposition to asbestos, radiation or other potential hazards. Initial workup included a bronchoscopy with endobronchial ultrasound and transbronchial needle aspiration (EBUS-TBNA). However, two consecutive bronchoscopies failed to deliverer a malignant cytology sample of the tumor for further workup. A subsequent fluorodeoxyglucose positron emission tomography-computed tomography (FDG-PET-CT) revealed a hypermetabolic tumor of the left lung, various bone metastases of the spine and a singular metastasis of the left adrenal gland (Figure 1). Finally, the histology of the tumor was obtained through drainage of a pleural effusion of the left lung. Pathological examination revealed an adenocarcinoma of the lung and the initial staging of the patient resulted in cT1 cN1 cM1c and UICC IVB. Comprehensive molecular diagnostics fulfilling standards of the national Network for Genomic Medicine (Germany) were performed. Targeted next generation sequencing (NGS) with a TSO500 panel (Illumina) was performed to detect single nucleotide variants and small insertions or deletions in 523 genes recurrently affected by mutations in various cancer types. This analysis further evaluated copy number variants of 59 genes, microsatellite instability and tumor mutation burden. Additionally, the Archer FusionPlex Lung panel was used to detect fusion transcripts of 17 genes including ALK, ROS1, RET and NTRK1-3. Fluorescence *in situ* hybridization (FISH) was performed to detect MET amplifications. These studies revealed an EGFR Exon 18 mutation (p.Glu709_Thr710delinsAsp), a neomorph U2AF1 mutation and a likely inactivating mutation in

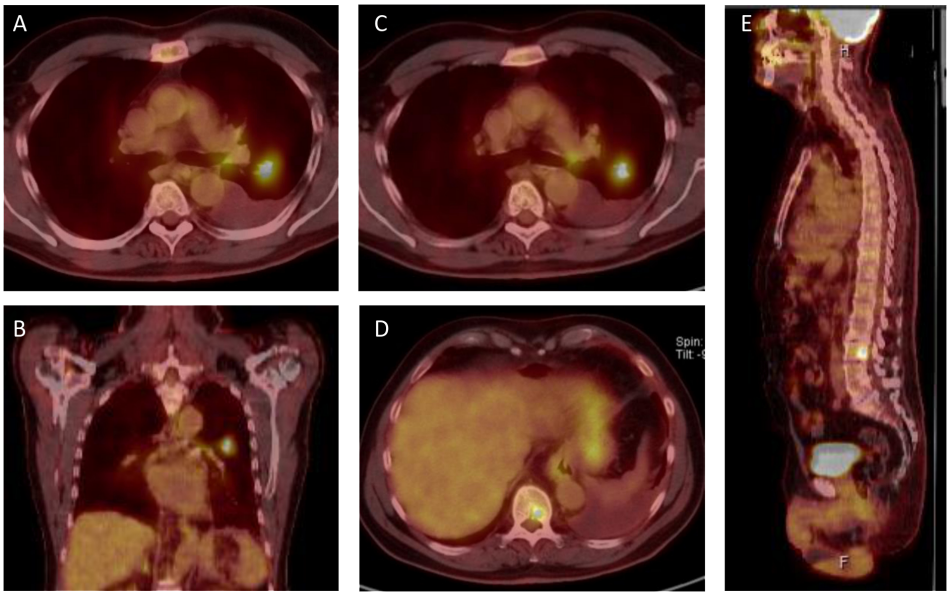


FIGURE 1
PET-CT scans from initial presentation in october 2021. **(A, B)** The patient presents with a ^{18}F-fluourodeoxyglucose (FDG)-positive lesion in the left superior lobe and an ipsilateral pleural effusion. **(C)** ipsilateral FDG positive lymphnode in the aortopulmonary window. **(D, E)** FDG positive bone lesions in the 10th thoracic vertebra **(D)** and the 2nd lumbar vertebra **(E)**, resulting in a clinical classification of T1 N1 M1c, Stage IVB, according to the 8th UICC edition.

PPC6, a negative regulator of MEK. Further, likely and known inactivating mutations in ATM, AR, DDX41 and variants of unknown significance in six further genes were detected (Table 1). No ALK, ROS1, RET or NTRK1/2/3 fusion transcripts and no MET amplifications were found. Tumor mutation burden was 8.6 variants/megabase pair (Mbp). At the same time, there was no expression of programmed death-ligand 1 (PD-L1) on tumor cells. After primary diagnosis of the NSCLC in September 2021, we initiated standard of care first-line treatment. The initial regimen was Cisplatin (75 mg/m^2) and Pemetrexed (500 mg/m^2) administered every three weeks, starting mid October 2021. The patient received two cycles of therapy in total without any major side effects. The bone metastases were additionally treated with intravenous infusions of zolendronic acid every other month, commencing in October 2021. The decision to waive radiation therapy in this patient was based on the absence of significant symptom burden associated with bone metastases such as pain or hypercalcemia. Furthermore, there were no osteolytic lesions at risk of fracturing detectable. The patient exhibited good tolerance to zolendronic acid, which was utilized as an adjunctive therapy alongside all systemic treatments thereafter. To monitor therapeutic success, we conducted a computed tomography (CT) scan in December 2021. Unfortunately, this follow-up scan revealed a progression of the primary tumor according to RECIST criteria. Although none of the distant metastases progressed, the patient's pleural effusion required more frequent drainage. As the patient furthermore suffered from severe nausea and vomiting from cisplatin, we decided to end chemotherapy and initiate TKI-therapy with Osimertinib. This decision was based on case reports previously describing the use of TKIs for this particular EGFR mutation with variable success (9–12). We began treatment with Osimertinib at the beginning of December 2021, starting with 80 mg taken orally once daily. The patient tolerated the administration of Osimertinib well and did not have any clinical signs of side effects or toxicities at first follow-up. Nevertheless, it was necessary to reduce Osimertinib dosing to 40 mg daily as the patient developed worsening thrombocytopenia (nadir of 114 giga/l) three weeks in to his TKI treatment. After dose reduction, the thrombocyte count remained stable at >120 giga/l. We conducted a short-term CT follow-up examination in January 2022, which revealed comprehensive therapeutic response of the NSCLC to Osimertinib therapy. The various bone metastases displayed increasing sclerosis compatible with a notable therapeutic response. The aforementioned pleural effusion likewise regressed. Additionally, the patient continued to tolerate Osimertinib without any further notable toxicities. Follow-up CT scans were conducted in March and August of 2022, which showed stable disease based on Response Evaluation Criteria in Solid Tumors (RECIST). However, the patient again developed a progressive pleural effusion in August 2022. The effusion was initially solely monitored using ultrasound. Regrettably, tumor progression was eventually noted on a further follow-up CT scan in November 2022 and Osimertinib therapy was discontinued. The pleural effusion was now treated with pleurodesis. Again, malignant NSCLC cells were detectable in the pleural fluid and we repeated a comprehensive pathological and molecular workup using NGS (TSO500). Here, PD-L1 status could be assessed to 5% on tumor cells in the newly acquired sample but otherwise the mutation pattern was identical to the initial analysis we conducted. As the patient reported a history of 40 pack years and no prior treatment with immunotherapy, the decision was made for a chemo-immunotherapy re-induction third line therapy regimen. The therapy was initiated in late November 2022 and consisted of Carboplatin AUC 5 (550mg absolute dose), Pemetrexed 500 mg/m^2 and Pembrolizumab 200mg administered every three weeks (Figure 2). The patient received two cycles of this treatment and tolerated it well. In January 2023, a follow-up CT scan revealed a mixed response to the applied cycles of chemo-immunotherapy. The patient again exhibited progressive pleural effusions on both sides, while the primary tumor in the left lung remained constant. The size of mediastinal lymph nodes was decreasing, but some lymph nodes in the retroperitoneal and mediastinal regions showed minor

TABLE 1 Genetic alterations detected.

Gens	Coding change	Protein change	VAF %	Functional impact
EPHA5	c.560A>G	p.Asp187Gly	34	unknown
DDX41	c.618del	p.Ile207PhefsTer15	51	likely inactivating
EGFR	c.2127_2129del	p.Glu709_Thr710delinsAsp	39	known activating
MET	c.1958C>T	p.Ser653Phe	40	unknown
PTPRD	c.3325C>T	p.Arg1109Cys	32	unknown
PPP6C	c.306del	p.Phe102LeufsTer19	13	likely inactivating
ARID5B	c.1490T>C	p.Ile497Thr	51	unknown
ATM	c.6115G>A	p.Glu2039Lys	61	known inactivating
ATM	c.9164G>C	p.Trp3055Ser	25	unknown
U2AF1	c.101C>T	p.Ser34Phe	36	known activating
AMER1	c.1143_1145del	p.Glu387del	90	unknown
AR	c.2453del	p.Pro818GlnfsTer5	72	likely inactivating

VAF, variant allele frequency.

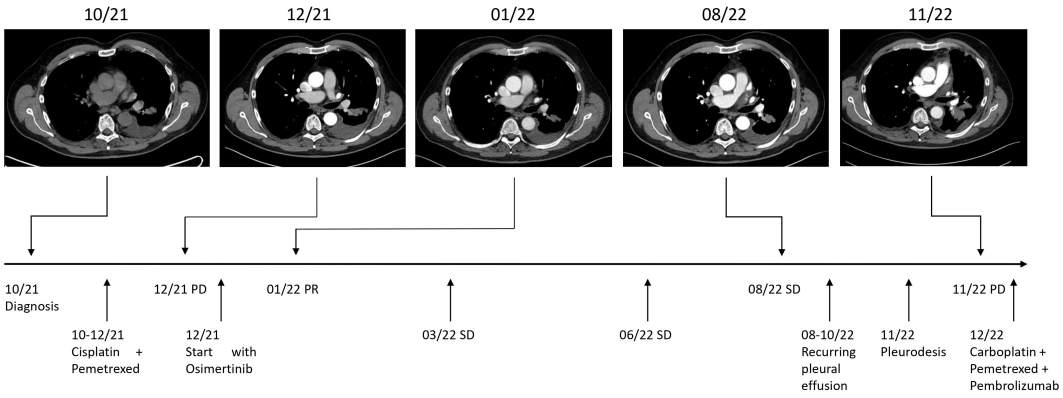


FIGURE 2
Timeline. Maximum diameters of lesion at diagnosis in 10/21: 15 x 23 mm, at PD (progressive disease) in 12/21: 18 x 30 mm, at PR (partial remission) in 01/22: 14 x 27 mm, and at PD 11/22: 37 x 41 mm. SD (stable disease).

progression. At the same time, bone lesions remained stable compared to previous CT scans and no further distant metastases were detectable. The CT examination was classified as stable disease based on RECIST. As the patient consistently showed good therapy tolerance, two additional cycles of chemo-immunotherapy were administered with unchanged dosage. In February 2023, another follow-up CT scan yet again showed a stable disease state based on RECIST. Both the primary lung tumor and lymph nodes displayed no significant changes in size. Notably, the bone metastases demonstrated progressive sclerosis further indicating therapy response. After completing four cycles of chemo-immunotherapy, Carboplatin and Pemetrexed were discontinued, while Pembrolizumab monotherapy was continued every three weeks. As of May 2023, the patient underwent another CT follow-up examination, which once more showed a stable disease state with a minor reduction of primary lung tumor size. No new distant metastases or other irregularities were observed. As of June 2023, the patient is currently continuing Pembrolizumab monotherapy.

3 Discussion and conclusion

For patients with classical EGFR mutations in exon 19 and 21 TKI therapies have had a remarkable impact on progression free and overall survival. The optimal therapy for rare EGFR Exon 18 p.Glu709_Thr710delinsAsp mutated NSCLC patients has yet not been determined. This case report emphasizes the importance of sophisticated genetic testing via NGS in NSCLC patients. Unfortunately, not all lung cancer patients receive NGS prior to therapy initiation. At the same time, growing knowledge and therapeutic possibilities with newly developed TKIs are becoming more challenging for clinicians as it becomes increasingly complex to make ideal treatment decisions. This is especially true for applying targeted therapies in rare mutations with unclear or yet unknown clinical implication. For this reason, various databases have been developed during recent years to catalog available knowledge on rare mutations and support clinicians in making appropriate treatment decisions. However, it is also imperative for clinicians to share new insights in applying targeted therapies. Numbers of patients with uncommon mutations will continue to be considerably low and structured clinical trials for such mutations will likely remain rare. Therefore, case reports may offer valuable insight for such mutations. Additionally, structured large-scale national or international investigations such as by the national network for genomic medicine in Germany or the French ERMETIC-IFCT network (13, 14) are even more important. Regarding this clinical case, it is notable that we conducted a second NGS examination with pleural fluid obtained at tumor progression in November 2022. This examination revealed an identical mutation pattern in comparison to the assay we conducted at first diagnosis. The only marker differing from the initial workup was PD-L1. Nevertheless, the tumor progressed regardless of Osimertinib therapy. Several previous case reports have been published describing the use of TKI-therapy in NSCLC patients with rare EGFR mutations. However, the effectiveness of these treatments have often been limited. Moreover, all patients in these reports received either first- or second-generation EGFR-TKIs and their clinical characteristics and demographics differed significantly from the patient described in this report. Ackermann et al. (15) presented the case of an 88-year-old female non-smoker who received Erlotinib and exhibited a partial response lasting for 4 months. Sousa et al. (10) described the case of a 66-year-old female with a smoking history who was treated with Gefitinib and showed a progression-free survival of 4 months and an overall survival of 24 months. Furthermore, Xu et al. (4) conducted an analysis of Chinese patients with various rare EGFR mutations, comparing the effectiveness of first or second generation EGFR-TKIs to chemotherapy or a combination of chemotherapy and TKIs. This study suggested that a combination of first generation TKIs and chemotherapy could be equally effective as treatment with Afatinib as a second generation TKI alone. Additionally, Wei et al. (12) reported successful treatment of EGFR Exon 18 Insertion p.Glu709_Thr710delinsAsp mutated NSCLC with Afatinib, followed by Almonertinib after tumor progression. The progression-free survival for Afatinib was 23 months, which was nearly twice as long as in our reported case. However, the patient in this report had different demographics and clinical characteristics, including Asian ethnicity and different tumor stage. Previously both

Osimertinib and Afatinib have shown efficacy in clinical trials with rare EGFR mutations. During the LUX Lung trials, patients treated with Afatinib showed an overall response rate (ORR) of up to 70% whereas patients in the UNICORN study treated with Osimertinib showed an ORR of 60% (5, 6, 8). However, none of the patients included carried an EGFR Exon 18 Insertion p.Glu709_Thr710delinsAsp mutation making it yet unclear to judge which TKI provides the greatest therapeutic benefit for this particular mutation. However, as NSCLC commonly metastasizes to the brain, we decided to implement Osimertinib instead of Afatinib due to its superior intracerebral efficacy. Our decision was furthermore based on its more favorable profile regarding adverse effects. Additionally, the use of immunotherapy as initial treatment for this patient is similarly debatable. Our decision was to refrain from administering immunotherapy as first line treatment due to the identified EGFR mutation of uncertain clinical significance. In addition, the absence of PD-L1 expression in the tumor cells of the pleural effusion likewise influenced our decision. However, the extent to which these tumor cells from the pleural fluid resemble the primary NSCLC lung tumor remains likewise debatable. Nonetheless, initiating immunotherapy upfront would have been justifiable in this case, given the patients smoking history of 40 pack years. Another point of discussion revolves around the re-induction therapy regimen following treatment failure of Osimertinib. Applying the IMPOWER150 regimen, comprising of Carboplatin, Paclitaxel, Bevacizumab and Atezolizumab would have also been a viable therapeutic option. In conclusion, the optimal treatment approach for this particular mutation remains undecided and might also depend on individual patient characteristics. To our knowledge, this is the first description of a successful therapeutic response for Osimertinib treatment in an EGFR Exon 18 p.Glu709_Thr710delinsAsp mutated NSCLC patient. This case report contributes to the understanding of this rare mutation and we would like to propose Osimertinib as a feasible treatment option.

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 review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article. The authors are accountable for all aspects of this work. Tissues were used in accordance with the ethics committee of the University of Giessen and German center for lung research.

Author contributions

BE, WS and FG initiated and supervised the study. MC, AB and PA collected data, and drafted the manuscript. BE, AB, SG, RS, WS and FG revised the manuscript for important intellectual content. All authors reviewed and edited the manuscript.

Funding

This study was supported by the German research foundation (DFG). Grant Number: GU 405/14-1.

Acknowledgments

We thank all physicians, nurses, technicians and other hospital staff involved with diagnosing and treating the patient.

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

1. Altorki NK, Markowitz GJ, Gao D, Port JL, Saxena A, Stiles B, et al. The lung microenvironment: an important regulator of tumour growth and metastasis. *Nat Rev Cancer* (2019) 19(1):9–31. doi: 10.1038/s41568-018-0081-9

2. Wu JY, Yu CJ, Chang YC, Yang CH, Shih JY, Yang PC. Effectiveness of tyrosine kinase inhibitors on "uncommon" epidermal growth factor receptor mutations of unknown clinical significance in non-small cell lung cancer. *Clin Cancer Res* (2011) 17 (11):3812–21. doi: 10.1158/1078-0432.CCR-10-3408

3. Wu SG, Chang YL, Lin JW, Wu CT, Chen HY, Tsai MF, et al. Including total EGFR staining in scoring improves EGFR mutations detection by mutation-specific antibodies and EGFR TKIs response prediction. *PloS One* (2011) 6(8):e23303. doi: 10.1371/journal.pone.0023303

4. Xu H, Yang G, Li W, Li J, Hao X, Xing P, et al. EGFR Exon 18 mutations in advanced non-small cell lung cancer: a real-world study on diverse treatment patterns and clinical outcomes. *Front Oncol* (2021) 11:713483. doi: 10.3389/fonc.2021.713483

5. Bar J, Peled N, Schokrpur S, Wolner M, Rotem O, Girard N, et al. UNcommon EGFR mutations: international case series on efficacy of Osimertinib in real-life practice in first-LiNe setting (UNICORN). *J Thorac Oncol* (2023) 18(2):169–80. doi: 10.1016/j.jtho.2022.10.004

6. Cho JH, Lim SH, An HJ, Kim KH, Park KU, Kang EJ, et al. Osimertinib for patients with non-small-cell lung cancer harboring uncommon EGFR mutations: a multicenter, open-label, phase II trial (KCSG-LU15-09). *J Clin Oncol* (2020) 38(5):488–95. doi: 10.1200/JCO.19.00931

7. Yang JC, Schuler M, Popat S, Miura S, Heeke S, Park K, et al. Afatinib for the treatment of NSCLC harboring uncommon EGFR mutations: a database of 693 cases. *J Thorac Oncol* (2020) 15(5):803–15. doi: 10.1016/j.jtho.2019.12.126

8. Yang JC, Sequist LV, Geater SL, Tsai CM, Mok TS, Schuler M, et al. Clinical activity of afatinib in patients with advanced non-small-cell lung cancer harbouring uncommon EGFR mutations: a combined post-hoc analysis of LUX-Lung 2, LUX-Lung 3, and LUX-Lung 6. *Lancet Oncol* (2015) 16(7):830–8. doi: 10.1016/S1470-2045(15)00026-1

9. Klughammer B, Brugger W, Cappuzzo F, Ciuleanu T, Mok T, Reck M, et al. Examining treatment outcomes with erlotinib in patients with advanced non-small cell lung cancer whose tumors harbor uncommon EGFR mutations. *J Thorac Oncol* (2016) 11(4):545–55. doi: 10.1016/j.jtho.2015.12.107

10. Sousa AC, Silveira C, Janeiro A, Malveiro S, Oliveira AR, Felizardo M, et al. Detection of rare and novel EGFR mutations in NSCLC patients: Implications for treatment-decision. *Lung Cancer* (2020) 139:35–40. doi: 10.1016/j.lungcan.2019.10.030

11. Van Acker L, Stevens D, Vermaelen K, Surmont V. Afatinib for the treatment of advanced non-small-cell lung cancer harboring an epidermal growth factor receptor exon 18 E709_T710delinsD mutation: a case report. *J Med Case Rep* (2021) 15(1):562. doi: 10.1186/s13256-021-02994-0

12. Wei Y, Cui Y, Guo Y, Li L, Zeng L. A lung adenocarcinoma patient with a rare EGFR E709_T710delinsD mutation showed a good response to Afatinib treatment: A case report and literature review. *Front Oncol* (2021) 11:700345. doi: 10.3389/fonc.2021.700345

13. Beau-Faller M, Prim N, Ruppert AM, Nanni-Metéllus I, Lacave R, Lacroix L, et al. Rare EGFR exon 18 and exon 20 mutations in non-small-cell lung cancer on 10 117 patients: a multicentre observational study by the French ERMETIC-IFCT network. *Ann Oncol* (2014) 25(1):126–31. doi: 10.1093/annonc/mdt418

14. Ng PK, Li J, Jeong KJ, Shao S, Chen H, Tsang YH, et al. Systematic functional annotation of somatic mutations in cancer. *Cancer Cell* (2018) 33(3):450–62.e10. doi: 10.1016/j.ccell.2018.01.021

15. Ackerman A, Goldstein MA, Kobayashi S, Costa DB. EGFR delE709_T710insD: a rare but potentially EGFR inhibitor responsive mutation in non-small-cell lung cancer. *J Thorac Oncol* (2012) 7(10):e19–20. doi: 10.1097/JTO.0b013e3182635ab4

frontiers | Frontiers in Oncology

TYPE Study Protocol
PUBLISHED 21 August 2023
DOI 10.3389/fonc.2023.1212526

Check for updates

OPEN ACCESS

EDITED BY
Hongda Liu,
Nanjing Medical University, China

REVIEWED BY
Yuting Ke,
Massachusetts Institute of Technology, United States
Cui Youbin,
Jilin University, China

*CORRESPONDENCE
Amit Mehndiratta
amit.mehndiratta@keble.oxon.org

RECEIVED 26 April 2023
ACCEPTED 31 July 2023
PUBLISHED 21 August 2023

CITATION
Baidya Kayal E, Ganguly S, Sasi A, Sharma S, DS D, Saini M, Rangarajan K, Kandasamy D, Bakhshi S and Mehndiratta A (2023) A proposed methodology for detecting the malignant potential of pulmonary nodules in sarcoma using computed tomographic imaging and artificial intelligence-based models.
Front. Oncol. 13:1212526.
doi: 10.3389/fonc.2023.1212526

COPYRIGHT
© 2023 Baidya Kayal, Ganguly, Sasi, Sharma, DS, Saini, Rangarajan, Kandasamy, Bakhshi and Mehndiratta. 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 proposed methodology for detecting the malignant potential of pulmonary nodules in sarcoma using computed tomographic imaging and artificial intelligence-based models

Esha Baidya Kayal[1], Shuvadeep Ganguly[2], Archana Sasi[2], Swetambri Sharma[2], Dheeksha DS[3], Manish Saini[3], Krithika Rangarajan[4], Devasenathipathy Kandasamy[3], Sameer Bakhshi[2] and Amit Mehndiratta[1,5]*

[1]Centre for Biomedical Engineering, Indian Institute of Technology Delhi, New Delhi, India, [2]Medical Oncology, Dr. B.R.Ambedkar Institute Rotary Cancer Hospital, All India Institute of Medical Sciences, New Delhi, Delhi, India, [3]Department of Radiodiagnosis, All India Institute of Medical Sciences, New Delhi, Delhi, India, [4]Radiodiagnosis, Dr. B.R.Ambedkar Institute Rotary Cancer Hospital, All India Institute of Medical Sciences, New Delhi, Delhi, India, [5]Department of Biomedical Engineering, All India Institute of Medical Sciences, New Delhi, Delhi, India

The presence of lung metastases in patients with primary malignancies is an important criterion for treatment management and prognostication. Computed tomography (CT) of the chest is the preferred method to detect lung metastasis. However, CT has limited efficacy in differentiating metastatic nodules from benign nodules (e.g., granulomas due to tuberculosis) especially at early stages (<5 mm). There is also a significant subjectivity associated in making this distinction, leading to frequent CT follow-ups and additional radiation exposure along with financial and emotional burden to the patients and family. Even 18F-fluoro-deoxyglucose positron emission technology-computed tomography (18F-FDG PET-CT) is not always confirmatory for this clinical problem. While pathological biopsy is the gold standard to demonstrate malignancy, invasive sampling of small lung nodules is often not clinically feasible. Currently, there is no non-invasive imaging technique that can reliably characterize lung metastases. The lung is one of the favored sites of metastasis in sarcomas. Hence, patients with sarcomas, especially from tuberculosis prevalent developing countries, can provide an ideal platform to develop a model to differentiate lung metastases from benign nodules. To overcome the lack of optimal specificity of CT scan in detecting pulmonary metastasis, a novel artificial intelligence (AI)-based protocol is proposed utilizing a combination of radiological and clinical biomarkers to identify lung nodules and characterize it as benign or metastasis. This protocol includes a retrospective cohort of nearly 2,000–2,250 sample nodules (from at least 450 patients) for training and testing and an ambispective cohort of nearly 500 nodules (from 100 patients; 50

patients each from the retrospective and prospective cohort) for validation. Ground-truth annotation of lung nodules will be performed using an in-house-built segmentation tool. Ground-truth labeling of lung nodules (metastatic/benign) will be performed based on histopathological results or baseline and/or follow-up radiological findings along with clinical outcome of the patient. Optimal methods for data handling and statistical analysis are included to develop a robust protocol for early detection and classification of pulmonary metastasis at baseline and at follow-up and identification of associated potential clinical and radiological markers.

KEYWORDS

sarcoma, lung metastases, machine learning, deep learning, artificial intelligence, malignancy, radiomics, early diagnosis

1 Introduction

The presence of lung metastases in patients with solid tumors is considered as an important criterion to direct appropriate treatment management and to further prognosticate. Computed tomography (CT) is the standard of care to detect pulmonary metastases and for staging of cancers (especially for intermediate- or high-grade tumors). Indeterminate pulmonary nodules are a very common finding and are often encountered in such clinical scenarios. However, CT is not very appropriate in differentiating metastatic nodules from benign nodules. This issue is even more glaring in developing countries in Asia and Africa where tuberculosis is highly prevalent, and it can be challenging, if not impossible, to differentiate tubercular granuloma from metastatic nodules. The benign and the malignant nodules, especially those at early stages (<5 mm) show very close resemblance to each other and there is a significant subjectiveness involved in making this distinction, requiring frequent follow-up imaging. This leads to increased financial and emotional burden over the healthcare facility and patients as well as unnecessary radiation exposure to patients. Even with the advent of 18F-fluoro-deoxyglucose positron emission technology-computed tomography (18F-FDG PET-CT), while extrapulmonary metastases may be additionally detected (1), the lung nodule conundrum remains a clinical problem. This is because not all metastatic nodules show FDG uptake on PET-CT, and at the same time, active benign nodules such as granulomas might show FDG uptake (1). In addition, small nodules (<5 mm) cannot be reliably evaluated using 18F-FDG PET-CT. Hence, currently there is no non-invasive imaging technique that can reliably characterize the malignant potential of lung nodules.

Lung is one of the favored sites of metastasis in sarcomas and 62%–83% of patients present with lung metastases during their entire disease course, with lung also being the most common site of relapse (2–6). Among patients with sarcomas of extremity, approximately 20% of patients develop isolated pulmonary metastasis at some point in their disease progression (2, 3). In patients who have multiple lung nodules detected with CT, 73% are reported to be metastases (7, 8). Hence, patients with sarcomas, especially those from low- to medium-income countries (LMICs), can provide an ideal platform to develop a model to differentiate metastases from benign nodules in the lung. Various consensus clinical guidelines exist for defining pulmonary nodules as metastatic. The Children Oncology Group definition of evident metastatic disease differs according to sarcoma type and is an area of constant evolution (9), which is left to the discretion of the treating physician (9, 10). For the subcentimeter lesions, optimal management remains unclear (11). The availability of thin slice CT technology introduced further uncertainty by detecting nodules <5 mm in diameter, thereby increasing the frequency of positive tests (12). There is no accurate non-invasive modality to determine the malignant potential of such smaller pulmonary nodules, and invasive biopsy of these small nodules is often not feasible. Moreover, in daily practice, overlapping radiologic features of metastases are frequently encountered, which makes distinction from other non-malignant pulmonary diseases difficult. For example, osteosarcoma metastases may appear as benign calcified pulmonary nodules, though as many as 40% of osteosarcoma lesions are not calcified and unusual forms of metastasis are also observed (7, 13).

To overcome the lack of optimal specificity of CT scan in detecting metastasis, it is imperative to develop a non-invasive imaging-based computational model to differentiate pulmonary nodules between benign and malignant etiologies. A lot of work has been reported for the detection and classification of primary lung cancer using machine learning (ML) and deep learning (DL) (14–20); however, predicting the malignant potential of the lung nodules in clinical scenarios of possible metastases from solid tumors is still an unmet need. Therefore, in this protocol study, we are proposing to utilize ML/DL techniques to identify lung nodules and develop a prediction model to characterize them as benign nodules or malignant lung metastases. The objectives of the proposed protocol are as follows: (a) identification of clinical and radiological markers for differentiation of lung metastases from benign lung nodules; (b) training and testing of classification model for lung metastases using retrospective training cohort; and (c) clinical validation of the developed classification model using both retrospective and prospective validation cohort.

2 Methods and analysis

2.1 Literature review on application of machine learning- and deep learning-based techniques for lung nodule characterization

Numerous studies have been reported for classification of primary lung cancer and benign lung nodules using ML/DL techniques (14–20). Extraction of effective discriminative features for lung nodule characterization is challenging due to complex anatomical structures in thoracic CT images. Many reported methods have shown improved classification performance toward this direction, and strategically, these methods can be categorized into two groups: traditional ML-based methods using feature engineering and DL-based methods.

The former methods employ feature engineering to extract various handcrafted features, like size, shape, margin, intensity, and statistical textural features in ROIs; next, identify effective discriminating features applying feature selection strategies (21, 22) and finally develop a classification model using ML algorithms, viz., logistic regression (LR), support vector machine (SVM), K-nearest neighbors (KNN), random forest (RF), linear discriminant analysis (LDA), deep neural network (DNN), and Adaboost (14–17). Various shape, margin, intensity, and textural features using GLCM in 2D axial plane and 3D volume of lung nodule (23–28), textural features using phylogenetic diversity (25), and wavelets (29, 30) were used to classify malignant and benign lung nodules using SVM, LDA, DNN, and naïve Bayes classifiers. Textural features using local binary pattern (LBP) and discrete cosine transform were extracted and used to train SVM and KNN models for lung nodule characterization (31). Shape, intensity, statistical textural features using GLCM, Gabor filters, and LBP were used to train RF models to classify lung nodules as malignant and benign (32, 33), and SVM, KNN, and LR models were used to classify different nodule types like solid, semi-solid, and ground-glass, respectively. A comparative study evaluating various ML models showed that an ensemble classifier combining SVM and RF produced the best classification performance for malignant nodules compared to KNN, LDA, and AdaBoost using shape, size, and texture-based features (34). Similarly, morphological and statistical features were applied to an ensemble of three classifiers utilizing multilayer perceptron (MLP), KNN, and SVM to classify benign or malignant lung nodules (35).

While handcrafted features need expert domain knowledge for pulmonary CT images, DL does not require explicit features extraction; it reveals intrinsic structural properties in input data by applying brain-inspired computing and has showed notable improvement in medical image analysis (18–20). Various frameworks based on convolutional neural networks (CNNs) have been developed for lung nodule detection and an elaborated review can be found in Refs (36–39). Applying reinforcement learning, region proposal network (40), faster region-based CNN (41), various advanced CNN models (42–45), and ensemble learners using multiple CNN models (46) have been designed for lung nodule detection and false-positive reduction. CNN-based dense convolutional binary-tree networks (47) and spatial pyramid dilated network (48) were developed to derive useful features from image data to discriminate malignant pulmonary nodules from benign nodules. Optimal Ensemble framework combining multiple CNN models using ResNet, AlexNet, DenceNet, InceptionNet, and SqueezeNet (49–53), transfer learning-based system (54–57), and hybrid CNN-based system (58–62) have been reported for classifying malignant and benign lung nodules. A research group proposed multi-view collaborative deep CNN models for incorporating prior knowledge about the association of nodule's malignancy and its heterogeneity (49) and further showed learning from ambiguous labels for more accurate lung nodule malignancy prediction (51). 3D Deep CNN and SVM with multiple kernel learning algorithms was proposed to fuse the DL features with clinical information for lung nodule diagnosis (61). 3D CNN and RF were used to combine CT imagery with biomarker annotation and volumetric radiomics features for lung nodule malignancy prediction (62). CNN with adaptive morphology and textural features (63), 3D segmentation attention network-based systems integrating asymmetric convolution with a gradient boosting machine (64), and 3D non-local network-based systems incorporating channel attention and adaptive network growth algorithm (65) were reported for lung nodule classification. Lung nodule classification was performed using features learned from two deep 3D customized mixed link network with gradient boosting machine (66). Studies have shown improved classification accuracy for malignant nodules using optimal deep feature selection from different CNN-based convolution layers and fusion of the deep features for the final classifier (67–69). A multi-scale cost-sensitive neural network was proposed to mitigate the issue of insufficient labeled data and class imbalance (70). A soft activation mapping-based method meta-learning scheme was reported for interpretable lung nodule classification (71) and a meta ordinal set was further generated by the same research group by developing meta ordinal weighting network to explore the ordinal relationship between the data for lung nodule classification (72). Recently, DL models based on transformers (73, 74) or combined with CNN and transformers (75, 76) have been successfully applied for lung nodule detection and classification. A self-supervised region-based 3D transformer model was developed to identify lung nodules among a set of candidate regions (73). A local focus scheme was incorporated into a deformable dilated transformer to develop a multi-granularity dilated transformer to focus on the more discriminative local features to classify lung nodules in CT scans (74). TransUnet was developed based on the transformer to encode feature representations of input CT scans and the Unet network to decode the hidden feature for outputting the final classification results (75). Res-Trans was developed using convolutional operations to extract local features and transformer blocks with self-attention to capture global features (76).

Few research gaps have been identified from the above literature survey for lung nodule characterization. Firstly, all the studies reported development of detection and classification models for primary lung cancer nodules, while characterizing metastatic lung nodules, which are also prevalent among patients with various

primary cancers, have not or rarely been explored. Secondly, it has been observed that clinical features, which are also informative for lung nodule characterization (6, 77, 78), have not been utilized by most of the reported ML and DL methods. Thirdly, temporal changes in lung nodules during treatment, which may be captured by radiological and/or radiomics features from follow-up CT scans and might be useful for predicting malignancy at early stage, have not been considered by the earlier studies. Fourthly, most of the studies have used publicly available retrospective thoracic CT datasets for training and validation of the proposed models; however, prospective datasets may be better suited to test the generalizability of the model's performance. These issues have been addressed in the current study protocol.

2.2 Study design

It is an ambispective cohort study with a retrospective training cohort and a prospective validation cohort. Patients with bone and soft tissue sarcomas who registered for treatment between January 2011 and December 2023 in the Medical Oncology Clinic of Dr. B.R.Ambedkar Institutional Rotary Cancer Hospital (IRCH), All India Institute of Medical Sciences (AIIMS) New Delhi, India will be included in this study. CT scans at baseline and follow-up with the presence of lung nodules and clinical data available in institutional databases will be used as training and testing datasets for the proposed classification model. For the validation cohort, patients with bone sarcomas and soft tissue sarcomas will be prospectively recruited from the Medical Oncology Clinic of Dr. B.R.A. IRCH, AIIMS New Delhi, India and these prospectively performed CT scans and prospectively collected clinical data will be used for the validation of the proposed classification model.

2.3 Sample size

For a two-class classification problem, considering a group ratio of 1:1, to achieve an area under the receiver operating characteristics curve of 0.8 with a 20% absolute error margin in a two-sided 95% confidence interval, a minimum of 34 samples will need to be investigated. However, for supervised learning models, studies have shown that increasing sample size beyond 1,000/class demonstrated no further significant improvements in the overall classification accuracy (79). For any patient, each annotated nodule in the lung will serve as a sample in the training/testing process during model development. In the proposed model, CT scans of a minimum of 500 patients will be used retrospectively, and considering an average of 5 nodules (range: 1–10) per CT scan, a minimum of ~2,500 samples of lung nodules will be annotated. For validation of the proposed model, CT datasets of a minimum of 50 patients will be prospectively collected. For the development of the proposed prediction model, retrospective CT datasets will be analyzed, and for validation of the proposed model, both retrospective and prospective CT datasets will be analyzed. A total of 2,000–2,250 sample nodules from at least 450 patients from the retrospective cohort will be used for training and testing the prediction model for characterizing lung metastases of sarcoma. Nearly 250 nodules from 50 patients from the retrospective cohort and 250 sample nodules from 50 prospective cohorts comprising a total of 500 nodules will be used exclusively for validating the proposed prediction model.

2.4 Data collection

Patients with bone and soft tissue sarcomas who registered for treatment during January 2011 to December 2023 in the outpatient department (OPD) of Dr. B.R.Ambedkar IRCH, AIIMS New Delhi, India will be considered for inclusion in this project.

Inclusion criteria for the retrospective cohort (registered during January 2011–December 2021) will be as follows: (a) patients with biopsy-proven bone or soft tissue sarcoma and (b) having lung nodule(s) in the chest CT scan at the time of presentation or in the course of treatment. Exclusion criteria for the retrospective cohort will be as follows: (a) patients who could not be adequately followed up and decision about lung nodule(s)'s malignancy could not be reached and (b) patients did not undergo chest CT scan in the course of treatment.

For the prospective cohort (registered during January 2022–December 2023), inclusion criteria will be as follows: (a) patients with biopsy-proven bone or soft tissue sarcoma and (b) having lung nodule(s) in the chest CT scan at the time of presentation. Exclusion criteria for the prospective cohort will be as follows: (a) patient did not undergo CT scan in the course of treatment, (b) patients whose CT scan could not be retrieved, (c) refusal for informed consent, and (d) patient for whom decision about the nature of the lung nodule(s) could not be reached within the follow-up time. For the prospective cohort, follow-up time will be at least 1 year or till dropout due to death or other reason, whichever is earlier.

For the development of the proposed classification model, a retrospective dataset of patients who registered during 2011–2018 will be used. For validation, a retrospective dataset of patients who registered during 2019–2021 and a prospective dataset of patients who will be recruited during 2022–2023 in the Medical Oncology Clinic of Dr. B.R.Ambedkar IRCH, AIIMS New Delhi, India will be used.

2.5 Ground-truth annotation and labeling of lung nodules

2.5.1 Ground-truth region of interest for lung nodule

Conventionally, the demarcation of ground-truth region of interest (ROI) for lung nodules is performed by an expert radiologist by freehand manual drawing of ROI on CT scan using annotation software after thoroughly scanning the whole CT slices. As manual annotation tasks are very tedious and time-consuming for radiologists and due to severe workload and increased number of scans, radiologists' decision-making may suffer from human error and inter- and intra-observer subjectivity. To mitigate these challenges, a semiautomated segmentation of lung nodules requiring only a minimal human input has been developed in-

house at the Center for Biomedical Engineering, IIT Delhi, India. Using the developed semi-automated segmentation tool, demarcation of ground-truth ROI for lung nodules has been initiated under the Department of Radiodiagnosis, AIIMS, Delhi, India.

The developed lung nodule segmentation algorithm is based on the region-growing method and morphological image processing algorithms. It requires only manual input for selecting a seed-point inside the nodule on a CT image and then performs automatic segmentation of the nodule across the slices of the CT image. The algorithm is capable of automatically segmenting out the lung nodules having different shapes, sizes, locations, and characteristics like solitary, juxta-pleural, ground-glass, and juxta-vascular nodules and has shown promising results in limited datasets of 50 patients for initial assessment. This segmentation tool will be used by the radiologists for ground-truth annotation of lung nodules. Examples of segmentation results for lung nodules using the developed algorithm are shown in Figure 1. This segmentation will be verified by two expert radiologists with 12 and 20 years of experience, respectively. Any disagreements will be resolved by mutual discussion to build a consensus.

2.5.2 Ground-truth labeling of lung nodule as metastatic or benign

Ground-truth labeling of lung nodules as metastatic or benign in CT scans will be performed based on a biopsy report if available or consensus decision depending on the clinical outcome of the patient and/or radiological changes in follow-up scans as summarized in Figure 2. At baseline imaging (prior to administration of chemotherapy), if consensus opinion at baseline suggests that the nodule can be clearly labeled as benign (no known malignancy, presence of calcification within the nodule in a patient with non-osteogenic sarcoma, history of tuberculosis, and presence of other features of tuberculosis), it will be labeled as benign. The label assigned at benign will be revised on the basis of follow-up imaging. After 6–12 months, if a lung nodule designated as benign at baseline remains stable, then it is assigned a definitive label of a benign lung nodule. On the other hand, if a lung nodule designated as malignant at baseline shows a size reduction after 6–12 months without undergoing chemotherapy, it is also assigned a definitive label of a benign lung nodule. If a malignant lung nodule at baseline shows a size reduction after chemotherapy or shows a size increase with or without chemotherapy, it is assigned a label of malignant lung nodule after 6–12 months. In contrast, if a malignant nodule at baseline remains stable during the next 6–12 months with or without chemotherapy, it is considered as a stable nodule and observed over the next 2–3 years for any change in size. If the size of the stable nodule continues to be stable over this time with or without chemotherapy, it is concluded as a benign nodule, likely granuloma. If the size of the stable nodule changes after 2–3 years whether being treated with chemotherapy or not, it is concluded as malignant.

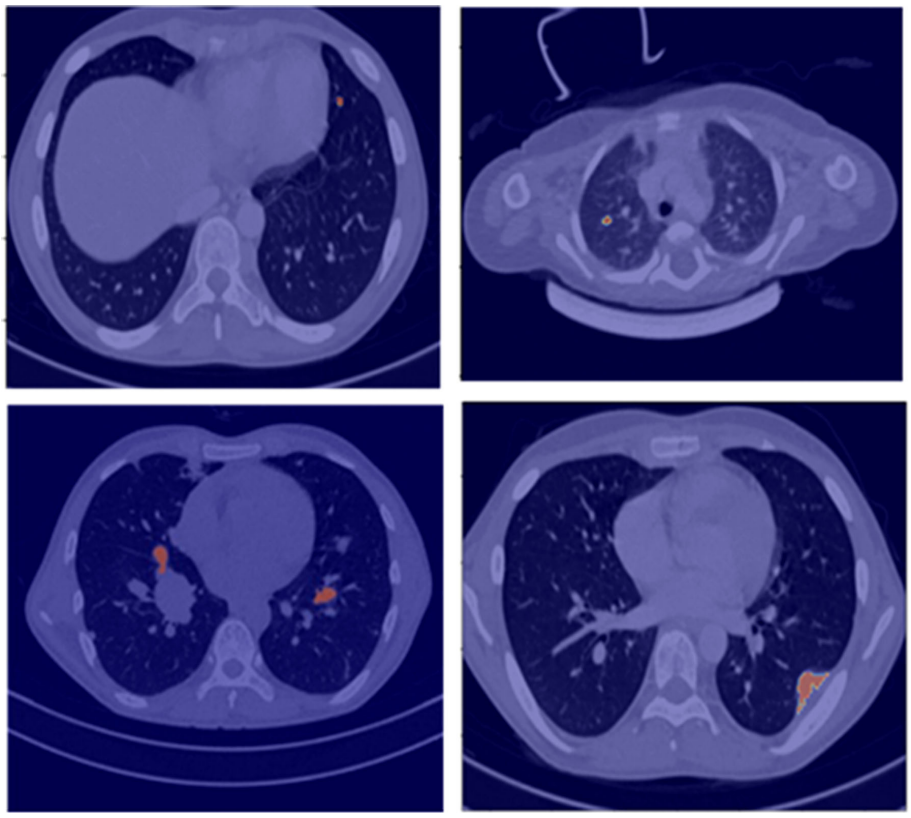


FIGURE 1
Segmentation results (red overlay) of lung nodules for a representative patient (25 years, male) with sarcoma.

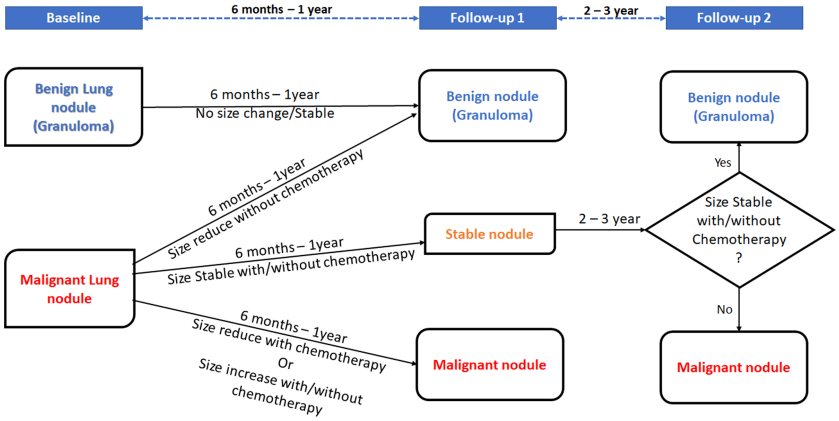


FIGURE 2
Decision rule for malignant lung nodules depending on patient outcome and/or radiological changes.

2.6 Workflow

The proposed study is composed of the following steps and summarized in the flowchart in Figure 3:

a) Retrospective data collection

Screening of available retrospective CT scans and clinical data of soft tissue sarcoma and bone sarcoma patients with known clinical follow-up details registered at Dr. B.R.Ambedkar IRCH, AIIMS New Delhi, India, between January 2011 and December 2021 will be performed. We anticipate that datasets of at least 500 patients will be collected.

b) Prospective data collection

For validation of the proposed model, CT scans and clinical data of nearly 100 patients will be prospectively collected at Dr. B.R.Ambedkar IRCH, AIIMS New Delhi, India between January 2022 and December 2023.

c) Ground-truth preparation

Ground- truth annotation and collation of all individual CT scans at baseline and follow-ups will be performed using the in-house semi-automated lung nodule segmentation tool (Figure 1). Ground-truth labeling of lung nodules as metastatic or benign will be performed using histopathological results of biopsy or metastatectomy (if available) or will be based on multi-disciplinary opinion based on baseline and/or follow-up

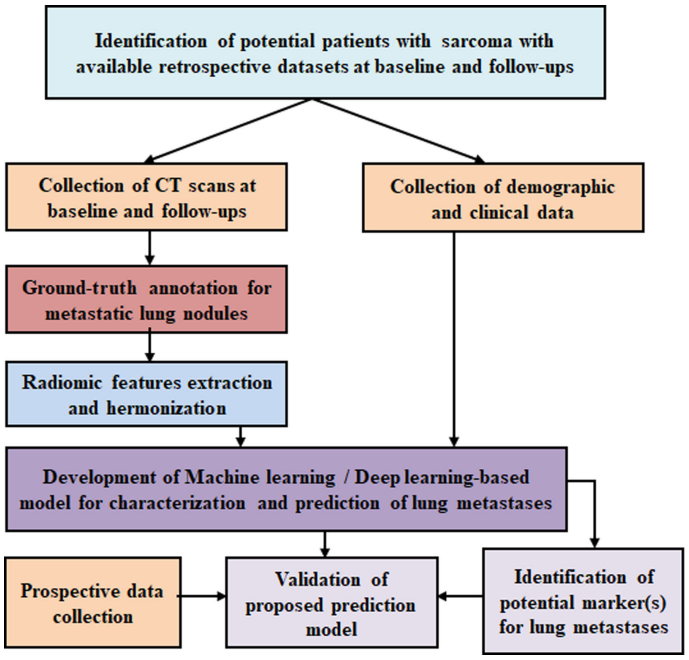


FIGURE 3
Flowchart of the proposed study protocol.

radiological findings along with the clinical outcome of the patient (Figure 2).

d) Radiological feature extraction

Radiological characteristics of lung nodules will be reviewed on CT scans at baseline and follow-up and relevant features like (i) nodule size, (ii) nodule position, (iii) change of size during follow-up, (iv) types of calcifications, (v) air bronchogram, (vi) surrounding ground-glass opacity, (vii) surrounding fibrosis, (viii) no feeding vessel sign, and (ix) bilaterality will be captured and analyzed.

Extraction of various radiomics features from the annotated ground-truth lung nodules will be performed to train the proposed classification model for identification of lung metastasis. The radiomics features consist of the following: (i) shape-based features, (ii) first-order textural features from histogram, (iii) second-order textural features from gray-level cooccurrence matrix (GLCM), and (iv) higher-order textural features from neighborhood gray-tone difference matrix (NGTDM) and run length matrix (RLM) will be evaluated. A detailed tentative list of radiomics features is listed in Table 1. Studies have shown that handcrafted features from tissues surrounding the lung nodule provide global representation of nodule CT images and can be informative for lung nodule characterization (80, 81). Therefore, the radiomics features in lung parenchyma surrounding the lung nodule will also be evaluated and analyzed. Temporal changes in semantic radiological and radiomics features in lung nodules will be evaluated from follow-up CT images (if available) during treatment and will be used along with baseline features. CT data are intrinsically dependent on the protocol acquisition parameters and pixel values in CT data are directly related to the physical characteristic of the tissues having different linear attenuation coefficients. Therefore, harmonization processes will be applied on extracted radiomics features to make them independent of scanner-specific parameter dependencies.

e) Clinical features

The historical records of the eligible patient cohort will be reviewed, and all the relevant clinical details including demographic profile, histological variants, treatment details, and disease outcomes will be captured and analyzed. Data pertaining to the following clinical characteristics and laboratory parameters at baseline will be included: (i) age at the time of diagnosis, (ii) gender, (iii) histologic subtype of sarcoma (osteosarcoma, primitive neuroectodermal tumor, or soft tissue sarcoma), (iv) histological grade, (v) site of primary disease, (vi) the presence of metastases to other sites, (vii) symptom duration prior to presentation, (viii) performance status, (ix) hemoglobin, (x) total leucocyte count (TLC), (xi) platelet count, (xii) serum lactate dehydrogenase (LDH), (xiii) serum alkaline phosphatase (ALP), (xiv) serum C-reactive protein (CRP), and (xv) serum albumin. The baseline factors may aid in estimation of the disease burden and the likelihood of lung metastases. In addition, details pertaining to chemotherapy and local site therapy modalities received (surgery and/or radiotherapy) including the timeline of receipt, therapeutic responses, and event-free survival (EFS) and overall survival (OS) will be collected. The radiological changes in the lung nodules will be assessed in the context of the timeline of therapy received and clinical responses to therapy. The degree of clinician suspicion for lung metastases based on retrospective file review will be mentioned

TABLE 1 List of radiomics features to be evaluated for lesion identification, characterization, diagnosis, monitoring prognosis or predicting outcome of patient.

SN	Radiomics Features		Description if any for easy understanding or describing importance of feature
	Methods	Features	
1	**Shape**	Volume	Volume in 3D
2		Eccentricity	Measure of aspect ratio, which is the ratio of the length of the major axis to the length of minor axis.
3		Axis of least inertia	It is defined as the line for which the integral of the square of the distances to points on the shape boundary is a minimum. It is unique to the shape.
4		Euler number	The difference between the number of contiguous parts and the number of holes on a shape.
5		Minimum bounding rectangle	Smallest rectangle that contains every point in the shape.
6		Circularity ratio	The ratio of the area of a shape to the area of a circle having the same perimeter.
7		Convexity	The ratio of perimeters of the convex hull over that of the original contour.
8		Solidity	The extent to which the shape is convex or concave.
9	**Histogram**	Mean	Average intensity of image
10		Variance	Contrast in intensity levels present in the image
11		Skewness	Measure of the asymmetry of histogram with respect to mean
12		Kurtosis	Measure of relative flatness and shape of histogram

(Continued)

TABLE 1 Continued

SN	Radiomics Features		Description if any for easy understanding or describing importance of feature
	Methods	Features	
13		Energy	Measure of uniformity; maximum value is 1 for constant image
14		Entropy	Measure of variability; minimum value is 0, for constant image
15	**Gray-Level Cooccurrence Matrix (GLCM)**	Energy	Measure of image homogeneity
16		Contrast	Measure of local image variation
17		Entropy	Measure of randomness of gray levels
18		Homogeneity	Measure of local homogeneity
19		Correlation	Measure of image linearity
20		Sum average	Measure of overall image brightness
21		Variance	Measure of spread of gray-level distribution
22		Dissimilarity	Measure the coarseness in image
23		Autocorrelation	Measure the repetitive nature of the texture primitive
24	**Neighborhood gray-tone difference matrix (NGTDM)**	Coarseness	Measure spatial rate of change in intensity, thus measure the level of coarseness.
25		Contrast	Measure intensity difference between neighboring regions
26		Busyness	Measure the rapid changes of intensity from one pixel to its neighboring regions
27		Complexity	Measure the rapid spatial changes in intensity
28		Strength	Measure the boldness or distinctiveness of the texture primitives
29	**Gray-level run length method (RLM)**	Short Run Emphasis (SRE)	Measures the distribution of short runs and is expected large for fine textures
30		Long Run Emphasis (LRE)	Measures distribution of long runs and is expected large for coarse structural textures
31		Gray-Level Nonuniformity (GLN)	Measure the gray-level nonuniformity
32		Run Length Nonuniformity (RLN)	Measure the nonuniformity of the length of runs
33		Run Percentage (RP)	Measures the homogeneity and the distribution of runs of an image in a specific direction.
34		Low Gray-Level Run Emphasis (LGRE)	Measures the distribution of low gray-level values
35		High Gray-Level Run Emphasis (HGRE)	Measures the distribution of high gray-level values
36		Short Run Low Gray-Level Emphasis (SRLGE)	Measures the joint distribution of short runs and low gray-level values.
37		Short Run High Gray-Level Emphasis (SRHGE)	Measures the joint distribution of short runs and high gray-level values.
38		Long Run Low Gray-Level Emphasis (LRLGE)	Measures the joint distribution of long runs and low gray-level values.
39		Long Run High Gray-Level Emphasis (LRHGE)	Measures the joint distribution of long runs and high gray-level values.

as “high”, “unclear”, or “low”. Clinical high suspicion of metastases is based on the progressive size of lung nodules in patients along with documented clinical deterioration or death or change of line of therapy. The degree of radiologist suspicion based on routine radiological findings will be mentioned separately by the radiologist. If there is a discrepancy between the two, a multidisciplinary discussion between the radiologist and the clinician will be carried out to reach a consensus. The ground truth as to whether the patient has lung metastases or not on imaging will be mentioned subsequently. For this study, OS will be defined as the time from treatment initiation till death from any cause. EFS will be defined as the time from treatment initiation till disease relapse/progression or death from any cause.

f) Feature selection

Dimensionality and redundancy reduction and identification of potential radiological and clinical feature(s) will be performed for

training the proposed model. Feature selection methods like Fisher's score, Lasso regularization, RF importance, and Recursive feature elimination will be applied on radiological and clinical feature sets separately for evaluating the importance of the features for the classification task and the final feature selection will be performed as the requirement of the training algorithm.

g) Machine learning/deep learning models for detection and classification of metastatic lung nodule

A number of studies have been reported for the classification of benign and malignant lung nodules using ML/DL techniques in primary lung cancer (14–20); however, classification of metastatic lung nodule using artificial intelligence (AI) has been rarely addressed. Various ML-based classification models for primary lung cancer have been reported, while SVM (23–25, 27, 29, 31, 34, 82) was applied mostly along with the other ML technologies like LR (82), KNN (31, 35, 82), RF (32, 33), LDA (27, 34), DNN (28, 30), and Adaboost (34). DL-based frameworks based on CNNs have been developed for lung nodule detection (36). Various advanced CNN models (41–45) and ensemble learners using multiple CNN models (46, 49–53), transfer learning-based systems (54–57), and hybrid CNN-based systems (58–62) have been reported for classifying malignant and benign lung nodules. CNN with adaptive morphology and textural features (63), deep feature selection from different convolution layers (67–69), and 3D segmentation attention network-based systems (64, 65) were reported for lung nodule classification. Recently, DL models based on transformers (73, 74) along with CNN (75, 76) have reported promising results for lung nodule detection and classification.

The plan in the current protocol is to train the proposed model using CT datasets of nearly 450 patients from a retrospective cohort. The aim is to detect and classify the metastatic lung nodules at baseline, i.e., even before commencement of treatment and at follow-up. A separate analysis will be carried out to train the classification model to predict the metastasis even for the smaller nodule (<5 mm) at its very early stage. For the classification model, ML-based algorithms like (i) multivariate LR, (ii) SVM, (iii) LDA, and (iv) RF and DL-based frameworks like (i) transformers with CNN, (ii) 3D CNN (CNN), (iii) recurrent CNN, and (iv) Spatial pyramid Pooling CNNs will be tested and validated. After features selection, relevant radiological features and relevant clinical features will be used separately and in combination to train the ML-based algorithms. For each ML algorithm, three separate training models will be prepared using (i) selected radiological features, (ii) selected clinical features, and (iii) selected radiological and clinical features in combination. For ML model training data normalization, data noise reduction and regularization techniques like ridge regularization and lasso regularization with k-fold cross-validation will be applied as applicable for the model to avoid overfitting and maintain generalizability of the model. The prediction accuracies of different training models for all ML-based algorithms will be compared to identify the best-performing prediction model. For DL-based models, convolutional layers will be used to extract the features from the labeled CT images, then the clinical features will be concatenated with the extracted features vectors from convolutional layers, which will be further feed to the fully connected layers to train the prediction model. Appropriate hyper-parameters like learning rate, dropout, batch size, loss function, momentum, and optimizer will be applied during DL model training. Data augmentation is an important step to overcome insufficiency of labeled data, prevent overfitting, and increase the training accuracy of a DL model. Many literatures reported that use of geometric transformations (55, 60, 65, 76, 83), kernel filtering, color and noise augmentation (50), and Generative Adversarial Networks (84, 85) for augmenting pulmonary CT data improved the accuracy for lung nodule characterization. Data augmentation may introduce inconsistency among data distribution between training and test data. To refine the trained model and maintain the model accuracy, the DL model will be trained on augmented data first, followed by data without any data augmentation. Prediction performance of applied DL-based techniques will be compared to determine the optimal prediction model balancing computational cost and accuracy. For ML-based model implementation, MATLAB® (MathWorks Inc., v2018, Philadelphia, USA) and Python version 3.9 (Python Software Foundation, https://www.python.org/) will be used, and for DL-based model development, programming environments PyTorch (https://pytorch.org/), TensorFlow (https://www.tensorflow.org/), and Keras (https://keras.io/) will be used.

h) Identification of potential marker(s) for lung metastases

The proposed model will be able to identify the potential clinical and/or radiological marker(s) for classifying the metastatic lung nodules from benign ones. The proposed prediction model for detecting lung metastases will be developed and tested for clinical and radiological marker(s) separately and in combination, and the importance of both in combination will be determined. This will help to improve the model performance, reducing the misdiagnosis and false-positive results particularly for the early-stage smaller metastatic lung nodules.

i) Validation

The validation of the proposed ML/DL-based prediction model will be performed using retrospective as well as prospective clinical and CT datasets of a total of 100 patients, 50 patients each from the retrospective and prospective cohort.

2.7 Statistical analysis

a) Student's *t*-test will be used for continuous variables and the Chi-square test will be used for categorical variables. A *p*-value of <0.05 will be taken as significant.

b) Dimensionality reduction and identification of unique features to train the classification model for lung metastasis detection.

c) ML methods to identify independent radiological and clinical feature(s) as markers of lung metastases.

d) ROC curve analysis will be used to find which feature or combination of features would best classify the lung metastases from the other existing benign lung nodules.

e) True-negative and true-positive rate, positive predictive value or precision, recall, F score, and average false-positive rate per patient will be used to evaluate the performance of the proposed models.

Data analysis and development of an ML-based classification model for lung metastases detection will be performed at the laboratory facility of the Centre for Biomedical Engineering, IIT Delhi, India.

3 Discussion

The identification of metastatic disease at presentation is vital to the clinician since the expected prognosis and treatment outcomes are different in metastatic disease from that in localized disease. Furthermore, it allows the clinician to decide the intent of treatment with greater clarity. Thus, determining the nature of lung nodules in a clinical scenario of possible pulmonary metastases is of great importance.

Out of all the medical imaging techniques, CT is considered to be one of the most effective means of detecting lung cancer early (86, 87). Clinicians need to diagnose malignant nodules accurately by reading the patient's lung CT image; however, reading a large number of CT images is not only time-consuming, and there is also a high probability of misdiagnosis. There has been a lot of research work done regarding the differentiation between benign and malignant solid lung nodules. Lung nodules can be evaluated according to diameter, area, or volume. Results from the literature agreed that volume measurement is a method with a better performance in nodule sizing, as well as in assessing the nodule's growth (88). Mehta et al. (89) added volumetric nodule measurement to an existing prediction model for malignancy estimation of nodules, showing an improvement in the number of nodule classification. There are a number of other factors like tumor image intensity, shape, and texture that help in determining probability of malignancy in lung nodules (90–97). When evaluating individuals with lung nodules, the probability of malignancy is estimated on the basis of patient-related clinical factors like primary tumor grade, tumor size, and histology type (6, 78) and nodule characteristics, including size (77). As regards morphological characteristics of nodules, besides small size, diffuse, central, laminated, or popcorn calcifications, fat tissue density and perifissural location have been recognized as indicative of benign lesions. It has been found that pleural tags and contour may be identified as independent predictors of pulmonary metastases (98) or higher mean attenuation and larger diameter are significant predictors for pulmonary metastases, while higher mean attenuation is a significant predictor for small non-calcified pulmonary metastases (99). Studies have reported that radiological semantic features like lobulation, spiculation, subtlety, calcification, and texture were relevant along with automatically calculated image-based radiomics features for malignancy classification of pulmonary nodules, which is found to be in line with current clinical practice (83, 100). Inspired by the aforementioned work, in this study, both semantic radiological and clinical features of lung nodules will be combined with radiomics features to develop the proposed classification ML/DL predictive model for metastatic lung nodule detection. Harmonization of radiomics features is an important aspect to obtain standardization and reproducibility of the developed model. In this study, extracted radiomics features will be harmonized by the methods following the recent studies (101–104) as applicable.

ML/DL has achieved a series of satisfactory results in the field of medical imaging, and it also has made great progress in the detection and classification of lung nodules (14–20) and the prediction of nodule growth (105, 106) using ML/DL techniques in primary lung cancer. However, there has been a dearth of work reported on detection of lung metastases. There is no accurate non-invasive modality available to determine the malignant potential of a smaller (<5 mm) pulmonary nodule. Furthermore, functional imaging like 18F-FDG PET/CT scans are also not sensitive for lesions <1 cm. Because this ML/DL-based system will be trained using thousands of lung scans, it will be optimized to detect tiny malignant areas that specialists might overlook in the daily busy clinical routine. On successful completion of the project, the proposed model might be capable of assisting the radiologist to make a diagnostic decision combining the radiological screening with clinical data. The proposed classification model may shift the current clinical practice paradigms by utilizing novel non-invasive radiomics features of CT in combination with ML/DL techniques for characterizing indeterminate lung nodules in the patients with sarcomas. A working model for accurate detection of lung metastases may be effectively applicable for patients with different kinds of primary malignancy as well, especially those with predilection for lung metastases. Early detection of metastatic disease may help in planning personalized treatment protocol and may improve the OS rate.

The strength of the proposed study design is in analyzing an ambispective cohort with separate derivation and validation cohort to develop the proposed prediction and classification model. Moreover, combining radiological features and clinical features with various radiomics textural features may reduce the false-positive rate and help to produce a robust model. Furthermore, information about temporal changes in lung nodules during anti-cancer treatment will also be included into the feature map that will aid the prediction of the malignant potential of lung nodules. In addition, a number of ML and DL algorithms or a hybrid model combining both techniques will be implemented and tested to deliver an optimal performance.

There are few limitations in the proposed study. For demarcation of lung nodules in CT scans, an in-house-built segmentation tool has been used that has been tested on a limited number of 50 patients; however, the accuracy of the tool will be further evaluated in due course as larger data will be tested on the project. This project is multi-institutional and will be facilitated and carried out under AIIMS New Delhi and IIT Delhi in India. Retrospective and prospective datasets of patients with sarcoma will be collected from the institutional database at AIIMS New Delhi. In the future, a multi-institutional dataset may be considered for further improvement of the developed model. A comparison between limited ML and DL classification models will be performed for evaluating best performance; however, there might be a large number of ML/DL-based algorithms that need to be considered.

4 Ethics and dissemination

The ethical approval for the use of retrospective data and collection of prospective data of the patients with sarcomas has been obtained (IEC-234/09.04.2021, RP-03/2021) from the Institute Ethics Committee, AIIMS New Delhi, India. The details of data acquisition, processing, and sharing along with risks and benefits for participating in the study will be explained to all the patients before recruitment to the proposed project. Detailed patient information sheets written in

English and Hindi (regional) languages have been reviewed and approved by the Institutional Ethics Committee.

The proposed study challenges and seeks to shift the current clinical practice paradigms by utilizing novel non-invasive radiomics features of CT combined with physiological symptoms and clinical biomarkers. As a result, misdiagnosis and false-positive rate might be reduced along with reduction in the total number of follow-up CT scans, which will subsequently reduce the time, cost, anxiety, and radiation exposure of the patients. Upon validation, the proposed model will be deployed in the local hospital settings and will be applied in the routine treatment protocol.

Ethics statement

The studies involving human participants were reviewed and approved by the Institute Ethics Committee, All India Institute of Medical Sciences New Delhi, India. Written informed consent is obtained for the prospective cohort from the participants in case of adult subjects (18 years and older), from the legal guardian in case of those below 18 years, and additionally assent from children who are 8 years and above. The Institute Ethics Committee waives off consent for the use of the participants' data for the retrospective cohort.

Author contributions

EBK, SG, KR, DK, SB, and AM conceptualized the study. Data curation and investigation were performed by EBK, SG, AS, SS, DDS, MS, and DK. Resources were provided by KR, DK, SB, and AM. Original manuscript was written by EBK. Reviewing and editing the manuscript were performed by EBK, SG, AS, SS, KR, DK, SB, and AM. All authors contributed to the article and read and approved the submitted version.

Funding

This research received external funding from Indian Council of Medical Research (ICMR), New Delhi, India (AI-Adhoc/06/2022-AI Cell) and intramural project funding under Multi-Institutional Faculty Interdisciplinary Research Project scheme in collaboration with All India Institute of Medical Sciences, Delhi, India (AI-56) and Indian institute of Technology Delhi, India (MI02654).

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

1. Singh T, Sharma A, Sharma A, Bakhshi S, Patel C, Pandey A, et al. Utility of 18F-FDG-PET/CT in management and prognostication of treatment naïve late-stage soft tissue sarcomas. *Nucl Med Commun* (2021) 42(7):818–25. doi: 10.1097/MNM.0000000000001401

2. Gadd MA, Casper ES, Woodruff JM. Development and treatment of pulmonary metastases in adult patients with extremity soft tissue sarcoma. *Ann Surg* (1993) 218 (6):705–12. doi: 10.1097/00000658-199312000-00002

3. Billingsley KG, Burt ME, Jara E, Ginsberg RJ, Woodruff JM, Leung DHY, et al. Pulmonary metastases from soft tissue sarcoma analysis of patterns of disease and postmetastasis survival. *Ann Surg* (1999) 229(5):602–12. doi: 10.1097/00000658-199905000-00002

4. Biswas B, Rastogi S, Khan S, Shukla N, Deo S, Agarwala S, et al. Hypoalbuminaemia is an independent predictor of poor outcome in metastatic Ewing's sarcoma family of tumours: a single institutional experience of 150 cases treated with uniform chemotherapy protocol. *Clin Oncol* (2014) 26(11):722–9. doi: 10.1016/j.clon.2014.05.006

5. Iqbal N, Shukla N, Deo S, Agarwala S, Sharma D, Sharma M, et al. Prognostic factors affecting survival in metastatic soft tissue sarcoma: an analysis of 110 patients. *Clin Transl Oncol* (2016) 18(3):310–6. doi: 10.1007/s12094-015-1369-9

6. Nataraj V, Rastogi S, Khan SA, Sharma MC, Agarwala S, Vishnubhatla S, et al. Prognosticating metastatic osteosarcoma treated with uniform chemotherapy protocol without high dose methotrexate and delayed metastasectomy: a single center experience of 102 patients. *Clin Transl Oncol* (2016) 18(9):937–44. doi: 10.1007/s12094-015-1467-8

7. Ciccarese F, Bazzocchi A, Ciminari R, Righi A, Rocca M, Rimondi E, et al. The many faces of pulmonary metastases of osteosarcoma: Retrospective study on 283 lesions submitted to surgery. *Eur J Radiol [Internet]* (2015) 84(12):2679–85. doi: 10.1016/j.ejrad.2015.09.022

8. Gross BH, Glazer GM, Bookstein FL. Multiple pulmonary nodules detected by computed tomography: diagnostic implications. *J Comput Assist Tomogr* (1985) 9 (5):880–5. doi: 10.1097/00004728-198509000-00007

9. Meyers P. Trial of ADR, CDDP and MTX with and without ifos, with and without MTP-PE for treatment of osteogenic sarcoma(POG-9351). (1996), COG 7921.

10. Weiss A. COGARST1321: Pazopanib neoadjuvant trial in non-rhabdomyosarcoma soft tissue sarcomas (PAZNTIS): A phase II/III randomized trial of preoperative chemoradiation or preoperative radiation plus or minus pazopanib (NSC# 737754, IN. (2014). clinicaltrials.gov NCT No: NCT02180867.

11. Picci P, Vanel D, Briccoli A, Talle K, Haakenaasen U, Malaguti C, et al. Computed tomography of pulmonary metastases from osteosarcoma : The less poor technique. A study of 51 patients with histological correlation. *Ann Oncol* (2001) 12 (11):1601–4. doi: 10.1023/a:1013103511633

12. Ginsberg MS, Panicek DM. Subcentimeter pulmonary nodules detected in patients with sarcoma. *Sarcoma* (2000) 4(1–2):63–6. doi: 10.1155/S1357714X00000116

13. Rastogi R, Garg R, Thulkar S, Bakhshi S, Gupta A. Unusual thoracic CT manifestations of osteosarcoma: review of 16 cases. *Padiatr Radiol* (2008) 38(5):551–8. doi: 10.1007/s00247-007-0735-3

14. Jassim MM, Jaber MM. Systematic review for lung cancer detection and lung nodule classification : Taxonomy, challenges, and recommendation future works. *J Intell Syst* (2022) 31:944–64. doi: 10.1515/jisys-2022-0062

15. Gu D, Liu G, Xue Z. On the performance of lung nodule detection, segmentation and classification. *Comput Med Imaging Graph* (2021) 89(January):101886. doi: 10.1016/j.compmedimag.2021.101886

16. De Margerie-mellon C, Chassagnon G. Artificial intelligence: A critical review of applications for lung nodule and lung cancer. *Diagn Interv Imaging [Internet]* (2023) 104(1):11–7. doi: 10.1016/j.diii.2022.11.007

17. Khawaja A, Bartholmai BJ, Rajagopalan S, Karwoski RA, Varghese C, Maldonado F, et al. Do we need to see to believe ?— radiomics for lung nodule classification and lung cancer risk stratification. *J Thorac Dis* (2020) 12(6):3303–16. doi: 10.21037/jtd.2020.03.105

18. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med [Internet]* (2019) 25:954–61. doi: 10.1038/s41591-019-0447-x

19. Hua K-L, Hsu C-H, Hidayati SC, Cheng W-H, Chen Y-J. Computer-aided classification of lung nodules on computed tomography images via deep learning technique. *OncoloTargets Ther* (2022) 8:2015–22. doi: 10.2147/OTT.S80733

20. Song Q, Zhao L, Luo X, Dou X. Using deep learning for classification of lung nodules on computed tomography images. *J Healthc Eng* (2017) 2017:8314740. doi: 10.1155/2017/8314740

21. Wu P, Xia K, Yu H. Correlation coefficient based supervised locally linear embedding for pulmonary nodule recognition. *Comput Methods Programs BioMed* (2016) 136:97–106. doi: 10.1016/j.cmpb.2016.08.009

22. Ciompi F, Jacobs C, Scholten ET, Wille MMW, De Jong PA, Prokop M, et al. Bag-of-frequencies : A descriptor of pulmonary nodules in computed tomography images. *IEEE Trans Med Imaging* (2015) 34(4):962–73. doi: 10.1109/TMI.2014.2371821

23. Kumar A, Sudipta D, Anirvan M, Mandeep D, Khandelwal N. A combination of shape and texture features for classification of pulmonary nodules in lung CT images. *J Digit Imaging* (2016) 29(4):466–75. doi: 10.1007/s10278-015-9857-6

24. Akram S, Javed MY, Hussain A, Riaz F. Intensity-based statistical features for classification of lungs CT scan nodules using artificial intelligence techniques. *J Exp Theor Artif Intell ISSN* (2015) 27(6):737–51. doi: 10.1080/0952813X.2015.1020526

25. De Carvalho Filho AO, Silva AC, de Paiva AC, Nunes RA, Gattass M. Computer-aided diagnosis of lung nodules in computed tomography by using phylogenetic diversity , genetic algorithm , and SVM. *J Digit Imaging* (2017) 30:812–22. doi: 10.1007/s10278-017-9973-6

26. Armato SG, Altman MB, Wilkie J, Sone S, Li F, Doi K, et al. Automated lung nodule classification following automated nodule detection on CT : A serial approach. *Med P* (2011) 30(6):1188–97. doi: 10.1118/1.1573210

27. Gong J, Liu J, Sun X, Zheng B, Niea S. Computer-aided diagnosis of lung cancer: the effect of training datasets on classification accuracy of lung nodules. *Phys Med Biol* (2018) 63(3):035036. doi: 10.1088/1361-6560/aaa610

28. Anand SKV. Segmentation coupled textural feature classification for lung tumor prediction, In: *International conference on communication control and computing technologies*, Nagercoil, India (2010). pp. 518–24. doi: 10.1109/ICCCCT.2010.5670607

29. Orozco HM, Osiris O, Villegas V, Guadalupe V, Sánchez C. Automated system for lung nodules classification based on wavelet feature descriptor and support vector machine. *Biomed Eng Online* (2015) 14(9):1–20. doi: 10.1186/s12938-015-0003-y

30. Lakshmanaprabu SK, Nandan S, Shankar K, Arunkumar N. Optimal deep learning model for classification of lung cancer on CT images. *Futur Gener Comput Syst* (2018) 92:374–82. doi: 10.1016/j.future.2018.10.009

31. Rehman A, Kashif M, Abunadi I, Ayesha N. Lung cancer detection and classification from chest CT scans using machine learning techniques, In: *2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA)*, Riyadh, Saudi Arabia (2021). pp. 101–10. doi: 10.1109/CAIDA51941.2021.9425269

32. Li X, Li B, Tian L, Zhang L. Automatic benign and Malignant classification of pulmonary nodules in thoracic computed tomography based on RF algorithm. *intitute Eng Technol* (2018) 12(7):1072–291. doi: 10.1049/iet-ipr.2016.1014

33. Wu W, Hu H, Gong J, Li X, Huang G, Nie S. Malignant-benign classification of pulmonary nodules based on random forest aided by clustering analysis. *Phys Med Biol* (2019) 64(3):035017. doi: 10.1088/1361-6560/aafab0

34. Kaya A, Can AB. A weighted rule based method for predicting Malignancy of pulmonary nodules by nodule characteristics. *J Biomed Inform* (2015) 56:69–79. doi: 10.1016/j.jbi.2015.05.011

35. Farahani FV, Ahmadi A, Hossein M, Zarandi F. Hybrid intelligent approach for diagnosis of the lung nodule from CT images using spatial kernelized fuzzy c-means and ensemble learning learning. *Math Comput Simul* (2018) 149:48–68. doi: 10.1016/j.matcom.2018.02.001

36. Monkam P, Qi S, Ma HE, Yao Y, Qian WEI. Detection and classification of pulmonary nodules using convolutional neural networks : A survey. *IEEE Access* (2019) 7:78075–91. doi: 10.1109/ACCESS.2019.2920980

37. Gu Y, Chi J, Liu J, Yang L, Zhang B, Yu D, et al. A survey of computer-aided diagnosis of lung nodules from CT scans using deep learning. *Comput Biol Med* (2021) 137:104806. doi: 10.1016/j.compbiomed.2021.104806

38. Li R, Xiao C, Huang Y, Hassan H, Huang B. Deep learning applications in computed tomography images for pulmonary nodule detection and diagnosis : A review. *Diagnostics (Basel)* (2022) 12(2):298. doi: 10.3390/diagnostics12020298

39. Riquelme D, Akhloufi MA. Deep learning for lung cancer nodules detection and classification in CT scans. *AI* (2020) Ml):28–67. doi: 10.3390/ai1010003

40. Qin Y, Zheng H, Zhu Y, Yang J, Rd D, District M, et al. Simultaneous accurate detection of pulmonary nodules and false positive reduction using 3D CNNs, In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, AB, Canada (2018). pp. 1005–9. doi: 10.1109/ICASSP.2018.8462546

41. Tang H, Kim DR, Xie X. Automated pulmonary nodule detection using 3D deep convolutional neural networks, In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. (2017). pp. 3–6. doi: 10.48550/arXiv.1903.09876

42. Winkels M, Cohen TS. 3D G-CNNs for pulmonary nodule detection, In: *1st Conference on Medical Imaging with Deep Learning (MIDL 2018)*. (2018). pp. 1–11. doi: 10.48550/arXiv.1804.04656

43. Da Silva GLF, Valente TLA, Corres SA, De Paiva AC, Gattass M. Convolutional neural network-based PSO for lung nodule false positive reduction on CT images. *Comput Methods Programs Biomed* (2018) 162:109–18. doi: 10.1016/j.cmpb.2018.05.006

44. Monkam P, Qi S, Xu M, Han F, Zhao X, Qian W. CNN models discriminating between pulmonary micro – nodules and non – nodules from CT images. *BioMed Eng Online* (2018) 17(1):1–16. doi: 10.1186/s12938-018-0529-x

45. Jia AJ, Liu BJ, Gu CY. Computer-aided diagnosis of pulmonary nodules on CT scan images, In: *2018 10th International Conference on Modelling, Identification and Control (ICMIC)*, Guiyang, China (2018). pp. 2–4. doi: 10.1109/ICMIC.2018.8529939

46. Hamidian S, Sahiner B, Petrick N, Pezeshk A, Health R, Food US, et al. 3D convolutional neural network for automatic detection of lung nodules in chest CT. *Proc SPIE Int Soc Opt Eng* (2017) 10134:1013409. doi: 10.1117/12.2255795

47. Liu Y, Hao P, Zhang P, Xu X, Wu J, Chen WEI. Dense convolutional binary-tree networks for lung nodule classification. *IEEE Access* (2018) 6:49080–8. doi: 10.1109/ACCESS.2018.2865544

48. Zhang G, Luo Y, Zhu D, Xu Y, Sun Y, Lu J. Spatial pyramid dilated network for pulmonary nodule Malignancy classification, In: *2018 24th Int Conf Pattern Recognit*, Beijing, China (2018). pp. 3911–6. doi: 10.1109/ICPR.2018.8546106

49. Xie Y, Xia Y, Zhang J, Song Y. Knowledge-based collaborative deep learning for benign-Malignant lung nodule classification on chest CT. *IEEE Trans Med Imaging* (2019) 38(4):991–1004. doi: 10.1109/TMI.2018.2876510

50. Ahmed I, Chehri A, Jeon G, Piccialli F. Automated pulmonary nodule classification and detection using deep learning architectures. *IEEE/ACM Trans Comput Biol Bioinforma* (2022) PP(X):1–12. doi: 10.1109/TCBB.2022.3192139

51. Liao Z, Xie Y, Hu S, Xia Y. Learning from ambiguous labels for lung nodule Malignancy prediction. *IEEE Trans Med Imaging* (2022) 41(7):1874–84. doi: 10.1109/TMI.2022.3149344

52. Liu H, Cao H, Song E. Multi-model ensemble learning architecture based on 3D CNN for lung nodule Malignancy suspiciousness classification. *J Digit Imaging* (2020) 1:1242–56. doi: 10.1007/s10278-020-00372-8

53. Zhang G, Yang Z, Gong L, Jiang S, Wang L, Zhang H. Classification of lung nodules based on CT images using squeeze – and – excitation network and aggregated residual transformations. *Radiol Med [Internet]* (2020) 125(4):374–83. doi: 10.1007/s11547-019-01130-9

54. Nóbrega RVMd, Peixoto SA, Silva SPPda, Filho PPR. Lung nodule classification via deep transfer learning in CT lung images, In: *2018 IEEE 31st International Symposium on Computer-Based Medical Systems Lung*, Karlstad, Sweden (2018). pp. 244–9. doi: 10.1109/CBMS.2018.00050

55. Wu R, Liang C, Li Y, Shi X, Zhang J, Huang H. Self-supervised transfer learning framework driven by visual attention for benign – Malignant lung nodule classification on chest CT. *Expert Syst Appl* (2023) 215(November 2022):119339. doi: 10.1016/j.eswa.2022.119339

56. William I, Liawatimena S, Wawan T. Lung nodule detection and classification from Thorax CT-scan using RetinaNet with transfer learning. *J King Saud Univ - Comput Inf Sci* (2022) 34(3):567–77. doi: 10.1016/j.jksuci.2020.03.013

57. Huang H, Wu R, Li Y, Peng C. Self-supervised transfer learning based on domain adaptation for benign-Malignant lung nodule classification on thoracic CT. *IEEE J BioMed Heal Informatics* (2022) 26(8):3860–71. doi: 10.1109/JBHI.2022.3171851

58. Zhao X, Liu L, Qi S, Teng Y, Li J, Qian W. Agile convolutional neural network for pulmonary nodule classification using CT images. *Int J Comput Assist Radiol Surg* (2018) 13(4):585–95. doi: 10.1007/s11548-017-1696-0

59. Nishio M, Sugiyama O, Yakami M, Ueno S, Kubo T. Computer-aided diagnosis of lung nodule classification between benign nodule, primary lung cancer, and metastatic lung cancer at different image size using deep convolutional neural network with transfer learning. *PLoS One* (2018) 13(7):1–12. doi: 10.1371/journal.pone.0200721

60. Dodia S, Basava A, Anand MP. A novel receptive field-regularized V-net and nodule classification network for lung nodule detection. *Int J Imaging Syst Technol* (2021) 32(1):88–101. doi: 10.1002/ima.22636

61. Tong C, Liang B, Su Q, Yu M, Hu J, Bashir AK, et al. Pulmonary nodule classification based on heterogeneous features learning. *IEEE J Sel AREAS Commun* (2021) 39(2):574–81. doi: 10.1109/JSAC.2020.3020657

62. Mehta K, Jain A, Mangalagiri J, Menon S, Nguyen P. Lung nodule classification using biomarkers , volumetric radiomics , and 3D CNNs. *J Digit Imaging* (2021) 34(3):647–66. doi: 10.1007/s10278-020-00417-y

63. Halder A, Chatterjee S, Dey D. Biomedical Signal Processing and Control Adaptive morphology aided 2-pathway convolutional neural network for lung nodule classification. *BioMed Signal Process Control* (2022) 72(PB):103347. doi: 10.1016/j.bspc.2021.103347

64. Guo Z, Yang J, Zhao L, Yuan J, Yu H. 3D SAACNet with GBM for the classification of benign and Malignant. *Comput Biol Med* (2023) 153(January):106532. doi: 10.1016/j.compbiomed.2022.106532

65. Al-shabi M, Shak K, Tan M. ProCAN : Progressive growing channel attentive non-local network for lung nodule classification. *Pattern Recognit* (2022) 122:108309. doi: 10.1016/j.patcog.2021.108309

66. Nasrullah N, Sang J, Alam MS, Mateen M, Cai B, Hu H. Automated lung nodule detection and classification using deep learning combined with multiple strategies. *Sensors* (2019) 19(3722):1–19. doi: 10.3390/s19173722

67. Ali I, Muzammil M, Haq IUL, Khaliq AA, Abdullah S. Deep feature selection and decision level fusion for lungs nodule classification. *IEEE Access* (2021) 9:18962–73. doi: 10.1109/ACCESS.2021.3054735

68. Çoruh Ag, Yenigün B, Uzun Ç, Kahya Y, Büyükceran E, Orhan K, et al. A comparison of the fusion model of deep learning neural networks with human observation for lung nodule detection and classification. *Br J Radiol* (2021) 94:20210222. doi: 10.1259/bjr.20210222

69. Bruntha PM, Ahmed LJ, Pandian SIA, Abraham SS. Investigation of deep features in lung nodule classification. *2022 6th Int Conf Devices Circuits Syst* (2022) :67–70. doi: 10.1109/ICDCS54290.2022.9780716

70. Xu X, Wang C, Guo J, Gan Y, Wang J, Bai H, et al. MSCS-DeepLN : Evaluating lung nodule Malignancy using multi-scale cost-sensitive neural networks. *Med Image Anal* (2020) 65:101772. doi: 10.1016/j.media.2020.101772

71. Lei Y, Tian Y, Shan H, Zhang J, Wang G, Kalra MK. Shape and margin-aware lung nodule classification in low-dose CT images via soft activation mapping. *Med Image Anal* (2020) 60:101628. doi: 10.1016/j.media.2019.101628

72. Lei Y, Shan H, Zhang J. Meta ordinal weighting net for improving lung nodule classification. *IEEE Xplor* (2021) 14(9):1210–4. doi: 10.1109/ICASSP39728.2021.9413622

73. Niu C, Wang G. Unsupervised contrastive learning based transformer for lung nodule detection Unsupervised contrastive learning based transformer for lung nodule detection. *Phys Med Biol* (2022) 67:204001. doi: 10.1088/1361-6560/ac92ba

74. Wu K, Peng B. Multi-granularity dilated transformer for lung nodule classification via local focus scheme. *Appl Sci* (2023) 13(337):1–12. doi: 10.3390/app13010377

75. Wang H, Zhu H, Ding L. Accurate classification of lung nodules on CT images using the TransUnet. *Front Public Heal* (2022) 10(1060798):1–9. doi: 10.3389/fpubh.2022.1060798

76. Liu D, Liu F, Tie Y, Qi L, Wang F. Res-trans networks for lung nodule classification. *Int J Comput Assist Radiol Surg* (2022) 17(6):1059–68. doi: 10.1007/s11548-022-02576-5

77. Gould MK, Donington J, Lynch WR, Mazzone PJ, Midthun DE, Naidich DP, et al. Evaluation of individuals with pulmonary nodules : when is it lung cancer? Diagnosis and management of lung cancer, 3rd ed : American college of chest physicians. *Chest* (2013) 143(5 Suppl):e93S–120S. doi: 10.1378/chest.12-2351

78. Xie L, Huang W, Wang H, Zheng C, Jiang J. Risk factors for lung metastasis at presentation with Malignant primary osseous neoplasms : a population-based study. *J Orthop Surg Res* (2020) 15(1):32. doi: 10.1186/s13018-020-1571-5

79. Balki I, Amirabadi A, Levman J, Martel AL, Emersic Z, Meden B, et al. Sample-size determination methodologies for machine learning in medical imaging research : A systematic review. *Can Assoc Radiol J* (2019) 70(4):344–53. doi: 10.1016/j.carj.2019.06.002

80. Calheiros JLL, de Amorim LBV, de Lima LL, de Lima Filho AF, Ferreira Júnior JR, de Oliveira MC. The effects of perinodular features on solid lung nodule classification. *J Digit Imaging* (2021) 34(4):798–810. doi: 10.1007/s10278-021-00453-2

81. Dilger SKN, Mott SL, Smith BJ, Newell JD, Hoffman EA, Sieren JC. Improved pulmonary nodule classification utilizing quantitative lung parenchyma features quantitative lung parenchyma features. *J Med Imaging* (2015) 2(4):041004. doi: 10.1117/1.JMI.2.4.041004

82. Kawathekar ID, Anu Areeckal S. Performance analysis of texture characterization techniques for lung nodule classification. *J Physics:Conference Series* (2022) 2161:012045. doi: 10.1088/1742-6596/2161/1/012045

83. Marques S, Schiavo F, Ferreira CA, Pedrosa J, Cunha A. A multi-task CNN approach for lung nodule Malignancy classification and characterization. *Expert Syst Appl* (2021) 184(September 2020):115469. doi: 10.1016/j.eswa.2021.115469

84. Apostolopoulos ID, Papathanasiou ND, Panayiotakis GS. Classification of lung nodule Malignancy in computed tomography imaging utilising generative adversarial networks and semi-supervised transfer learning. *Biocybern BioMed Eng* (2021) 41(4):1243–57. doi: 10.1016/j.bbe.2021.08.006

85. Ghosal SS, Sarkar I, Elhallaoui I. Lung nodule classification using Convolutional Autoencoder and Clustering Augmented Learning Method (CALM). *HSDM@WSDM* (2020).

86. Welter S, Grabellus F, Bauer S, Schuler M, Eberhardt W. Growth patterns of lung metastases from sarcoma : prognostic and surgical implications from histology. *Interact Cardiovasc Thorac Surg* (2012) 15(July):612–7. doi: 10.1093/icvts/ivs269

87. Digesu CS, Wiesel O, Vaporciyan AA, Yolonda L. Management of sarcoma metastases to the lung. *Surg Oncol Clin N Am* (2017) 25(4):721–33. doi: 10.1016/j.soc.2016.05.005

88. Manickavasagam R, Selvan S. Automatic detection and classification of lung nodules in CT image using optimized neuro fuzzy classifier with cuckoo search algorithm. *J Med Syst* (2019) 43(3):77. doi: 10.1007/s10916-019-1177-9

89. Jennings SG, Winer-muram HT, Tarver RD, Farber MO. Lung tumor growth : assessment with CT — Comparison of diameter and cross-sectional area with volume measurements. *Radiology* (2004) 231(3):866–71. doi: 10.1148/radiol.2313030715

90. MacMahon H, Naidich DP, Goo JM, Lee KS, Leung ANC, Mayo JR, et al. Guidelines for Management of incidental Pulmonary nodules Detected on cT images: From the Fleischner Society 2017. *Radiology* (2017) 284(1):228–43. doi: 10.1148/radiol.2017161659

91. Bankier AA, MacMahon H, Goo JM, Rubin GD, Schaefer-Prokop CM, Naidich DP. Recommendations for Measuring Pulmonary nodules at cT: A Statement from the Fleischner Society. *Radiology* (2017) 285(2):162894. doi: 10.1148/radiol.2017162894

92. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Haibe-kains B, Grossmann P, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat C* (2014) 5(1):4006. doi: 10.1038/ncomms5644

93. Cook GJR, Yip C, Siddique M, Goh V, Chicklore S, Roy A, et al. Are pretreatment 18F-FDG PET tumor textural features in non–small cell lung cancer associated with response and survival after chemoradiotherapy? *J neuclear Med* (2013) 54(1):19–26. doi: 10.2967/jnumed.112.107375

94. Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res* (2003) 3(1):1157–82. doi: 10.1162/153244303322753616

95. Larici AR, Farchione A, Franchi P, Ciliberto M, Cicchetti G, Calandriello L, et al. Lung nodules : size still matters. *Eur Respir Rev [Internet]* (2017) 26(146):170025. doi: 10.1183/16000617.0025-2017

96. Shen J, Du H, Wang Y, Du L, Yang D, Wang L, et al. A novel nomogram model combining CT texture features and urine energy metabolism to differentiate single benign from Malignant pulmonary nodule. *Front Oncol* (2022) 12:1035307. doi: 10.3389/fonc.2022.1035307

97. He C, Liu J, Li Y, Lin L, Qing H. Quantitative parameters of enhanced dual-energy computed tomography for differentiating lung cancers from benign lesions in solid pulmonary nodules. *Front Oncol* (2022) 12:11027985. doi: 10.3389/fonc.2022.1027985

98. Guo R, Yan S, Wang F, Su H, Xie Q. A novel diagnostic model for differentiation of lung metastasis from primary lung cancer in patients with colorectal cancer. *Front Oncol* (2022) 12:1017618. doi: 10.3389/fonc.2022.1017618

99. Cho YJ, Kim WS, Choi YH, Ha JY, Lee S, Park SJ, et al. Computerized texture analysis of pulmonary nodules in pediatric patients with osteosarcoma : Differentiation of pulmonary metastases from non-metastatic nodules. *PloS One* (2019) 14(2):e0211969. doi: 10.1371/journal.pone.0211969

100. Hancock MC, Magnan JF. Lung nodule Malignancy classification using only radiologist-quantified image features as inputs to statistical learning algorithms : probing the Lung Image Database Consortium dataset with two statistical learning methods. *J Med Imaging (Bellingham)* (2017) 3(4):044504. doi: 10.1117/1.JMI.3.4.044504

101. Haga A, Takahashi W, Aoki S, Nawa K, Yamashita H, Abe O, et al. Standardization of imaging features for radiomics analysis. *J Med Investig* (2019) 66(1.2):35–7. doi: 10.2152/jmi.66.35

102. Fortin J, Parker D, Tunç B, Watanabe T, Elliott MA, Ruparel K, et al. Harmonization of multi-site diffusion tensor imaging data. *Neuroimage* (2018) 161:149–70. doi: 10.1016/j.neuroimage.2017.08.047

103. Pasini G, Bini F, Russo G, Comelli A, Marinozzi F, Stefano A. matRadiomics : A novel and complete radiomics framework, from image visualization to predictive model. *J Imaging* (2022) 8(8):221. doi: 10.3390/jimaging8080221

104. Bertolini M, Trojani V, Botti A, Cucurachi N, Galaverni M, Cozzi S, et al. Novel harmonization method for multi-centric radiomic studies in non-small cell lung cancer. *Curr Oncol* (2022) 29(8):5179–94. doi: 10.3390/curroncol29080410

105. Yang R, Hui D, Li X, Wang K, Li C, Li Z. Prediction of single pulmonary nodule growth by CT radiomics and clinical features — a one- year follow-up study. *Front Oncol* (2022) 12:1034817. doi: 10.3389/fonc.2022.1034817

106. Liao R, Li A, Yan H, Lin J, Liu H, Hou Y, et al. Deep learning-based growth prediction for sub-solid pulmonary nodules on CT images. *Front Oncol* (2022) 12:1002953. doi: 10.3389/fonc.2022.1002953

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 19 October 2023
DOI 10.3389/fonc.2023.1282335

Check for updates

OPEN ACCESS

EDITED BY
Qingqing Zhu,
The First Affiliated Hospital of Soochow University, China

REVIEWED BY
Rongrong Zhou,
Central South University, China
Yu-gang Huang,
Hubei University of Medicine, China

*CORRESPONDENCE
Fancai Lai
laifancai@fjmu.edu.cn
Fengqiang Yu
martinyu@fjmu.edu.cn

†These authors have contributed equally to this work

RECEIVED 24 August 2023
ACCEPTED 06 October 2023
PUBLISHED 19 October 2023

CITATION
Zhang L, Zhang X, Guan M, Yu F and Lai F (2023) In-depth single-cell and bulk-RNA sequencing developed a NETosis-related gene signature affects non-small-cell lung cancer prognosis and tumor microenvironment: results from over 3,000 patients.
Front. Oncol. 13:1282335.
doi: 10.3389/fonc.2023.1282335

COPYRIGHT
© 2023 Zhang, Zhang, Guan, Yu and Lai. 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.

In-depth single-cell and bulk-RNA sequencing developed a NETosis-related gene signature affects non-small-cell lung cancer prognosis and tumor microenvironment: results from over 3,000 patients

Liangyu Zhang[1,2†], Xun Zhang[1,2†], Maohao Guan[1,2†], Fengqiang Yu[1,2*] and Fancai Lai[1,2*]

[1]Department of Thoracic Surgery, the First Affiliated Hospital, Fujian Medical University, Fuzhou, China, [2]Department of Thoracic Surgery, National Regional Medical Center, the First Affiliated Hospital, Fujian Medical University, Fuzhou, China

Background: Cell death caused by neutrophil extracellular traps (NETs) is known as NETosis. Despite the increasing importance of NETosis in cancer diagnosis and treatment, its role in Non-Small-Cell Lung Cancer (NSCLC) remains unclear.

Methods: A total of 3298 NSCLC patients from different cohorts were included. The AUCell method was used to compute cells' NETosis scores from single-cell RNA-sequencing data. DEGs in sc-RNA dataset were obtained by the Seurat's "FindAllMarkers" function, and DEGs in bulk-RNA dataset were acquired by the DESeq2 package. ConsensusClusterPlus package was used to group patients into different NETosis subtypes, and the Enet algorithm was used to construct the NETosis-Related Riskscore (NETRS). Enrichment analyses were conducted using the GSVA and ClusterProfiler packages. Six distinct algorithms were utilized to evaluate patients' immune cell infiltration level. Patients' SNV and CNV data were analyzed by maftools and GISTIC2.0, respectively. Drug information was obtained from the GDSC1, and predicted by the Oncopredict package. Patient response to immunotherapy was evaluated by the TIDE algorithm in conjunction with the phs000452 immunotherapy cohort. Six NRGs' differential expression was verified using qRT-PCR and immunohistochemistry.

Results: Among all cell types, neutrophils had the highest AUCell score. By Intersecting the DEGs between high and low NETosis classes, DEGs between normal and LUAD tissues, and prognostic related genes, 61 prognostic related NRGs were identified. Based on the 61 NRGs, all LUAD patients can be divided into two clusters, showing different prognostic and TME characteristics. Enet regression identified the NETRS composed of 18 NRGs. NETRS significantly associated with LUAD patients' clinical characteristics, and patients at different NETRS groups showed significant differences on prognosis, TME characteristics, immune-related molecules' expression levels, gene mutation frequencies,

response to immunotherapy, and drug sensitivity. Besides, NETRS was more powerful than 20 published gene signatures in predicting LUAD patients' survival. Nine independent cohorts confirmed that NETRS is also valuable in predicting the prognosis of all NSCLC patients. Finally, six NRGs' expression was confirmed using three independent datasets, qRT-PCR and immunohistochemistry.

Conclusion: NETRS can serves as a valuable prognostic indicator for patients with NSCLC, providing insights into the tumor microenvironment and predicting the response to cancer therapy.

KEYWORDS

NEtosis, non-small-cell lung cancer, immune, single-cell, prognosis biomarker

Introduction

Human lung cancer is among the most deadly types of cancer and is associated with a high mortality rate and morbidity (1, 2). Lung cancer can be classified as SCLC (Small-Cell Lung Cancer) or NSCLC (Non-Small-Cell Lung Cancer). Among lung cancer cases, NSCLCs constitute the majority, and lung adenocarcinomas (LUADs) are the most common NSCLC type (3). The growth rate of LUAD is generally slower than that of squamous lung cancer (LUSC). However, it begins metastasizing earlier than LUSC (4). Molecular-targeted agents and immunotherapies are currently used to treat NSCLC, which are more effective and less harmful than traditional chemotherapy and radiotherapy (5). Although immunotherapy and targeted treatments offer advantages for a small proportion of patients, survival rates remain low. Consequently, finding reliable prognosticators for LUAD, or all NSCLC in general, is imperative.

NETs (Neutrophil Extracellular Traps) are histone and proteases-coated DNA structures released by neutrophils to trap microbes, and they are formed through a process known as NETosis (6, 7). NETs may play a role in non-infectious diseases, including autoimmune disease, coagulation disorders, acute injury, and cancer. There has also been research investigating its role in malignancies such as venous thromboembolism, invasive growth, and metastasis (8). Additionally, it's also known that NETs increase tumor cells' ability to metastasize within the bloodstream by enhancing the cell cycle (9). Researchers have demonstrated that NET-DNA, which is a component of NETs, promotes cancer metastasis through CCDC25, which is a transmembrane protein (10); and it has been shown that NETs formation triggers protumorigenic inflammatory responses and activates HCC metastasis (11). NETs are also reported to shield cancer cells from immune system's attack and reduce immunotherapy's efficacy (12, 13). Despite the growing importance of NETosis in cancer diagnosis and treatment, studies about its role in LUAD are limited.

Due to the importance of NETosis to LUAD, our research focused on developing a genetic pattern involving NETosis-associated genes. In the single-cell dataset, neutrophils scored highest for NETosis. Based on the NETosis score, all the cells were divided into two classes, and the class with higher NETosis had more active cell communication. Based on 61 prognostic related differently expressed NRGs, we identified two NETosis-related subtype, exhibited distinct prognosis and TME features. Then a novel NETosis-related Riskscore consist of eignteen NRGs was developed by the Enet algorithm. Significant variations were observed among LUAD patients in different NETRS categories in terms of immune cell infiltration, clinical features, prognosis, SNV and CNV variation frequencies, responsiveness to immunotherapy, and sensitivity to drug. By comparing NETRS with 20 published gene signatures, we found that NETRS was more powerful in predicting LUAD patients' prognosis. Additionally, NETRS was able to predict the prognosis of patients with various types of NSCLC. The findings of this study suggest that NETosis may play a crucial role in developing therapeutic approaches for individuals diagnosed with LUAD, and could offer fresh perspectives and sources for future investigations into the function of NETosis in NSCLC.

Materials and methods

Data acquisition

Clinical information and bulk RNA sequencing data for LUAD and all NSCLC patients, and data on single nucleotide variations (SNV) and copy number variations (CNV) for LUAD patients, were downloaded from the TCGA website (https://portal.gdc.cancer.gov/) (14). SNV data was processed by the maftools package, and CNV data were analyzed using GISTIC2.0 (15). Ten GEO datasets for NSCLC patients, including GSE72094-LUAD, GSE31210-LUAD, GSE8894-NSCLC, GSE42127-NSCLC, GSE68465-NSCLC, GSE41271-NSCLC, GSE74777-NSCLC, 3141-NSCLC, GSE30219-NSCLC, and GSE37745-NSCLC, were obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) (16). The TCGA-NSCLC dataset comprises clinical and RNA sequencing information for LUAD

and LUSC patients. The GSE72094 and GSE31210 datasets exclusively contain data for LUAD patients. Among the remaining eight GEO datasets, GSE74777 is the only one that solely includes data for LUSC patients, while the others all encompass data for patients with various types of NSCLC. These types include LUAD, LUSC, lung basal cell carcinoma, lung large cell carcinoma, and other NSCLC subtypes. It is important to note that the majority of these NSCLC subtypes are LUAD and LUSC. The GSE127465 dataset includes single-cell RNA-sequencing information for 7 primary LUAD samples. That data was downloaded from TISCH (17) and processed as described in previous (18). Gene sets for NETosis were compiled from previously published studies (Table S1) (19–21). Based on the AUCell method, the fraction of enrichment for NETosis-related gene expression was calculated in single cells (22). The TIDE score, which is an ICB response predictor for LUAD patients, was calculated on the TIDE website (http://tide.dfci.harvard.edu) (23), and an Immunotherapy cohort phs000452 was downloaded from the TIGER database (http://tiger.canceromics.org/#/) (24). From Thorsson V's study (25), we acquired data on multi-omics, such as neoantigen load and aneuploidy.

Cell - cell communication analysis

A comparison of intercellular communication frequencies and intensities between high and low networks was carried out in R-package 'CellChat' (26).

Consensus clustering

Based on the expression of the 61 prognosis-related NETosis-Related Genes (NRGs), the R package ConsensusClusterPlus was used to effectively cluster the LUAD patients into different clusters. All LUAD patients could be clustered into two clusters, demonstrating prognostic and immunoinfiltration differences.

Differential analysis

Using Seurat's "FindAllMarkers" function, genes differentially expressed (DEGs) between two NETRS classes were identified, and genes had |log2 (fold change)| > 0.25 with adjusted *p*-value (*P*adj) <0.01 were considered as DEGs. DESeq2 package was used to identify DEGs between normal and LUAD tumor samples, and genes with *P*adj < 0.01 and |log2(Fold-change)|> 1 were included.

Enrichment analysis

In order to understand the biological functions and potential signaling pathways associated with genes related to NETosis, GO and KEGG enrichment analysis were utilized. In order to uncover potential prognostic mechanisms related to NETRS, we performed enrichment analyses using GO, KEGG, and GSVA for genes with significant associations. GO and KEGG analyses were performed with the R package 'ClusterProfiler' (27), and GSVA was performed by 'GSVA' package (28). The sets of reference were named 'c5.all.v7.0.symbols.gmt'.

Immune infiltration analysis

A total of six algorithms were used to measure the extent of immune cell infiltration in the TCGA-LUAD dataset. TIMER, quantTIseq, MCP-counter, EPIC, and ESTIMATE algorithms are implemented using the R package 'IOBR' (29), while the ssGSEA algorithm is implemented by the R package "GSVA". We also compared immune-related molecule expression between patients with high and low NETRS in the TCGA-LUAD dataset.

Construction of the NETosis-related Riskscore

Genes associated with survival prognosis are identified through univariate Cox analysis, and NETRS was constructed through elastic network (Enet) algorithm. According to the C-Index, we adjust the α value in the Enet algorithm between 0.1 and 0.9, and finally determined 0.1 as the optimal α value. Based on the median NETRS value, patients with NSCLC were categorized into groups of high risk and low risk. PCA analysis was performed using the R package 'status', while the generation of time-dependent ROC curve was accomplished using the 'survminer' and 'timeROC' packages. In addition, through the R package 'rms', we construct a nomogram by combining NETRS with clinical factors. The calibration curve, time-ROC curve, and DCA curve were used to evaluate the nomogram.

Predicting potential drugs target NETRS

The GDSC1 database (https://www.cancerrxgene.org/) (30) provides information about drug sensitive data and corresponding gene expression matrix. Based on the R package 'oncoPredict' (31), we calculated the IC50 value, chich is an indicator of drug sensitivity, for each sample.

Validation of six NRGs' expression

Sample Collection: Two datasets (GSE19188 and GSE43458) containing LUAD tissues and paracancerous tissues were downloaded from the GEO database. In addition, anatomical samples involving 8 pairs of LUAD and corresponding paracancerous tissues were collected from the First Affiliated Hospital of Fujian Medical University. All patients provided written informed consent, and the study protocol was approved by the Ethics Committee of the First Affiliate Hospital of Fujian Medical University. qRT-PCR analysis: Total RNA was extracted using an RNA extraction kit (Vazyme, China) following the

manufacturer's instructions. The extracted RNA was then reverse transcribed into cDNA using the All-in-One First-Strand Synthesis MasterMix kit (iScience, China). For qRT-qPCR analysis, triplicate aliquots of each cDNA sample were prepared using Taq SYBR® Green qPCR Premix (iScience, China). The internal reference gene used in this study was β-Actin. The primers of the six NRGs and the internal reference gene were shown in Table S3.

Statistic analysis

R (version 4.1.1) was used for all statistical analysis. Using the Wilcoxon test or the t-test, the disparity between two groups was compared. In correlation analyses, Pearson or Spearman correlation coefficients were used. The K-M analysis was employed to forecast the disparity in overall survival between the low and high NETRS categories. Multivariate Cox regression analysis was conducted to examine the predictive significance of NETRS and clinicopathological features. If Ns- P is greater than or equal to 0.05, *- P is less than 0.05, **- P is less than 0.01, and ***- P is less than 0.001.

Results

Exploring NETosis-related cell type

In the GSE127465 single-cell sequencing dataset, we obtained a total of 26,655 cells after initial quality control. Based on the UMAP map, we observed seven samples with a relatively uniform distribution of cells, indicating no obvious batch effects (Figure 1A). The cells were then clustered into 25 clusters (Figure 1B), and the meta-data from TISCH database was utilized to identify 12 cell types (Figure 1C). Further analysis using the AUCell algorithm revealed that neutrophils exhibited the highest NETosis Score (Figures 1D, E). Subsequently, we divided each cell population into two groups, High-AUC and Low-AUC, based on their mean NETosis Score (Figure 1F). Examination of cell-cell communication showed that cells with a high NETosis Score demonstrated more frequent and stronger communication compared to cells with a low NETosis score (Figures 1G, H).

Consensus clustering identified two NETosis-related subtypes

With the help of the 'FindAllMarkers' function in the Seurat package, we identified differently expressed genes (DEGs) between high and low NETosis classes. GO analysis shows that these DEGs are mainly involved in functions related to "neutrophil activation", "myeloid leukocyte migration", "neutrophil chemotaxis", "neutrophil migration", "immune response-regulating signaling pathway" and "cytokine mediated signaling pathway" (Figure 2A); KEGG shows that these DEGs related signaling pathways including "neutrophil extracellular trap formation", "TNF signaling pathway", "Th17 cell differentiation", "Apoptosis", "IL-17 signaling pathway", and "Leukocyte transendothelial migration" (Figure 2B). As a result, these DEGs have a close connection to NETosis. To identify key genes among these DEGs, we intersected them with DEGs between normal tissues and LUAD in bulk datasets, and also with genes whose *P*-value in univariate cox regression were less than 0.04, and finally acquired 61 genes (Figure 2C). Most of the 61 genes were positively correlated, but some were negatively correlated, according to the correlation heatmap (Figure 2D). Next, we observed that the clustering effect was most optimal when k=2 (Figure 2E), allowing us to group all LUAD patients into two clusters based on these 61 genes. Most of these 61 genes exhibited high expression in cluster 1 (Figure 2F), which was associated with significantly better prognosis (Figure 2G), earlier stages, and lower mortality events (Figure 2H) compared to cluster 2. Furthermore, ssGSEA analysis indicated that cluster 1 had significantly higher infiltration of immune cells compared to cluster 2 (Figure 2I). Therefore, patients in these two NETosis clusters exhibited distinct characteristics.

Construction and validation of the NETosis-related Riskscore

To better predict patients' survival based on NETosis-related genes, we performed Elastic Network on these 61 prognostic related genes. By utilizing the TCGA-LUAD as the training dataset, it was observed that the prognostic model achieved its highest C-Index when the α value was set to 0.1 (Figure 3A). Therefore, we obtain a NETosis-Related Riskscore (NETRS) consisting of 18 NETosis-Related Genes (Figure 3B) and their corresponding coefficients (Figure 3C) by the Enet algorithm (α=0.1).

$$\begin{aligned}NETRS=&ALDH2*(-0.01837867)+ALOX5AP*(-0.01871386)+CCT6A*\\&(0.05085945)+CD69*(-0.01867372)+CKAP4*(0.02313153)+DDIT4*\\&(0.05508357)+DOCK4*(-0.03075094)+ERO1A*(0.09426655)+FBP1*\\&(-0.03484596)+FKBP4*(0.07161017)+KRT8*(0.06435152)+LDHA*\\&(0.10353750)+MS4A1*(-0.05941937)+S100P*(0.02683835)+SEC14L4*\\&(-0.02788501)+SLC16A3*(0.03103659)+SNX30*(-0.03191062)+UBE2S*\\&(0.01091161)\end{aligned}$$

The median NETRS value was used to categorize patients into two groups. This showed that patients in the high-NETRS group had a significantly worse prognosis than patients in the low-NETRS group, not only in the training set TCGA-LUAD (Figure 3D), but also in two external validation sets, namely GSE72094-LUAD (Figure 3E), and GSE31210-LUAD (Figure 3F). Based on the time-ROC curves, NETRS was able to predict patients' prognoses at 1, 3, and 4 years with an AUC value greater than 0.7, indicating satisfactory predictive ability (Figure 3G-I). Additionally, patients with high NETRS experienced a higher rate of mortality (Figure 3J-L), and the PCA analysis revealed noticeable differences between patients with a high and low NETRS (Figure 3M-O).

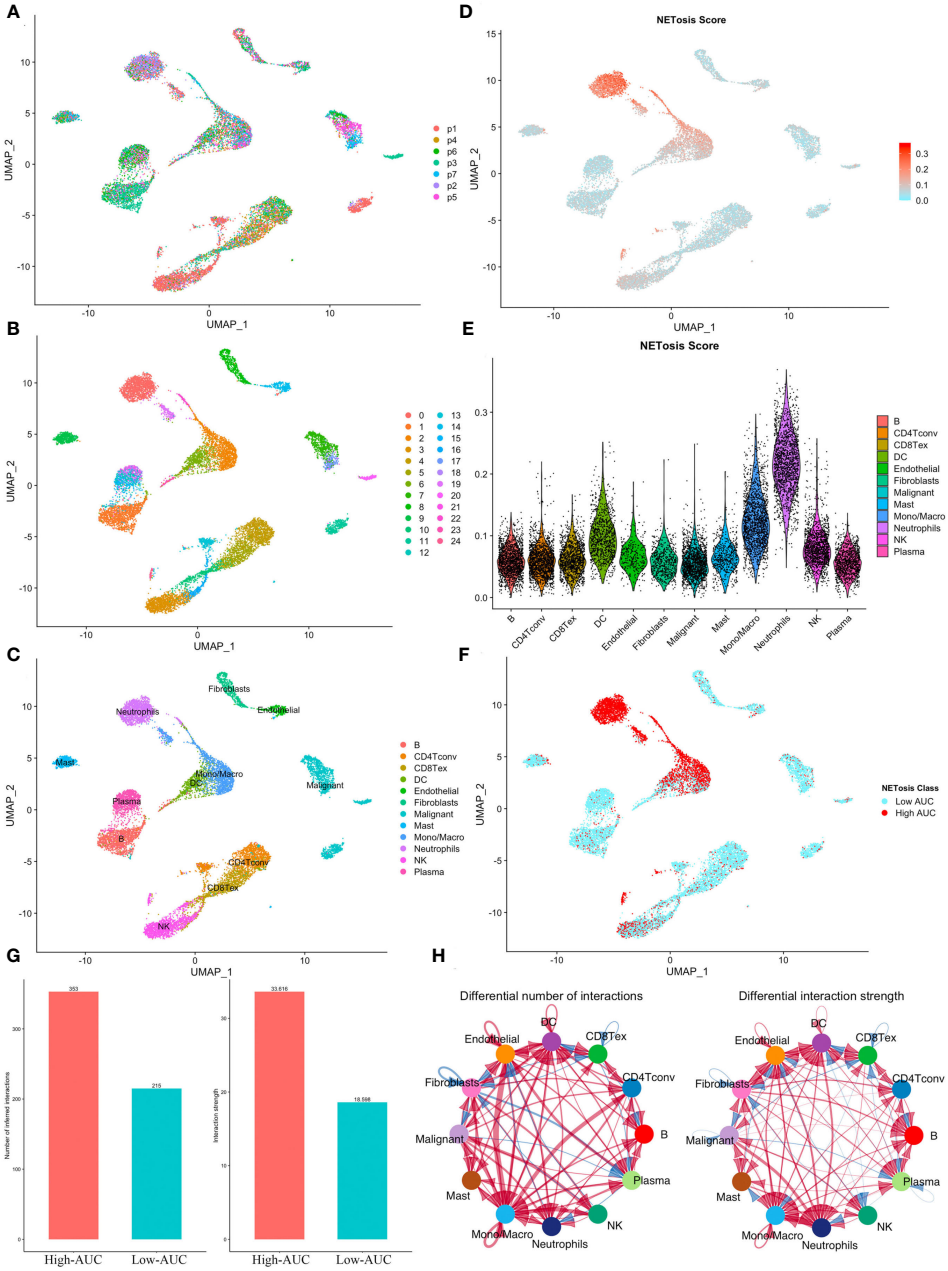


FIGURE 1
Identification of NETosis related active cells. **(A)** The UMAP plot shows the distribution of cells from 7 samples. **(B, C)** Cells were clustered into 25 clusters **(B)** and identified as 12 different types **(C)**. **(D)** The distribution of NETosis Score at the UMAP map. **(E)** Violin plot shows that neutrophils has highest NETosis Score. **(F)** Cells were classified into two clusters based on their NETosis score. **(G, H)** Comparison of the number and intensity of interaction between high and low NETosis Scores' cells using the bar plot **(G)** and the network diagram **(H)**.

NETRS strongly correlates with LUAD patients' clinical characteristics

The heatmap illustrates the expression patterns of the 18 genes comprising NETRS in the TCGA (Figure 4A), GSE72094 (Figure 4B), and GSE31210 (Figure 4C) datasets. We observed that in all datasets, ALDH2, ALOX5AP, CD69, DOCK4, FBP1, MS4A1, SEC14L4, and SNX30 were highly expressed in the low-NETRS group, while the remaining 10 genes were highly expressed in the high-NETRS group. In the TCGA-LUAD cohort, patients in the high-NETRS group exhibited a higher proportion of advanced T Stage (Figure 4D), N Stage (Figure 4E), Clinical Stage (Figure 4F), and deceased Survival status (Figure 4G). Furthermore, in the GSE72094 and GSE31210 cohorts, NETRS also increased with advanced stages (Figures 4H, J) and deceased status (Figures 4I, K). Additionally, we discovered that in the TCGA cohort, patients' Progress Free Survival (PFS) decreased as their NETRS increased (Figure 4L). Similarly, in the GSE31210 cohort, patients' Relapse Free Survival (RFS) followed a similar trend (Figure 4M). Surprisingly, in the GSE31210 cohort, NETRS achieved an

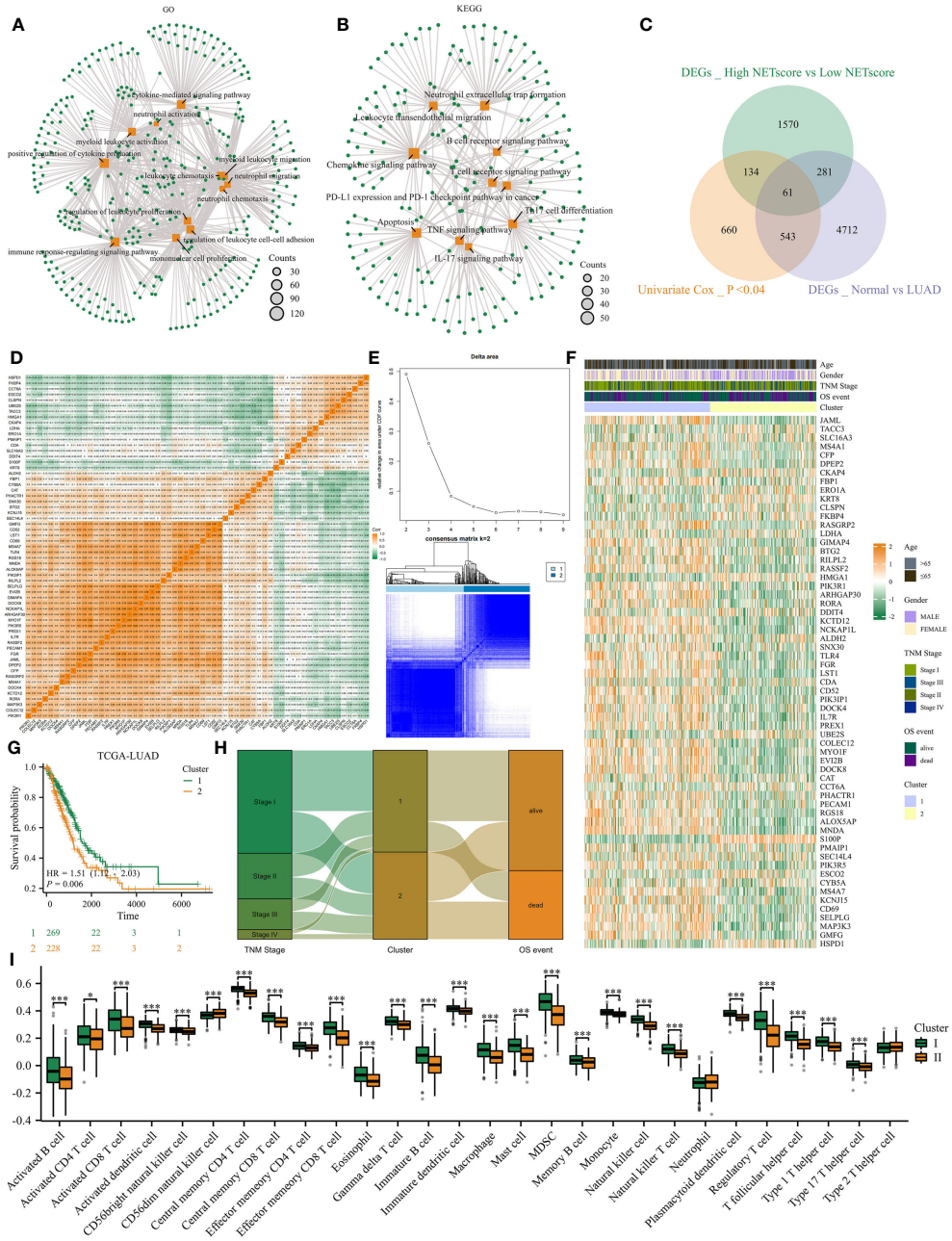


FIGURE 2
Consensus clustering identified two NETosis-related clusters. **(A, B)** GO **(A)** and KEGG **(B)** analysis on DEGs between high- and low- NETosis classes. **(C)** The Venn diagram. **(D)** The correlation heatmap illustrates the 61 genes' relationship with each other. **(E)** When k=2, the cluster effect was best. **(F)** The 61 genes' expression map between two clusters. **(G)** Two cluster patients' different prognosis. **(H)** The sankey graph shows clusters' association with clinical features. **(I)** Comparison of the immune infiltration level between cluster1 and cluster2, '*' means P value is less than 0.05, '***' means P value is less than 0.001.

impressive predictive power for patients' 1-year RFS, with a score of 0.943 (Figure 4M), highlighting its excellent predictive capability.

Comparison of NETRS with previously published prognostic models

In order to demonstrate the superiority of NETRS, we compared its predictive power with 20 previously published prognostic gene signatures for LUAD. These signatures consisted of functional genes related to lactic acid metabolism (PMID:36275729), mitotic spindle (PMID:37266661), autophagy (PMID:35529878), inflammation (PMID:35069695), cuproptosis (PMID:36353226), pyroptosis (PMID:36437954), and apoptosis (PMID:35571020). Genes' coefficients and calculation formulas for each signature can be found in the respective articles. Then we extracted the gene signatures from the found candidate studies and applied them to our three study cohorts (TCGA-LUAD, GSE72094-LUAD, and

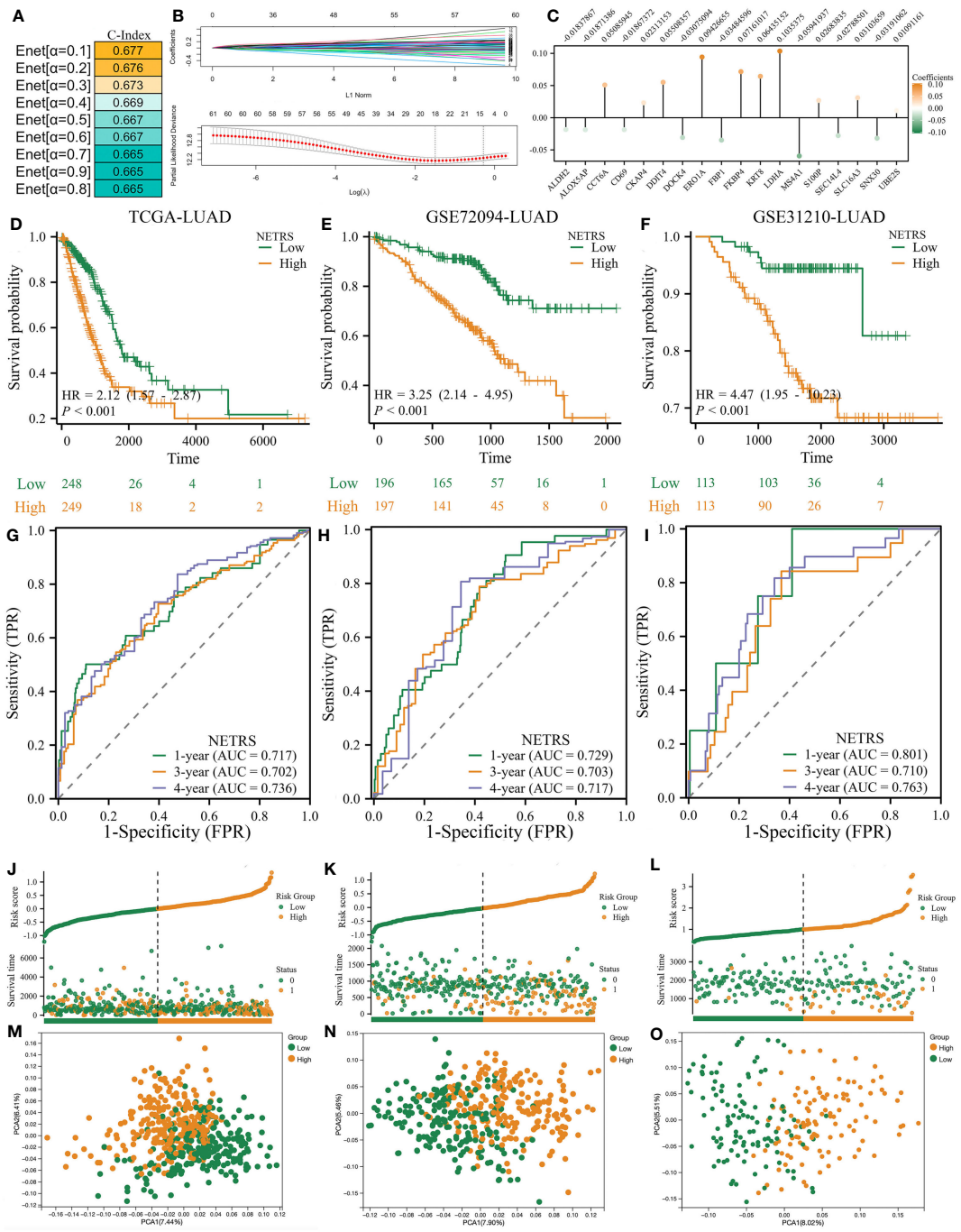


FIGURE 3
The NETRS was identified and confirmed. **(A)** When setting the α value as 0.1, the Enet model gets its highest C-index. **(B, C)** The Enet algorithm (α=0.1) identified 18 NETosis-related genes **(B)** and their corresponding coefficients **(C)**. **(D–F)** Patients' different prognosis between high- and Low-NETRS groups in TCGA **(D)**, GSE72094 **(E)**, and GSE31210 **(F)** sets. **(G–I)** The time-ROC curves shows the AUC value of NETRS in predicting patients' survival in TCGA **(G)**, GSE72094 **(H)**, and GSE31210 **(I)** sets. **(J–L)** Patients' different OS events between high- and low- NETRS groups in TCGA **(J)**, GSE72094 **(K)**, and GSE31210 **(L)** sets. **(M–O)** The PCA analysis showed that two NETRS group patients' characteristics differed extinct in the TCGA **(M)**, GSE72094 **(N)**, and GSE31210 () sets.

GSE31210-LUAD). By calculating the C-Index for each gene signature, it was observed that NETRS exhibited the highest C-Index among the three cohorts (Figures 5A, D, G). Univariate Cox regression analysis revealed that NETRS exhibited the highest HR value, indicating a higher risk compared to these signatures (Figures 5B, E, H). Moreover, in terms of predicting patients' prognosis at 1- and 2- years, NETRS outperformed almost all signatures, as indicated by the higher AUC value (Figures 5C, F, I). In the GSE31210 cohort, NETRS demonstrated slightly weaker performance than Li, Zhao et al.'s signatures in predicting patients'

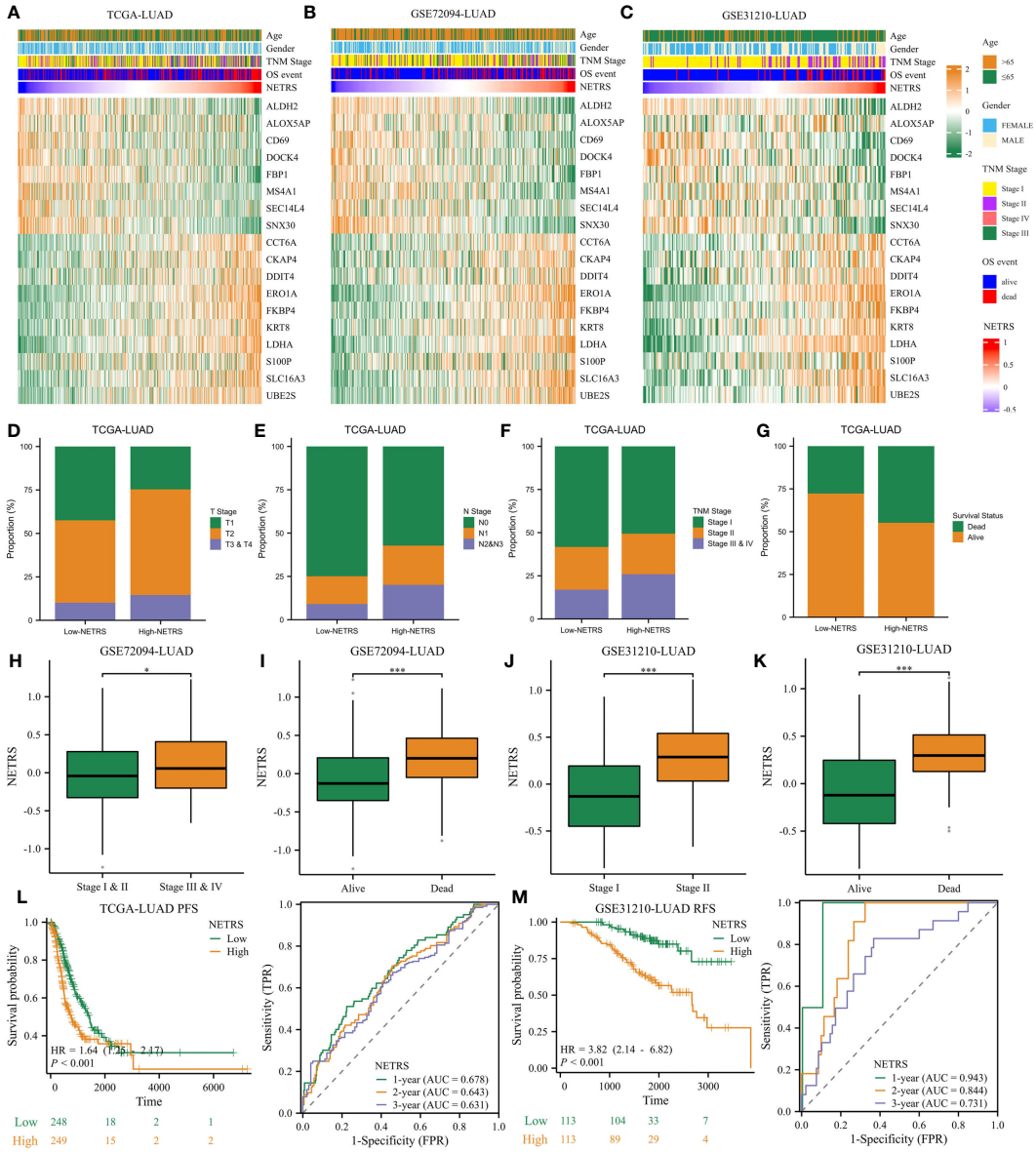


FIGURE 4
NETRS strongly correlates with LUAD patients' clinical features. **(A–C)** Heatmaps showed the expression pattern of the 18 genes that make up NETRS in the TCGA **(A)**, GSE72094 **(B)**, and GSE31210 **(C)** sets. **(D–G)** LUAD patients in the high-NETRS group exhibited higher proportion of advanced T **(D)**, N **(E)**, pathologic Stage **(F)**, and lethal OS event **(G)** in the TCGA-LUAD cohort. **(H, I)** In GSE72094 cohort, patients' NETRS increased with Stage progression **(H)** and lethal OS event **(I)**. **(J, K)** In GSE31210 cohort, patients' NETRS increased with Stage progression **(J)** and lethal OS event **(K)**. **(L)** NETRS's efficacy in predicting LUAD patients' PFS in TCGA cohort. **(M)** NETRS's efficacy in predicting LUAD patients' RFS in GSE31210 cohort. '*' means P value is less than 0.05, '***' means P value is less than 0.001.

1-year survival (Figure 5I). However, in other cohorts, the performance of Li, Zhao et al.'s signatures were significantly lower than that of NETRS, suggesting a chance effect. Additionally, by utilizing two machine learning algorithms, namely RandomForest and Boruta, we have further validated the importance of NETRS. Both algorithms consistently ranked NETRS as the most significant feature among all 21 signatures (Figures 5J–O). In general, NETRS exhibited superior predictive capabilities for the prognosis of LUAD patients compared to the other 20 gene signatures, thereby establishing itself as a more dependable prognostic indicator.

Construction and validation of a nomogram

NETRS was evaluated as an independent prognostic factor using univariate and multivariate Cox regression analysis. The results demonstrated that even after accounting for clinical factors, NETRS still significantly impacted the prognosis of LUAD patients. This suggests that NETRS can serve as an independent prognostic indicator, not only in the TCGA-LUAD cohort (Figures 6A, B), but also in the GSE72094 (Table 1) and GSE31210 (Table 2) cohorts. By

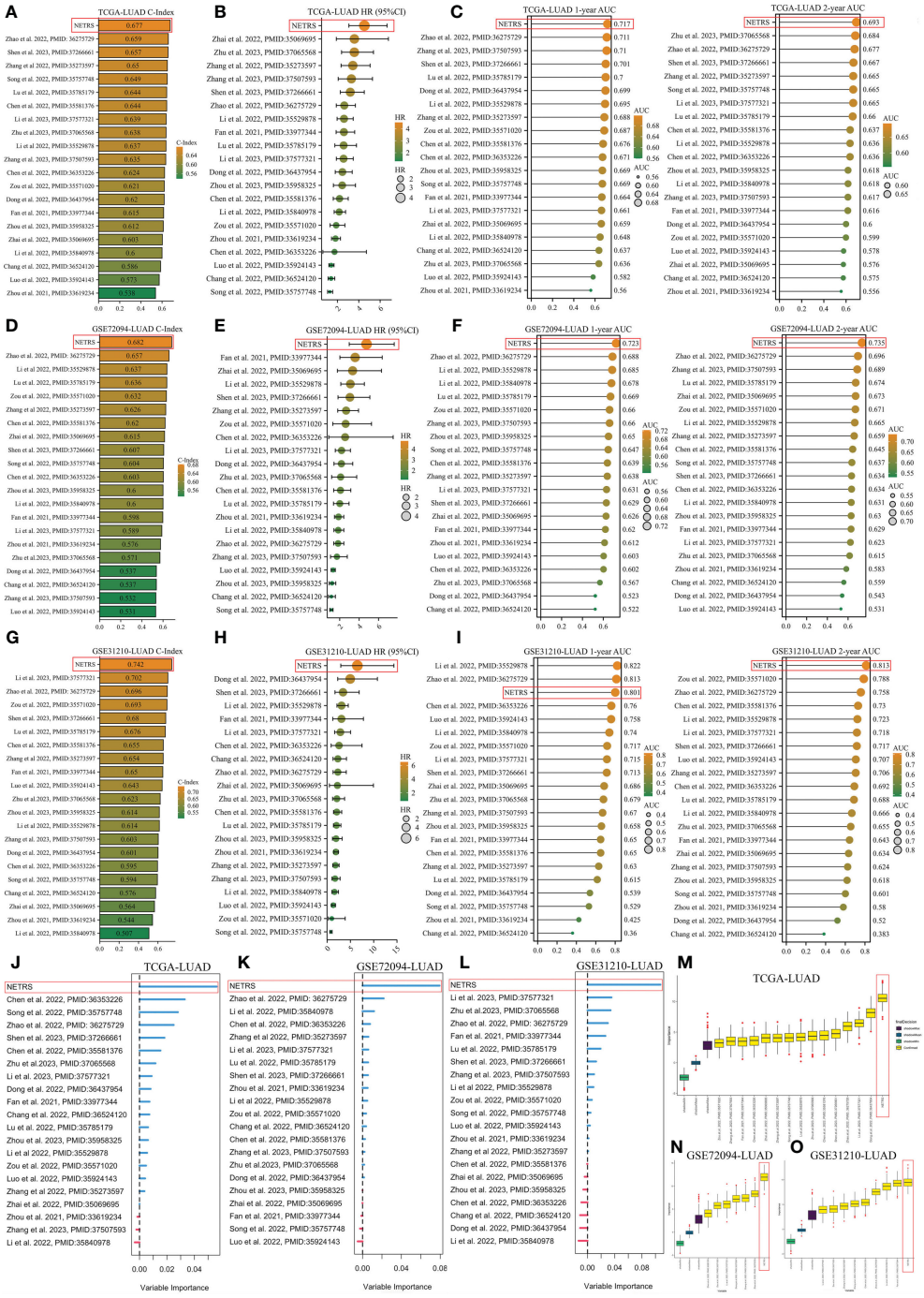


FIGURE 5
Comparing NETRS with previously published 20 gene signatures. **(A, D, G)** Comparison of the C-Index of NETRS and 20 gene signatures in TCGA **(A)**, GSE72094 **(D)**, and GSE31210 **(G)** cohorts. **(B, E, H)** Comparison of the HR value of NETRS and 20 gene signatures in TCGA **(B)**, GSE72094 **(E)**, and GSE31210 **(H)** cohorts. **(C, F, I)** Comparison of the AUC value of NETRS and 20 gene signatures in TCGA **(C)**, GSE72094 **(F)**, and GSE31210 **(I)** cohorts. **(J–L)** Exploring the importance of the 21 gene signatures in predicting patients' survival in TCGA **(J)**, GSE72094 **(K)**, and GSE31210 **(L)** cohorts using RandomForest. **(M–O)** Exploring the importance of the 21 gene signatures in predicting patients' survival in TCGA **(M)**, GSE72094 **(N)**, and GSE31210 **(O)** cohorts using Boruta.

combining NETRS with clinical factors like TNM Stage, age, and gender, we developed a nomogram to predict the prognosis of LUAD patients. The nomogram had a C-Index of 0.729 in TCGA (Figure 6C), and its calibration curve (Figure 6D) and time-ROC curve (Figure 6E) confirmed its reliable predictive ability. Additionally, the DCA curves indicated that when integrated with clinical characteristics, NETRS exhibited a more robust predictive capability (Figures 6F–H). Finally, based on the time-AUC curve, we can conclude that the nomogram had the strongest predictive power, followed by NETRS and TNM Stage, while age and gender had the least predictive impact (Figure 6I).

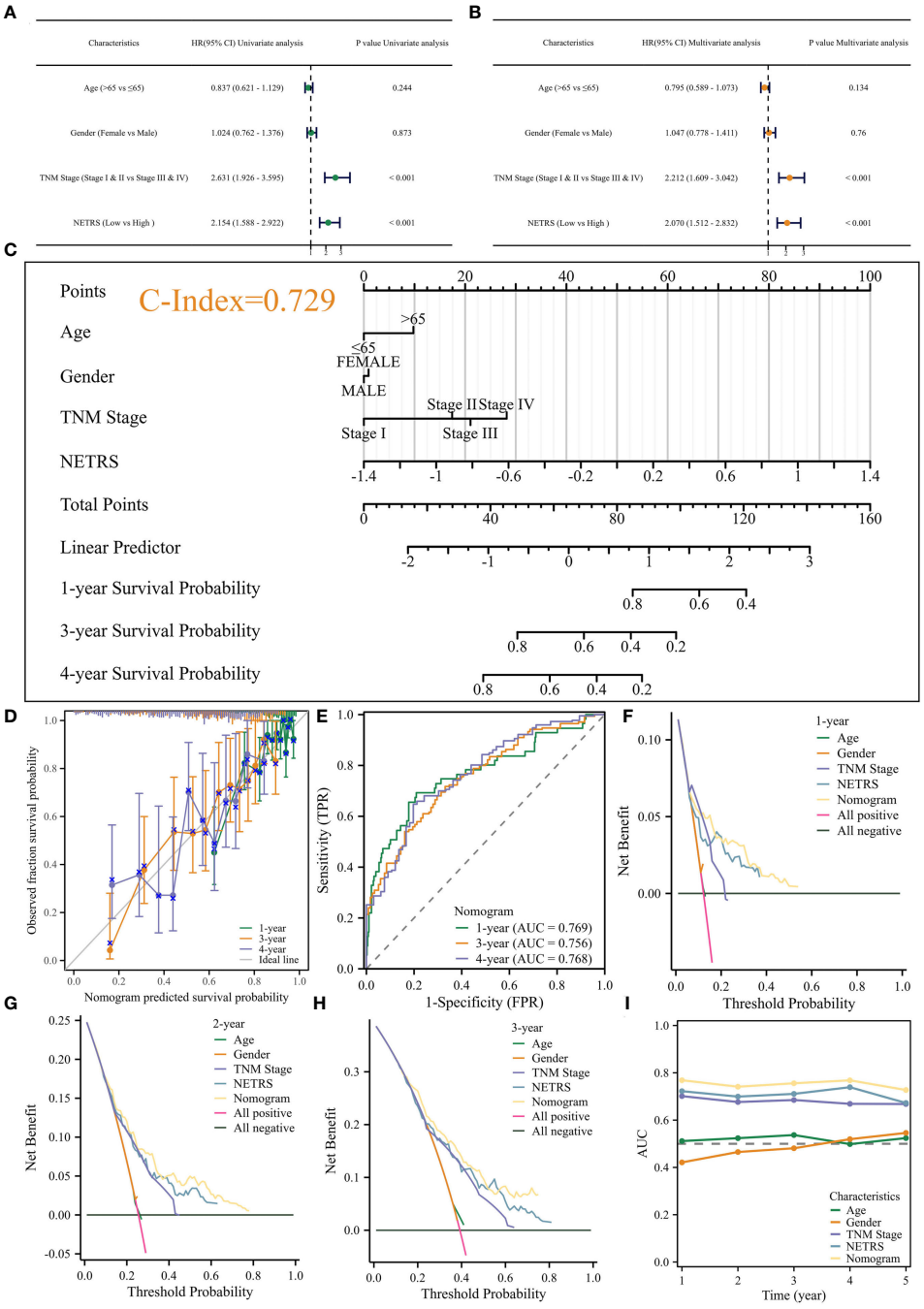


FIGURE 6
Construction of the Nomogram. **(A, B)** Uni- **(A)** and Multi- **(B)** variable Cox regression analysis identified NETRS as an independent prognostic factor. **(C–E)** The nomogram **(C)** and its calibration **(D)**, and ROC **(E)** curves. **(F–H)** The DCA curves showed that the nomogram has stronger power for predicting patients' 1- **(F)**, 2- **(G)**, and 3- **(H)** years' survival than single clinical factors and NETRS. **(I)** Time-AUC curves shows that after incorporating clinical factors, NETRS has stronger power for predicting patients' survival.

Exploring NETRS associated biological functions

In order to clarify the mechanism behind NETRS' excellent predictive abilities, further studies were carried out. Correlation analysis was performed to identify genes associated with NETRS, and we visualized the top 50 correlated genes (Table S2; Figure 7A). Genes positively correlated with NETRS were mainly involved in biological processes, such as 'Cell cycle', 'DNA replication', 'chromosome segregation', 'nuclear division', and 'p53 signaling pathway' (Figure 7B); while genes negatively correlated with NETRS were mainly concentrated in some immune-related functions such as 'macrophage activation', 'immune receptor activity', 'chemokine binding', 'MHC protein complex assembly', and 'mononuclear cell differentiation' (Figure 7C). Besides, cell cycle associated gene sets had higher activity in the high-NETRS group (Figure 7D), while immune-

TABLE 1 The NETRS in GSE72094 cohort was analyzed using both univariable and multivariable Cox regression analysis.

Characteristics	Univariate analysis		Multivariate analysis	
	Hazard ratio (95% CI)	P value	Hazard ratio (95% CI)	P value
Age (≤65 vs >65)	0.712 (0.464 - 1.093)	0.121	0.697 (0.451 - 1.075)	0.102
Gender (female vs male)	1.547 (1.065 - 2.246)	**0.022**	1.623 (1.110 - 2.374)	**0.012**
TNM Stage (I&II vs III&IV)	2.607 (1.736 - 3.914)	< **0.001**	2.655 (1.759 - 4.007)	< **0.001**
NETRS (low vs high)	3.218 (2.114 - 4.899)	< **0.001**	3.177 (2.082 - 4.849)	< **0.001**

The bold values means that the values were statistically significant, meaning that the p-value was less than 0.05.

function related gene sets scored higher in the low-NETRS group (Figure 7E), which echoing the results of GO and KEGG analysis.

Exploring NETRS at single-cell level

Figures 8A, B illustrate the distribution of cellular expression for 18 NETRS genes. It was observed that LDHA was expressed in seven different cell types, MS4A1 was exclusively expressed in B cells, S100P only in neutrophils, and FBP1 exclusively in mono/macrophages. SNX30, UBE2S, SEC14L4, FKBP4, and DOCK4 exhibited low expression levels across all cell types. Moving forward, we performed NETRS calculations at the single-cell level, which revealed that malignant cells exhibited the highest NETRS values (Figure 8C). By categorizing all cells into two groups based on the mean NETRS value (Figure 8D), we observed that the high NETRS group consisted of a larger proportion of tumor cells (Figure 8E), advanced T (Figure 8F) and N (Figure 8G) stages. Additionally, we validated enrichment analysis results in bulk datasets, demonstrating that cells with high NETRS scores exhibited elevated EMT (Figure 8H), cell cycle (Figure 8I), and DNA replication repair (Figure 8J) scores, indicative of more malignant characteristics.

NETRS significantly affects the TME

Due to the immune-hot characteristics observed in patients with low NETRS (Figure 7E), we conducted a comprehensive investigation into the relationship between NETRS and the tumor microenvironment (TME). After verifying six algorithms, we found that the low-NETRS group exhibited higher levels of immune cell infiltration (Figure 9A). This pattern was also observed in the single-cell dataset (Figure 9C). Furthermore, the low-NETRS group displayed significantly higher expression of various TME regulators, including immunoinhibitors, stimulators, chemokines, and MHC molecules (Figure 9D). Remarkably, we observed a significant positive correlation between most immune cells, while a significant negative correlation was evident between NETRS and the majority of immune cells (Figure 9B). Similar correlations were observed among TME regulators, with NETRS demonstrating significant negative associations with most of them (Figure 9E). Additionally, patients with low NETRS showed a marked reduction in the TIDE score, indicating a potentially greater benefit from immune checkpoint blockade (ICB) therapy for these individuals (Figure 9F). This hypothesis was validated in the ICB therapy cohort phs000452, where patients with low NETRS exhibited a higher proportion of ICB responders (Figure 9G) and showed improved prognosis (Figure 9H).

Multi-omics comparison between NETRS-high and NETRS-low groups in TCGA- LUAD

The top 20 genes with the highest mutation frequency were analyzed and visualized between groups with high and low NETRS scores (Figures 10A, B). Patients in the high-NETRS group had significantly higher gene mutation frequency (high-NETRS: TP53:60%, TTN: 54%, CSMD3: 49%, MUC16: 47%, RYR2: 40%;

TABLE 2 The NETRS in GSE31210 cohort was analyzed using both univariable and multivariable Cox regression analysis.

Characteristics	Univariate analysis		Multivariate analysis	
	Hazard ratio (95% CI)	P value	Hazard ratio (95% CI)	P value
Age (≤65 vs >65)	2.583 (1.313 - 5.083)	**0.006**	4.409 (2.145 - 9.066)	< **0.001**
Gender (female vs male)	1.519 (0.780 - 2.955)	0.219	1.379 (0.703 - 2.707)	0.350
TNM Stage (I vs II)	4.232 (2.175 - 8.236)	< **0.001**	3.226 (1.583 - 6.575)	**0.001**
NETRS (low vs high)	7.016 (2.720 - 18.096)	< **0.001**	5.983 (2.251 - 15.898)	< **0.001**

The bold values means that the values were statistically significant, meaning that the p-value was less than 0.05.



FIGURE 7
Enrichment analysis of potential biological functions associated with NETRS. **(A)** The heatmap shows the top 50 genes which are most correlated with NETRS. **(B, C)** GO and KEGG analysis showed the functions of genes positively **(B)** or negatively **(C)** correlated with NETRS. **(D, E)** GSVA analysis revealed the gene sets with high activity in the high NETRS group **(D)** and the gene sets with high activity in the low NETRS group **(E)**.

low-NETRS: TP53: 38%, TTN: 33%, CSMD3: 28%, MUC16: 32%, RYR2: 31%). Besides, NETRS posotively correlated with TMB (Figure 10C), and patients in high-NETRS group had higher TMB (Figure 10D). Survival analysis showed that TMB didn't affect patients' prognosis (Figure 10E). However, after combining TMB with NETRS, it can better stratify patients' survival (Figure 10F), and patients with Low NETRS + High TMB had relatively best prognosis. Patients in high-NETRS group had higher neoantigen load (Figure 10G), higher gene mutation rate (Figure 10H), high number of segments (Figure 10I), fraction altered (Figure 10J), aneuploidy score (Figure 10K), and homologous recombination defect (Figure 10L). Besides, we found that the CNV event also differed significantly between two NETRS groups (Figures 10M, N). Patients in the high-NETRS group had a higher frequency of CNV event, and most of them were amplification; while patients in the low-NETRS group had relatively lower frequency of CNV, and deletion had higher proportion. In addition, we found from the ChromPlots that patients in the high NETRS group had significantly higher G-Scores than patients in the low NETRS group (Figures 10O, P). Thus, high-NETRS patients with LUAD were more likely to display malignant characteristics.

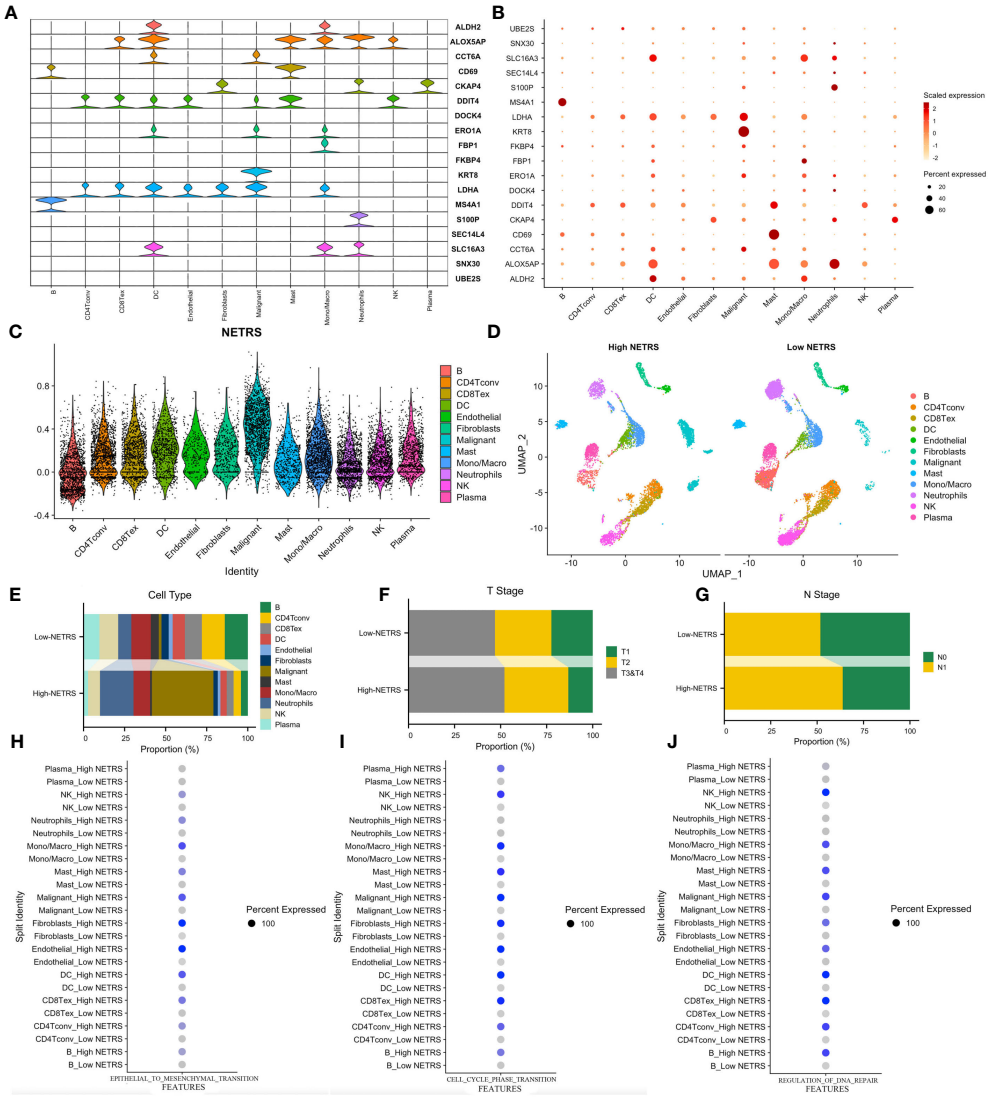


FIGURE 8
Exploring NETRS's distribution at cellular level. **(A, B)** The violin plot **(A)** and the dot plot **(B)** showed the 18 NETRS genes' expression at different cell types. **(C)** Different cells' NETRS level. **(D)** On the UMAP plot, the distribution of cells in high and low NETRS groups is shown. **(E–G)** Different cell **(E)**, T Stage **(F)**, and N Stage **(G)** proportion at different NETRS groups. **(H–J)** High-NETRS cells had higher EMT **(H)**, Cell cycle **(I)**, and DNA replication activity.

Identification of potential drugs targeting NETRS

Based on the drug information obtained from GDSC1, we investigated the relationship between NETRS and commonly used drugs for the clinical treatment of LUAD. Through correlation analysis, we found a negative correlation between NETRS and the IC50 values of 13 drugs (Figure 11A). These drugs, namely Docetaxel, Gefitinib, Vinorelbine, Cisplatin, Vinblastine, Paclitaxel, Gemcitabine, Etoposide, Methotrexate, Sorafenib, Mitomycin-C, Doxorubicin, and Afatinib, exhibited significantly lower IC50 values in the high-NETRS group compared to the low-NETRS group. This suggests that patients in the high-NETRS group may potentially benefit more from these drugs (Figures 11B–N).

NETRS is also valuable in predicting all NSCLC patients' survival

In light of these analyses, NETRS is a superior and robust prognosticator for patients with LUAD. However, NSCLC includes many subtypes, including LUAD, LUSC (lung squamous cell carcinoma), lung large cell carcinoma, lung basal cell carcinoma et al. Aiming to assess how well NETRS predicts the prognosis of patients with all forms of NSCLC, we collected nine independent datasets of NSCLC patients. In these nine datasets, except for the GSE74777 dataset only includes data for LUSC patients, all datasets contain data for patients with various NSCLC types. Figure 12A shows the expression of 18 NETRS genes in the TCGA-NSCLC cohort. In this cohort, NETRS increased with advanced T (Figure 12B), N

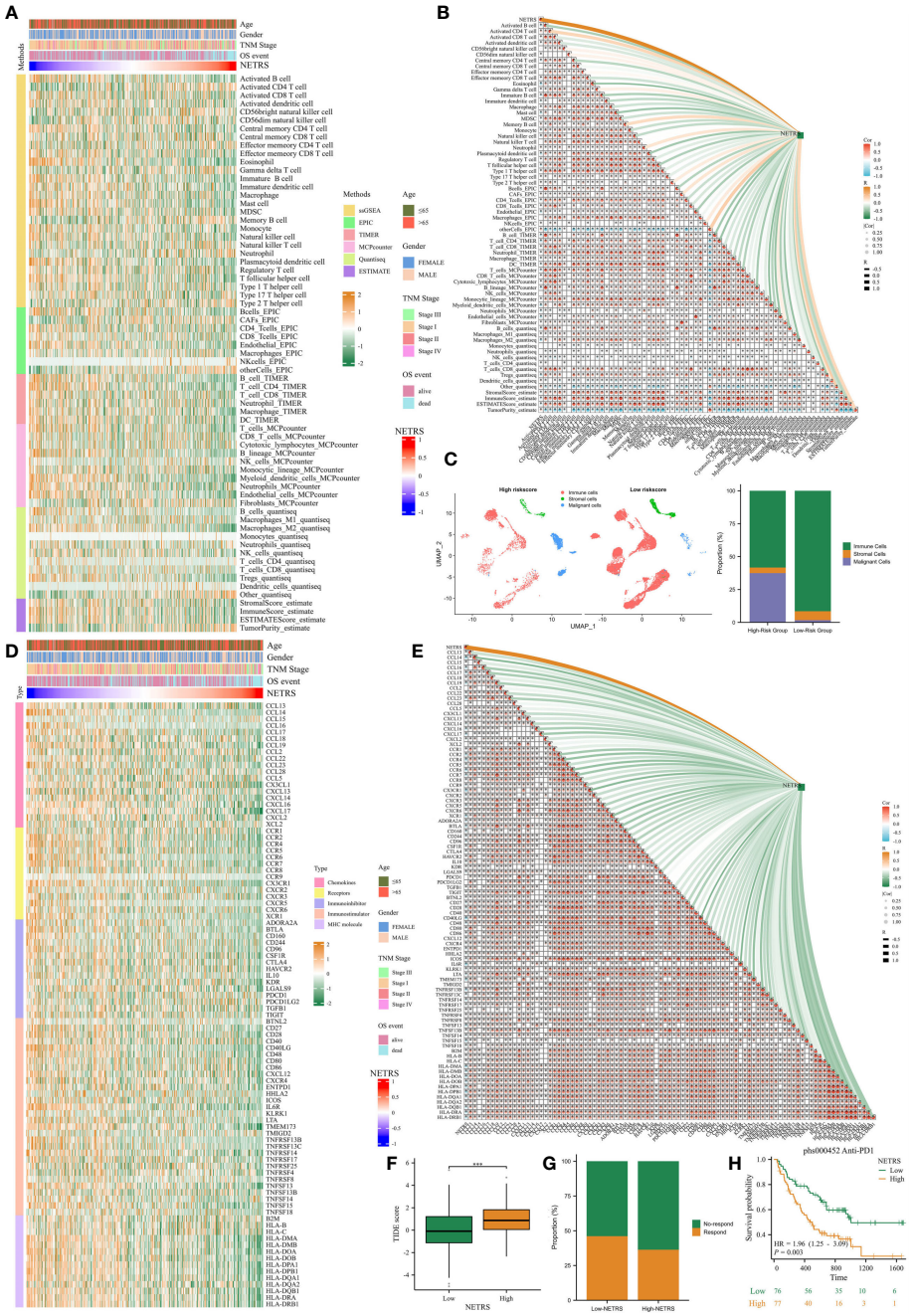


FIGURE 9
The effects of NETRS on the TME. **(A)** Immunocytes infiltrated significantly different between high- and low- NETRS groups. **(B)** NETRS negatively correlated with immune cells' infiltration. **(C)** In sc-RNA set, low-NETRS group had higher immune cell level. **(D)** TME modulators expressed significantly different between two NETRS groups. **(E)** NETRS negatively correlated with TME modulators' expression. **(F)** Patients with low NETRS had lower TIDE score. **(G)** The low-NETRS group had higher ICB responders' proportion. **(H)** Patients receiving immunotherapy with lower NETRS had better prognosis. '***' means P value is less than 0.001.

(Figure 12C), and TNM stages (Figure 12D), and dead survival event (Figure 12E), similar to the results in LUAD cohorts. Including the TCGA-NSCLC cohort, nine independent cohorts found that NSCLC patients with high NETRS had a significantly worse prognosis than those with low NETRS (all *P*-value < 0.01), and the ROC curves demonstrated that NETRS is highly predictive (Figures 12F–N). Therefore, NETRS is also valuable in predicting the prognosis of patients suffering from all types of NSCLC.

Validation of six NRGs' expression using three datasets, qRT-PCR, and Immunohistochemical staining

NETRS is comprised of 18 NRGs, of which six genes - ALOX5AP, DOCK4, CCT6A, MS4A1, SEC14L4, and SNX30 - have been minimally investigated in relation to LUAD. To further explore the potential involvement of the six NRGs in LUAD, we

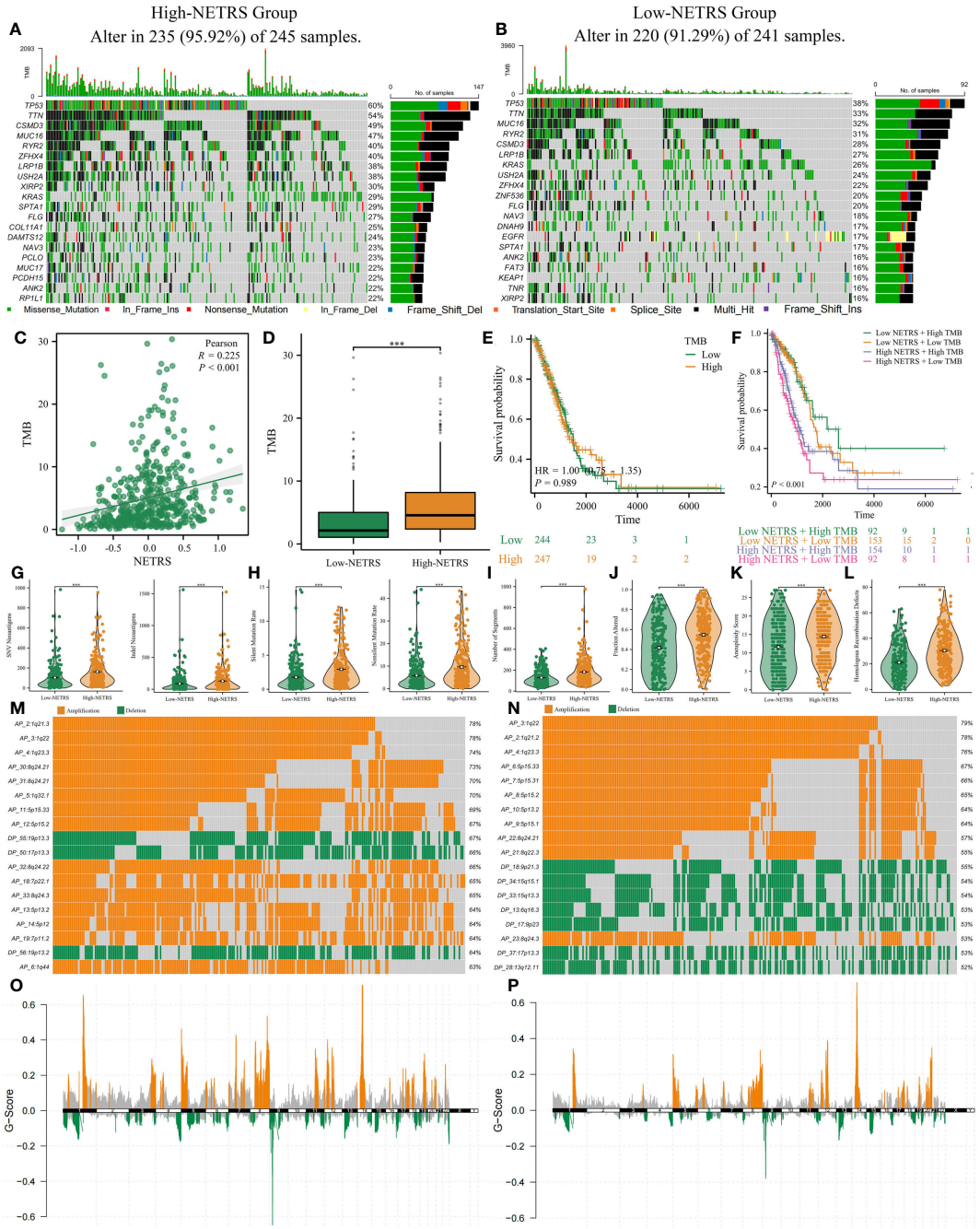


FIGURE 10
The multi-omics landscape differences between high and low NETRS groups. **(A, B)** High- **(A)** and low- **(B)** NETRS patients' somatic mutation frequency. **(C, D)** NETRS's correlation with TMB **(C)** and different NETRS groups patients' TMB difference **(D)**. **(E, F)** TMB **(E)** or TMB combined with NETRS's effect **(F)** on patients' survival. **(G–L)** High- and low- NETRS group patients' different neoantigen load **(G)**, gene mutation rate **(H)**, number of segments **(I)**, fraction altered **(J)**, aneuploidy score **(K)**, and homologous recombination defect **(L)**. **(M, N)** The top 18 CNV events in high- **(M)** and low- **(N)** NETRS groups. **(O, P)** ChromPlots show patients' G-score in high- **(O)** and low- **(P)** NETRS groups. '***' means P value is less than 0.001.

conducted a comparative analysis of their expression levels between LUAD and normal tissues. In three separate datasets, we observed high expression levels of DOCK4, ALOX5AP, SNX30, and SEC14L4 in normal lung tissues. Conversely, LUAD tissues showed high expression levels of CCT6A and MS4A1 (Figures 13A, C, E). Furthermore, the ROC curves demonstrated that all the six NRGs exhibited high diagnostic accuracy (Figures 13B, D, F).

Although we have unveiled the differential expression of the six NRGs, all the analyses were performed using publicly available databases. To further validate the expression of six NRGs and enhance the credibility of NETRS, we collected clinical samples

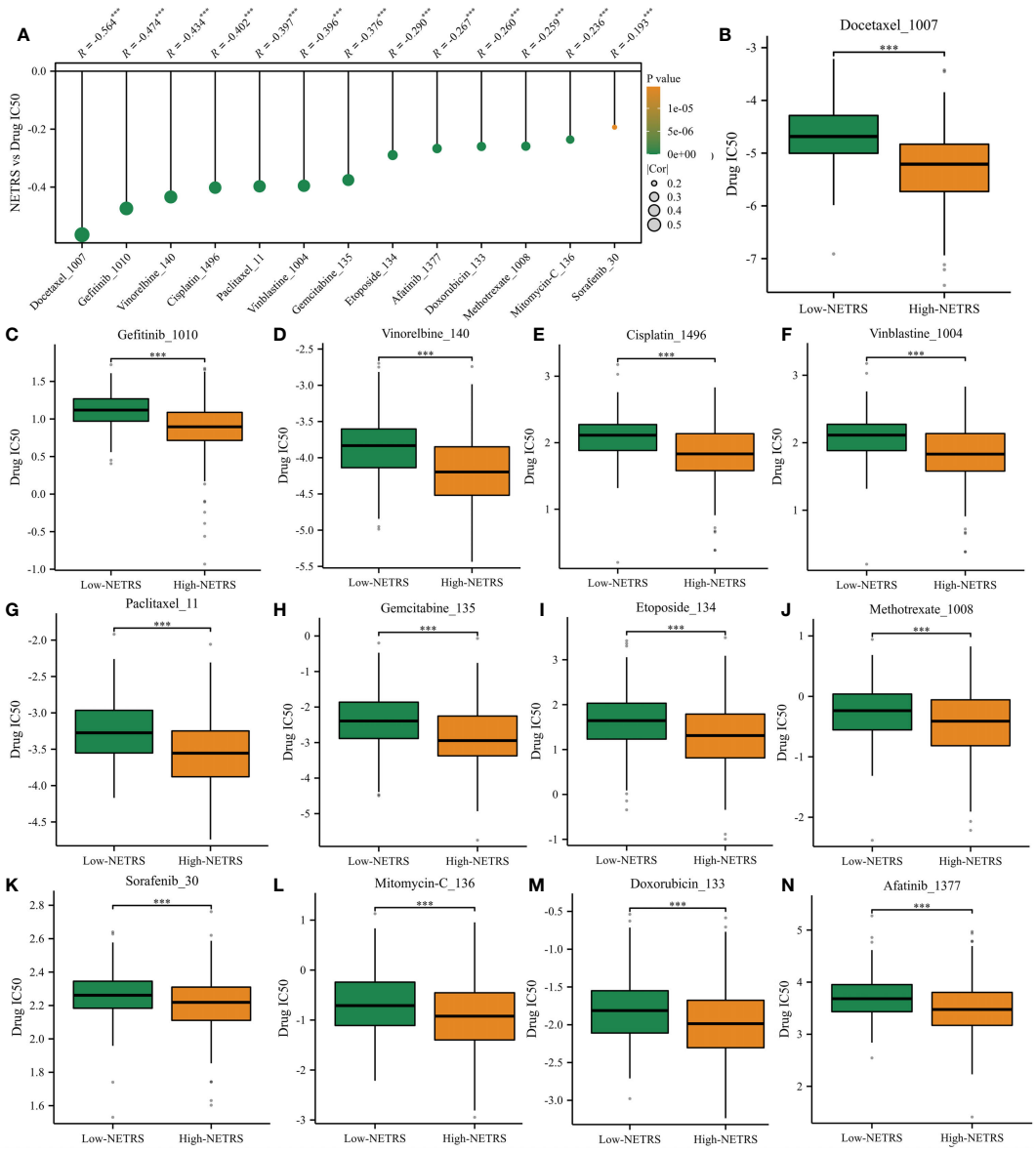


FIGURE 11
High-NETRS patients may more sensitivity to chemotherapy. **(A)** NETRS's relationship with drugs' IC50 value. **(B–N)** Comparison of the IC50 value of Docetaxel **(B)**, Gefitinib **(C)**, Vinorelbine **(D)**, Cisplatin **(E)**, Vinblastine **(F)**, Paclitaxel **(G)**, Gemcitabine **(H)**, Etoposide **(I)**, Methotrexate **(J)**, Sorafenib **(K)**, Mitomycin-C **(L)**, Doxorubicin, **(M)** and Afatinib **(N)** between high- and low- NETRS groups' patients.

and conducted qRT-PCR analysis. In line with the findings from publicly available databases, DOCK4, ALOX5AP, SNX30, and SEC14L4 exhibited high expression levels in normal lung tissues, while CCT6A and MS4A1 were highly expressed in LUAD tissues (Figure 13G). Finally, we selected ALOX5AP and CCT6A as representatives and validated their localization and protein expression in tissues by immunohistochemical staining. Consistent with the previous findings, immunohistochemical staining images demonstrated that ALOX5AP expression was decreased in LUAD (Figure 13H), whereas CCT6A expression was increased in LUAD (Figure 13I).

Discussion

There has been considerable progress in treating lung cancer, but it remains a very challenging tumor for medical professionals to treat adequately. The development of high-throughput sequencing technology has led to the discovery of more and more prognostic markers. The role of A newly detected type of programmed cell death is NETosis. NETosis in various tumor pathology is now well established, however it remains unclear what their precise molecular mechanism is (32, 33). This study comprehensively investigated NETosis-related genes in NSCLC due to the limited research on the role of NETosis in this context.

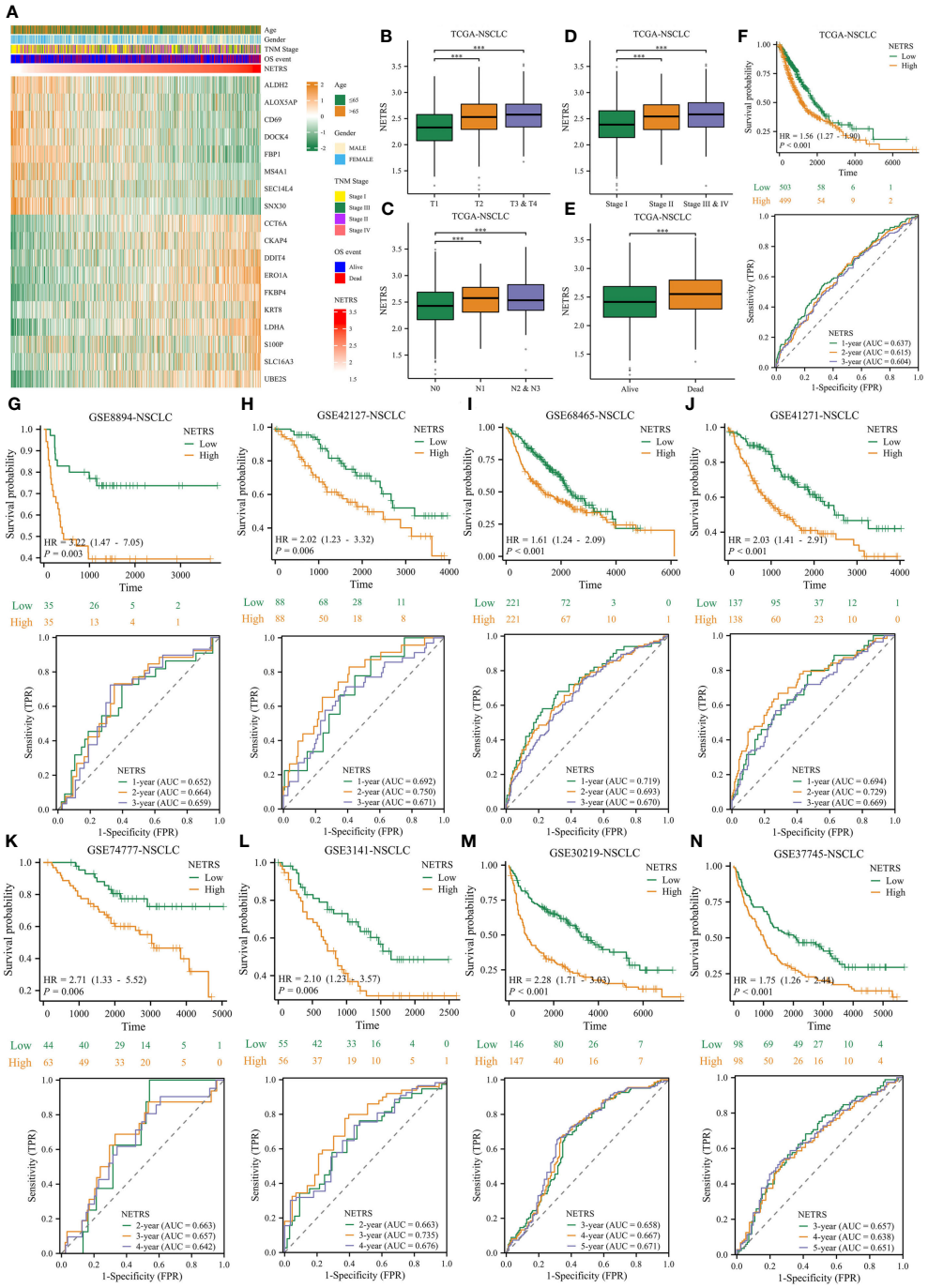


FIGURE 12
The value of NETRS in predicting the prognosis of NSCLC patients. **(A)** The 13 NETRS genes' expression pattern in TCGA-NSCLC cohort. **(B–E)** NSCLC patients' NETRS between different T **(B)**, N **(C)**, TNM **(D)** Stages, and survival events **(E)**. **(F–N)** The prognosis of NSCLC patients among high- and low- NETRS groups at TCGA-NSCLC **(F)**, GSE8894 **(G)**, GSE42127 **(H)**, GSE68465 **(I)**, GSE41271 **(J)**, GSE74777 **(K)**, 3141 **(L)**, GSE30219 **(M)**, and GSE307745 **(N)** sets and the corresponding time-dependent ROC curves. '***' means P value is less than 0.001.

In the first step, the AUCell algorithm was used to calculate the NETosis scores for various cell types. Our results indicated that neutrophils had the highest NETosis score, it may due to the fact that NETosis is a neutrophil-mediated programmed cell death. In the next step, we divided each population into two groups based on their mean NETosis score. According to cell-cell communication analysis, cells with high NETosis Score communicate more frequently and strongly than those with low NETosis Score. This suggests that cells with a high NETosis score may be more active in the TME, leading to anti-disease or pathogenic mechanisms. We then determined DEGs between high NETRS and low NETRS cells, and enrichment analysis showed that these genes were significantly correlated with NETosis. By intersecting DEGs between high and low NETRS cell groups in single-cell dataset, with DEGs between

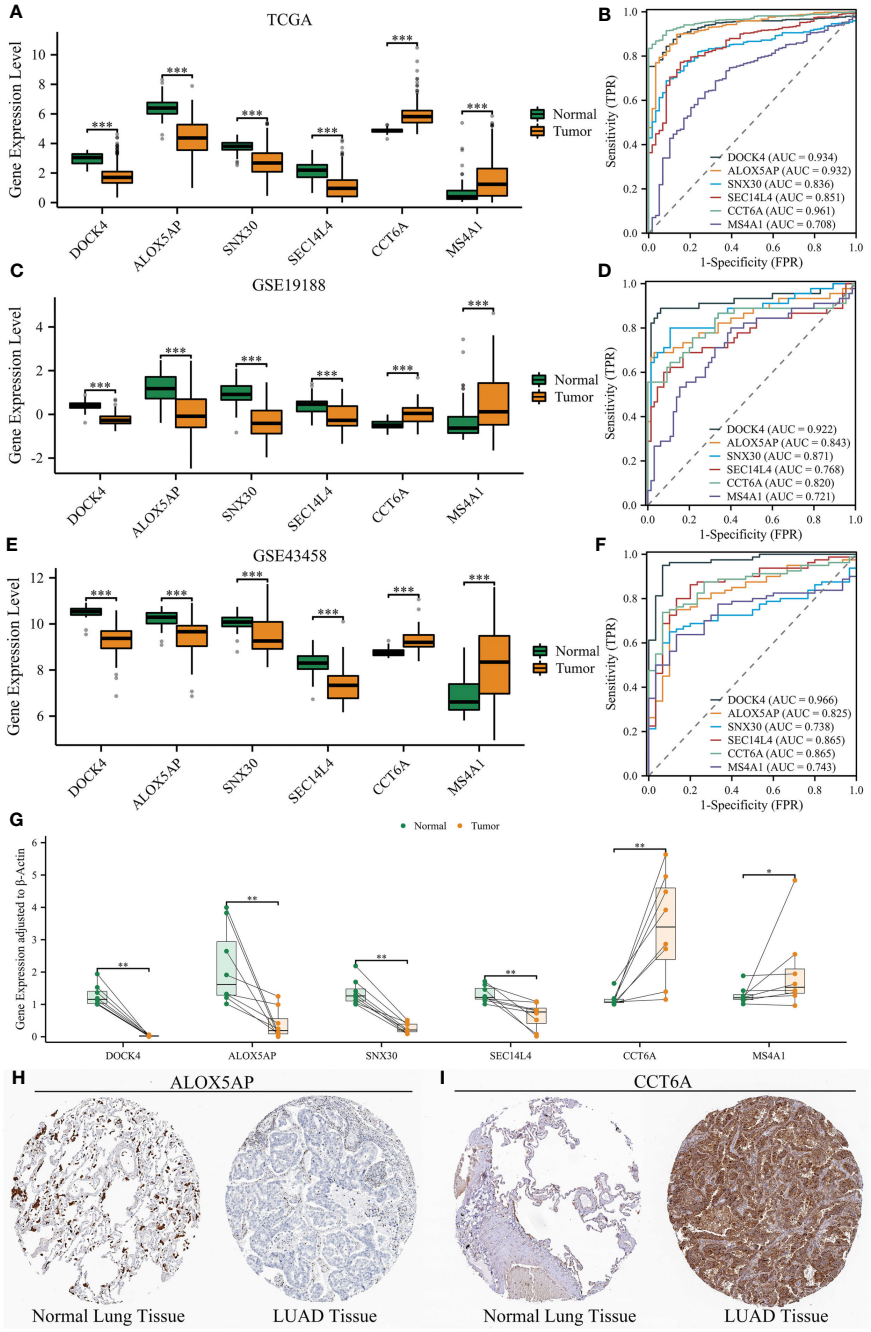


FIGURE 13
Validation of six NRGs' expression using three datasets and clinical samples. **(A, B)** Six NRGs' differential expression between normal lung and LUAD tissues **(A)**, and their diagnostic value **(B)** in the TCGA dataset. **(C, D)** Six NRGs' differential expression between normal lung and LUAD tissues **(C)**, and their diagnostic value **(D)** in the GSE19188 dataset. **(E, F)** Six NRGs' differential expression between normal lung and LUAD tissues **(E)**, and their diagnostic value **(F)** in the GSE43458 dataset. **(G)** The differential expression of six NRGs between normal lung and LUAD tissues by qRT-PCR analysis. **(H–I)** Immunohistochemical staining confirmed the decreased protein level of ALOX5AP **(H)** and increased protein level of CCT6A **(I)** in LUAD tissues. '*' means P value is less than 0.05, '**' means P value is less than 0.01, and '***' means P value is less than 0.001.

normal tissues and LUAD in bulk datasets, and also with prognostic related genes, 61 NETosis-related Genes (NRGs) were identified. Then based on these 61 NRGs, we identified two NETosis-related clusters. Patients in cluster1 had better prognosis and more active TME characteristics, and most of the 61 NRGs were highly expressed in cluster1, which suggested that cluster1 may more like to be a NETosis-like cluster. Cluster1 also demonstrated a significant improvement in prognosis, suggesting that higher NET activity in LUAD patients' TME may benefit tumor cell clearance. After that, the NETRS comprised of 18 NRGS was obtained using the Enet algorithm (α=0.1). NETRS proved superior predictive efficacy and was an independent prognostic factor in both training

TCGA-LUAD and test sets GSE72094-LUAD and GSE31210-LUAD. Additionally, significant differences were observed between patients categorized into high and low NETRS groups, as determined by PCA analysis. Further exploration was conducted to investigate the potential prognostic mechanism of NETRS, along with the disparities in immune cell infiltration, expression of immune-related molecules, SNV and CNV frequencies, TMB, Neoantigen load, and response to immunotherapy, and drug sensitivity among patients with high and low NETRS. The construction of nomograms revealed that the prognostic ability of NETRS was improved when integrated with clinical parameters, offering a novel approach to predict the prognosis of patients with LUAD. NETRS can also predict all NSCLC patients' prognosis, which is its another valuable feature: Among more than 2,000 NSCLC patients in nine independent cohorts, patients with high NETRS all had significantly worse outcomes than those with low NETRS, which demonstrated the multifaceted value of NETRS. Lastly, the differential expression of the six identified NRGs was verified using three independent datasets, along with clinical samples gathered from our own collection. These six NRGs have been relatively understudied in LUAD. This study is the first to confirm their expression patterns. Specifically, DOCK4, ALOX5AP, SNX30, and SEC14L4 showed elevated expression levels in normal lung tissues, while CCT6A and MS4A1 exhibited high expression in LUAD tissues. Therefore, this finding not only provides a basis for future investigations on these NRGs in LUAD, but also reinforces the efficacy and reliability of NETRS.

NETRS consisted of 18 NRGs: ALDH2, ALOX5AP, CCT6A, CD69, CKAP4, DDIT4, DOCK4, ERO1A, FBP1, FKBP4, KRT8, LDHA, MS4A1, S100P, SEC14L4, SLC16A3, SNX30, and UBE2S. The lasso regression coefficient of UBE2S, CKAP4, S100P, SLC16A3, CCT6A, DDIT4, KRT8, FKBP4, ERO1A, and LDHA is greater than 0, which were risk factors; while the regression coefficients of MS4A1, FBP1, SNX30, DOCK4, SEC14L4, ALOX5AP, CD69, and ALDH2 were less than 0, which were protective factors. Some of these genes have been extensively studied in relation to LUAD. Overexpression of ALDH2 (protective factor in our study) decreased migration, and proliferation in LUAD cells, but knockdown of ALDH2 increased these properties (34); the growth rate of lung cancer cells overexpressed with CKAP4 (risk factor in our study) was increased *in vivo*, while an antibody against the protein inhibited it (35); LUAD cells were impaired in their ability to invade, metastasize, and proliferate after FBP1 (protective factor in our study) was overexpressed (36); FKBP4 and S100P (risk factors in our study) promotes proliferation and migration of NSCLC cells and inhibits apoptosis, while promoting tumor growth *in vivo* (37, 38). Additionally, GO and GSVA enrichment analyses showed that genes positively associated with NETRS were significantly enriched in functions related to cell cycle. Studies shown that when EOR1A (risk factor in our study, also known as ERO1L) was depleted from NSCLC cells, the expression of factors associated with cell cycle is dramatically reduced (39). Thus, cell cycle-related functions enriched in patients with a high NETRS may as a result of these risky factors such as ERO1A. Besides, enrichment analysis also showed that genes negatively correlated with NETRS were distinctly enriched in functions related to the immune system. We found that CD69 (protective factor in our study) was highly expressed in low-NETRS groups. CD69 has been shown to play an important role in regulating TME, and it may play a role in affecting PD-1 treatment response (40). Thus, immune-related functions enriched in patients with a low NETRS may as a result of these protective factors such as CD69.

Cancer immunotherapy offers new hope to patients with cancerous growths, but evading the immune system remains a formidable challenge to treatment (41, 42). According to our research, individuals belonging to the low NETRS category exhibited elevated levels of immune cell infiltration, increased expression of immune-related substances, including well-known immune checkpoints like PDCD1, BTLA, CTLA4, along with chemokines/receptors and MHC molecules. Consequently, individuals with low levels of NETRS experience 'immune hot' symptoms accompanied by greater immune cell infiltration. Moreover, it was discovered that individuals possessing a low NETRS exhibited considerably reduced TIDE scores, which means that the TIDE algorithm anticipated that patients with a low NETRS would display heightened responsiveness to immunotherapy. This hypothesis was confirmed in the phs000452 immunotherapy cohort. Furthermore, we compared the sensitivity of patients in different NETRS groups to a number of drugs commonly used to treat NSCLC in clinics. These was a number of chemotherapy agents can effectively treat patients with NSCLC, such as paclitaxel, docetaxel, cisplatin, vincristine, and vinorelbine. Interestingly, these chemotherapy agents' IC50 value in patients with high-NETRS patients, which means patients with higher NETRS may more sensitive to these drugs. In addition, the IC50 value of gefitinib and afatinib, which target NSCLC patients with EGFR mutation, remain low in patients with high NETRS (43, 44). Thus, patients with high NETRS would likely benefit more from targeted and chemotherapy therapies, while those with low NETRS would benefit more from immunotherapy.

In spite of the fact that NETosis is a PCD mode which strongly associated with neutrophils, this is the first study to investigate how NETosis-related genesets are expressed by the cells of LUAD tissues, and confirmed that neutrophils has the highest NETosis score in LUAD. Through this, we gained a deeper understanding of NETosis, and provided some novel insights into the effects of NETosis on TME for future studies. This study is also the first to systematically investigate the prognostic role of NETosis-related genes in NSCLC, and the NETRS consisting of 18 NRGs demonstrated its powerful predictive power in 12 cohorts of nearly 3,300 NSCLC patients. In comparison to 20 previously published gene signatures, NETRS had higher C-Index, HR value, and greater accuracy in predicting 1- and 2-year survival of LUAD patients. According to two machine learning algorithms, NETRS had the greatest impact on the survival of LUAD patients compared to these 20 signatures. Thus, we have developed a gene signature that more accurately predicts LUAD prognoses. In addition, we found that when our predecessors applied single-cell sequencing to construct prognostic models, they didn't combined functional phenotypes, such as the study of Song et al. (PMID:35757748, PMID:35152302) and the study of Zhang et al. (PMID:37507593).

In these studies, prognostic models were only constructed based on the markers genes of B cells, NK cells, and DCs, while the relationship between functional phenotypes and cells wasn't explored. Besides, some studies that targeting functional phenotypes didn't utilize single-cell sequencing sets, such as the study of Zhao et al. (PMID:36275729) and the study of Li et al. (PMID:35529878). Therefore, we believe that combining single-cell sequencing with functional phenotypes to construct prognostic models can provide new insights into future prognostic studies.

Although the study demonstrated NETRS's remarkable capability in forecasting prognosis and gauging therapy response in LUAD patients, it still has certain constraints. Initially, all the information in this research originated from publicly accessible databases, and the effectiveness of NETRS was not validated using our own group of subjects. A second reason is that our study only validated the expression patterns of six NRGs and did not extensively investigate their role in LUAD. In our future study, we will confirm the effectiveness of NETRS in larger groups, and additionally investigate the NRGs' risky or protective role in LUAD further.

To sum up, this research has devised a novel NETRS (NETosis-related Riskscore) which can accurately forecast the prognosis of patients with NSCLC and their reaction treatment. It offers fresh perspectives for future investigations on genes associated with NETosis and the combining use of bulk- and single-cell RNA-sequencing data.

Data availability statement

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

Ethics statement

The studies involving humans were approved by The Ethics Committee of the First Affiliate Hospital of Fujian Medical University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

LZ: Data curation, Formal Analysis, Investigation, Methodology, Software, Writing – original draft. XZ: Data curation, Investigation, Methodology, Software, Writing – original draft. MG: Data curation, Investigation, Methodology, Software, Writing – original draft. FY: Conceptualization, Supervision, Writing – review & editing. FL: Conceptualization, Supervision, Writing – review & editing.

Funding

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

Acknowledgments

We thanks all the online databases.

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.2023.1282335/full#supplementary-material

References

1. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Piñeros M, et al. Estimating the global cancer incidence and mortality in 2018: globocan sources and methods. *Int J Cancer* (2019) 144(8):1941–53. doi: 10.1002/ijc.31937

2. Lu T, Yang X, Huang Y, Zhao M, Li M, Ma K, et al. Trends in the incidence, treatment, and survival of patients with lung cancer in the last four decades. *Cancer Manag Res* (2019) 11:943–53. doi: 10.2147/cmar.S187317

3. Gandhi L, Rodríguez-Abreu D, Gadgeel S, Esteban E, Felip E, De Angelis F, et al. Pembrolizumab plus chemotherapy in metastatic non-small-cell lung cancer. *N Engl J Med* (2018) 378(22):2078–92. doi: 10.1056/NEJMoa1801005

4. Novello S, Barlesi F, Califano R, Cufer T, Ekman S, Levra MG, et al. Metastatic non-small-cell lung cancer: esmo clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol* (2016) 27(suppl 5):v1–v27. doi: 10.1093/annonc/mdw326

5. Youlden DR, Cramb SM, Baade PD. The international epidemiology of lung cancer: geographical distribution and secular trends. *J Thorac Oncol* (2008) 3(8):819–31. doi: 10.1097/JTO.0b013e31818020eb

6. Jorch SK, Kubes P. An emerging role for neutrophil extracellular traps in noninfectious disease. *Nat Med* (2017) 23(3):279–87. doi: 10.1038/nm.4294

7. Ou Q, Fang JQ, Zhang ZS, Chi Z, Fang J, Xu DY, et al. Tcpc inhibits neutrophil extracellular trap formation by enhancing ubiquitination mediated degradation of peptidylarginine deiminase 4. *Nat Commun* (2021) 12(1):3481. doi: 10.1038/s41467-021-23881-8

8. Efrimescu CI, Buggy PM, Buggy DJ. Neutrophil extracellular trapping role in cancer, metastases, and cancer-related thrombosis: A narrative review of the current evidence base. *Curr Oncol Rep* (2021) 23(10):118. doi: 10.1007/s11912-021-01103-0

9. Szczerba BM, Castro-Giner F, Vetter M, Krol I, Gkountela S, Landin J, et al. Neutrophils escort circulating tumour cells to enable cell cycle progression. *Nature* (2019) 566(7745):553–7. doi: 10.1038/s41586-019-0915-y

10. Yang L, Liu Q, Zhang X, Liu X, Zhou B, Chen J, et al. DNA of neutrophil extracellular traps promotes cancer metastasis via ccdc25. *Nature* (2020) 583 (7814):133–8. doi: 10.1038/s41586-020-2394-6

11. Yang LY, Luo Q, Lu L, Zhu WW, Sun HT, Wei R, et al. Increased neutrophil extracellular traps promote metastasis potential of hepatocellular carcinoma via provoking tumorous inflammatory response. *J Hematol Oncol* (2020) 13(1):3. doi: 10.1186/s13045-019-0836-0

12. Zhang Y, Chandra V, Riquelme Sanchez E, Dutta P, Quesada PR, Rakoski A, et al. Interleukin-17-induced neutrophil extracellular traps mediate resistance to checkpoint blockade in pancreatic cancer. *J Exp Med* (2020) 217(12):e20190354. doi: 10.1084/jem.20190354

13. Jiang ZZ, Peng ZP, Liu XC, Guo HF, Zhou MM, Jiang D, et al. Neutrophil extracellular traps induce tumor metastasis through dual effects on cancer and endothelial cells. *Oncoimmunology* (2022) 11(1):2052418. doi: 10.1080/2162402x.2022.2052418

14. Wang Z, Jensen MA, Zenklusen JC. A practical guide to the cancer genome atlas (Tcga). *Methods Mol Biol* (2016) 1418:111–41. doi: 10.1007/978-1-4939-3578-9_6

15. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. Gistic2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* (2011) 12(4):R41. doi: 10.1186/gb-2011-12-4-r41

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

17. Sun D, Wang J, Han Y, Dong X, Ge J, Zheng R, et al. Tisch: A comprehensive web resource enabling interactive single-cell transcriptome visualization of tumor microenvironment. *Nucleic Acids Res* (2021) 49(D1):D1420–d30. doi: 10.1093/nar/gkaa1020

18. Zhang L, Guan M, Zhang X, Yu F, Lai F. Machine-learning and combined analysis of single-cell and bulk-rna sequencing identified a dc gene signature to predict prognosis and immunotherapy response for patients with lung adenocarcinoma. *J Cancer Res Clin Oncol* (2023). doi: 10.1007/s00432-023-05151-w

19. Şenbabaoğlu Y, Gejman RS, Winer AG, Liu M, Van Allen EM, de Velasco G, et al. Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger rna signatures. *Genome Biol* (2016) 17(1):231. doi: 10.1186/s13059-016-1092-z

20. Papayannopoulos V. Neutrophil extracellular traps in immunity and disease. *Nat Rev Immunol* (2018) 18(2):134–47. doi: 10.1038/nri.2017.105

21. Zhang Y, Guo L, Dai Q, Shang B, Xiao T, Di X, et al. A signature for pan-cancer prognosis based on neutrophil extracellular traps. *J Immunother Cancer* (2022) 10(6): e004210. doi: 10.1136/jitc-2021-004210

22. Van de Sande B, Flerin C, Davie K, De Waegeneer M, Hulselmans G, Aibar S, et al. A scalable scenic workflow for single-cell gene regulatory network analysis. *Nat Protoc* (2020) 15(7):2247–76. doi: 10.1038/s41596-020-0336-2

23. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med* (2018) 24(10):1550–8. doi: 10.1038/s41591-018-0136-1

24. Chen Z, Luo Z, Zhang D, Li H, Liu X, Zhu K, et al. Tiger: A web portal of tumor immunotherapy gene expression resource. *Genomics Proteomics Bioinf* (2022) 29: S1672-0229(22)00099-7. doi: 10.1016/j.gpb.2022.08.004

25. Liu X, Wu J. History, applications, and challenges of immune repertoire research. *Cell Biol Toxicol* (2018) 34(6):441–57. doi: 10.1007/s10565-018-9426-0

26. Fang Z, Tian Y, Sui C, Guo Y, Hu X, Lai Y, et al. Single-cell transcriptomics of proliferative phase endometrium: systems analysis of cell-cell communication network using cellchat. *Front Cell Dev Biol* (2022) 10:919731. doi: 10.3389/fcell.2022.919731

27. Yu G, Wang LG, Han Y, He QY. Clusterprofiler: an R package for comparing biological themes among gene clusters. *Omics* (2012) 16(5):284–7. doi: 10.1089/omi.2011.0118

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

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

30. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of drug sensitivity in cancer (Gdsc): A resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res* (2013) 41(Database issue):D955–61. doi: 10.1093/nar/gks1111

31. Maeser D, Gruener RF, Huang RS. Oncopredict: an R package for predicting *in vivo* or cancer patient drug response and biomarkers from cell line screening data. *Brief Bioinform* (2021) 22(6):bbab260. doi: 10.1093/bib/bbab260

32. Zhu T, Zou X, Yang C, Li L, Wang B, Li R, et al. Neutrophil extracellular traps promote gastric cancer metastasis by inducing epithelial–Mesenchymal transition. *Int J Mol Med* (2021) 48(1):127. doi: 10.3892/ijmm.2021.4960

33. van der Windt DJ, Sud V, Zhang H, Varley PR, Goswami J, Yazdani HO, et al. Neutrophil extracellular traps promote inflammation and development of hepatocellular carcinoma in nonalcoholic steatohepatitis. *Hepatology* (2018) 68 (4):1347–60. doi: 10.1002/hep.29914

34. Li K, Guo W, Li Z, Wang Y, Sun B, Xu D, et al. Aldh2 repression promotes lung tumor progression via accumulated acetaldehyde and DNA damage. *Neoplasia* (2019) 21(6):602–14. doi: 10.1016/j.neo.2019.03.008

35. Nagoya A, Sada R, Kimura H, Yamamoto H, Morishita K, Miyoshi E, et al. Ckap4 is a potential exosomal biomarker and therapeutic target for lung cancer. *Transl Lung Cancer Res* (2023) 12(3):408–26. doi: 10.21037/tlcr-22-571

36. Wang Z, He T, Lv W, Hu J. Down-regulation of fbp1 in lung adenocarcinoma cells promotes proliferation and invasion through slug mediated epithelial mesenchymal transformation. *Transl Cancer Res* (2023) 12(2):236–46. doi: 10.21037/tcr-22-2200

37. Meng W, Meng J, Jiang H, Feng X, Wei D, Ding Q. Fkbp4 accelerates Malignant progression of non-small-cell lung cancer by activating the akt/mtor signaling pathway. *Anal Cell Pathol (Amst)* (2020) 2020:6021602. doi: 10.1155/2020/6021602

38. Hsu YL, Hung JY, Liang YY, Lin YS, Tsai MJ, Chou SH, et al. S100p interacts with integrin α7 and increases cancer cell migration and invasion in lung cancer. *Oncotarget* (2015) 6(30):29585–98. doi: 10.18632/oncotarget.4987

39. Shi X, Wu J, Liu Y, Jiang Y, Zhi C, Li J. Ero1l promotes nsclc development by modulating cell cycle-related molecules. *Cell Biol Int* (2020) 44(12):2473–84. doi: 10.1002/cbin.11454

40. Hu ZW, Sun W, Wen YH, Ma RQ, Chen L, Chen WQ, et al. Cd69 and sbk1 as potential predictors of responses to pd-1/pd-L1 blockade cancer immunotherapy in lung cancer and melanoma. *Front Immunol* (2022) 13:952059. doi: 10.3389/fimmu.2022.952059

41. Hegde PS, Chen DS. Top 10 challenges in cancer immunotherapy. *Immunity* (2020) 52(1):17–35. doi: 10.1016/j.immuni.2019.12.011

42. Yu S, Wang R, Tang H, Wang L, Zhang Z, Yang S, et al. Evolution of lung cancer in the context of immunotherapy. *Clin Med Insights Oncol* (2020) 14:1179554920979697. doi: 10.1177/1179554920979697

43. Hosomi Y, Morita S, Sugawara S, Kato T, Fukuhara T, Gemma A, et al. Gefitinib alone versus gefitinib plus chemotherapy for non-small-cell lung cancer with mutated epidermal growth factor receptor: nej009 study. *J Clin Oncol* (2020) 38(2):115–23. doi: 10.1200/jco.19.01488

44. Harvey RD, Adams VR, Beardslee T, Medina P. Afatinib for the treatment of egfr mutation-positive nsclc: a review of clinical findings. *J Oncol Pharm Pract* (2020) 26 (6):1461–74. doi: 10.1177/1078155220931926

frontiers | Frontiers in Oncology

TYPE Case Report
PUBLISHED 21 November 2023
DOI 10.3389/fonc.2023.1265749

Check for updates

OPEN ACCESS

EDITED BY
Hongda Liu,
Nanjing Medical University, China

REVIEWED BY
Naotaka Ogasawara,
Aichi Medical University School of Medicine, Japan
Minghan Yang,
New York University,
United States

*CORRESPONDENCE
Conghua Xie
chxie_65@whu.edu.cn
Yan Gong
yan.gong@whu.edu.cn

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

RECEIVED 23 July 2023
ACCEPTED 24 October 2023
PUBLISHED 21 November 2023

CITATION
Wang M, Chen G, Luo J, Fan Z, Liu Y, Xie C and Gong Y (2023) Case Report: Genetic profiling of small intestine metastasis from poorly differentiated non-small cell lung cancer: report of 2 cases and literature review of the past 5 years.
Front. Oncol. 13:1265749.
doi: 10.3389/fonc.2023.1265749

COPYRIGHT
© 2023 Wang, Chen, Luo, Fan, Liu, Xie and Gong. 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.

Case Report: Genetic profiling of small intestine metastasis from poorly differentiated non-small cell lung cancer: report of 2 cases and literature review of the past 5 years

Mengqin Wang[1†], Gang Chen[1†], Jiang Luo[1], Zhipeng Fan[1], Yuying Liu[1], Conghua Xie[1,2*] and Yan Gong[1,3*]

[1]Department of Radiation and Medical Oncology, Zhongnan Hospital of Wuhan University, Wuhan, China, [2]Hubei Key Laboratory of Tumor Biological Behaviors, Zhongnan Hospital of Wuhan University, Wuhan, China, [3]Tumor Precision Diagnosis and Treatment Technology and Translational Medicine, Hubei Engineering Research Center, Zhongnan Hospital of Wuhan University, Wuhan, China

Background: Poorly differentiated non-small cell lung cancer (NSCLC) is characteristic of high rate of distant metastasis and late stages at diagnosis. Small intestine metastasis is a rare but severe complication of lung cancer with a high rate of mortality. However, there is currently a lack of genetic profile studies on the small intestine metastasis of lung cancer.

Case presentations: We present 2 cases of male patients in their 60s with primary NSCLC of low differentiation, initially with no distant metastasis detected. Biopsy samples were obtained from the primary pulmonary lesions, and both patients received systematic radiotherapy (RT) and chemotherapy. However, both cases presented with abdominal pain and distension, and immunohistochemistry of small intestine biopsy samples obtained by endoscopy confirmed lung cancer metastasis. Next generation sequencing was used to explore the genetic profiles from the biopsy samples of both the primary pulmonary lesions and small intestine metastases. The correlated genes responsible for the small intestine metastasis from poorly differentiated NSCLC in these 2 patients included TP53, LRP1B, and FGFR2. The reports of small intestine metastasis from poorly differentiated NSCLC with the past 5 years were systematically reviewed and summarized subsequently.

Conclusions: Poorly differentiated NSCLC with small intestine metastases, while rare, substantially impacts the prognosis and poses major challenges for diagnosis and treatment. Through comparisons of genetic profiles between patients and in the same patient before and after metastasis, we identified the mutations in genes such as TP53, LRP1B, and FGFR2, which were correlated with

the occurrence and progression of poorly differentiated NSCLC, as well as its small intestinal metastasis. This discovery has the potential to guide clinicians in developing personalized treatment plans through the manipulation of targeted and radiation therapies.

KEYWORDS

poorly differentiated NSCLC, small intestine metastasis, genetic profiling, radiotherapy, gene mutation

1 Introduction

Global cancer statistics revealed that lung cancer is the leading cause of cancer-related mortality worldwide (1). Non-small cell lung cancer (NSCLC) accounts for the majority (approximately 85%) of lung cancer cases. Compared with well differentiated ones, poorly differentiated NSCLC had an increased mortality by 83% and a 2.1-fold increase of recurrence due to its rapid growth rate and high metastasis propensity (2). More than 60% poorly differentiated NSCLC patients suffer from distant metastasis (3), including lymph nodes, liver, adrenal glands, bone, and brain. The occurrence of gastrointestinal tract metastasis is relatively uncommon, with only 1.77% reported, but it severely impacts the survival of lung cancer patients (4). A retrospective analysis of 366 lung cancer patients with gastrointestinal metastasis revealed that over half of them passed away within 3 months from diagnosis, with a median overall survival of 2.8 months (5). Among them, small intestine metastasis was further rare, reported in 11.9% of gastrointestinal cases, but substantially shortened the survival of lung cancer patients (6).

With technological development, next generation sequencing (NGS) is increasingly applied to detect genetic alterations and guide personalized treatments for patients in clinic. It has been demonstrated that mutations in driver genes such as EGFR, ALK, and KRAS are associated with metastases in various sites in NSCLC, including the brain, bone, liver, and lungs (7), but not the intestine. Reports and studies on lung cancer with small intestine metastasis remain relatively limited. Early detection or prediction is critical and necessary to improve efficacy and patient survival.

In this report, we presented 2 cases of poorly differentiated NSCLC with small intestine metastasis. By juxtaposing their NGS results, we unearthed the most correlated genes, including but not limited to, TP53, LRP1B, and FGFR2. Investigating genetic profiles in NSCLC with small intestine metastasis is crucial for early detection and prevention, and may provide insights to the underlying molecular mechanisms and potential therapeutic targets.

2 Case presentations

2.1 Case 1

The patient, a 61-year-old male, presented to the local hospital in April 2022 with complaints of dizziness and fatigue. A left upper lobe partial resection with lymph node clearance was performed after a 21 mm x 15 mm mass was found on a chest computed tomography (CT). Postoperative pathology revealed infiltrating adenocarcinoma with low to moderate differentiation. Immunohistochemistry revealed partial positive staining for CK7, focal positive staining for NapsinA, and TTF-1, negative staining for CK20. After thorough examination, the diagnosis of the malignant tumor in the left lung was confirmed as pT1cN0M0. Four cycles of PP regimen (Cisplatin + Pemetrexed) chemotherapy were given postoperatively, followed by radiotherapy (RT) of 56 GY/25 F from September 10 to November 14, 2022.

On January 6, 2023, the patient was admitted to our hospital due to abdominal pain and distension. CT examination of the abdomen and pelvis showed segmental thickening of the small intestinal wall in the right upper abdomen and multiple enlarged lymph nodes in the pelvic and inguinal regions (Figure 1A). On January 16, 2023, endoscopic examination of the small intestine revealed an ulcerated tumor in the jejunum, occupying approximately 2/3 of the lumen, located approximately 100 cm from the pylorus (Figure 1B). Histopathological examination of the biopsy specimen showed a high-grade cancer-like lesion in the submucosal layer of the small intestine. Immunohistochemistry demonstrated weakly positive expression of CK, positive expression of CK7 and Vimentin, negative expression of NapsinA, TTF-1, CK20. The diagnosis of undifferentiated cancer was considered.

From January 20 to May 19, 2023, the patient underwent 6 cycles of maintenance therapy with Tislelizumab. The therapeutic response was assessed every 2 cycles and classified as stable disease (SD) (Figure 1C).

2.2 Case 2

The patient, a 65-year-old man, was admitted to the hospital on April 6, 2021 due to a one-month history of coughing and phlegm. Chest CT scans revealed soft tissue nodules in the upper and middle lobes of the right lung, as well as soft tissue masses in the right pulmonary hilum and multiple enlarged lymph nodes in the mediastinum and right pulmonary hilum. No other malignant tumors or signs of metastasis were found in other parts of the body on PET/CT. A sample of the right lung mass and the 7th groups of lymph nodes were taken for pathological examination, which revealed poorly differentiated cancer. Immunohistochemistry

10.3389/fonc.2023.1265749

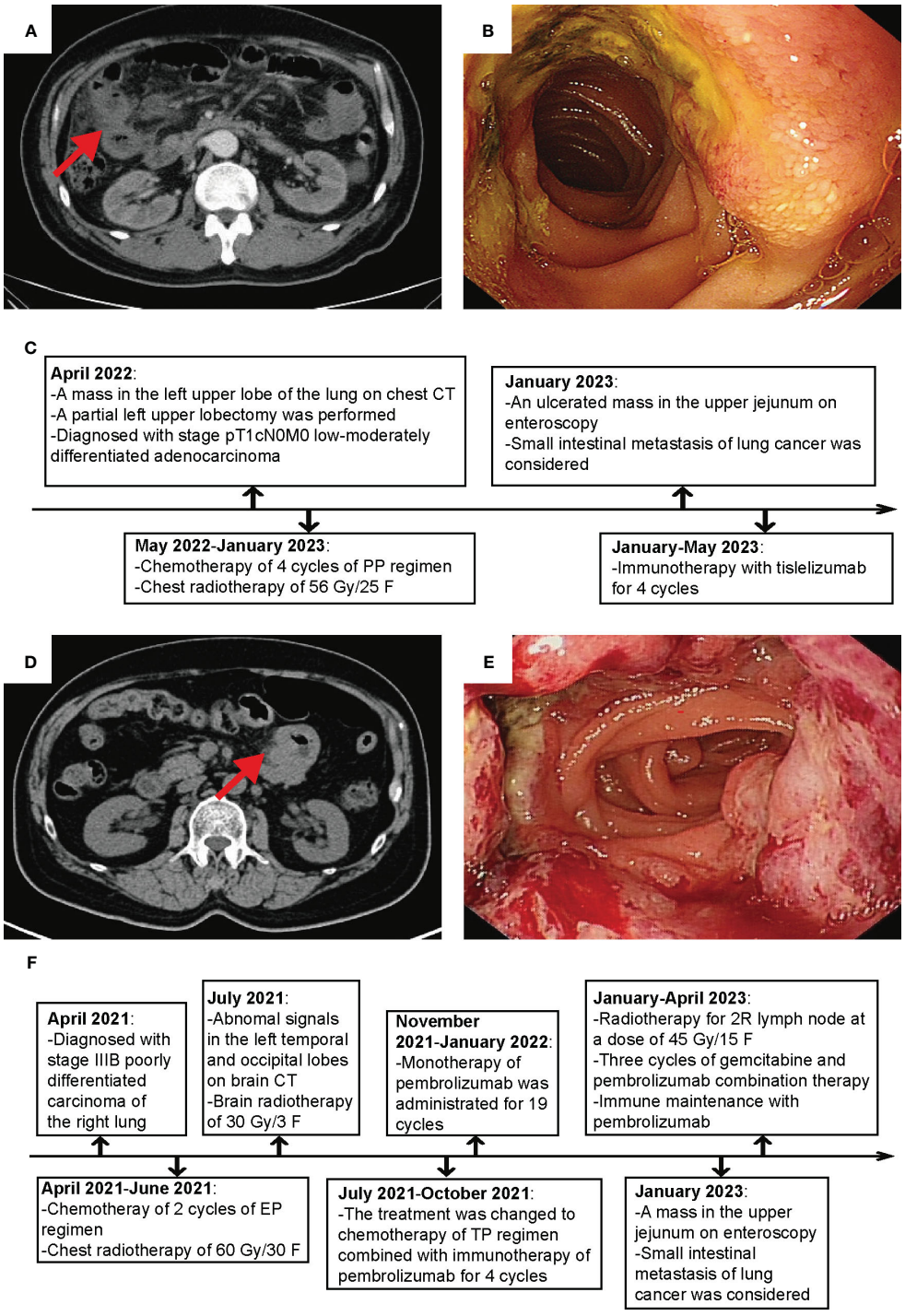


FIGURE 1
Clinical and imaging features of the 2 cases. **(A)** Abdominal CT image of Patient 1. **(B)** Small intestine endoscopy image of Patient 1. **(C)** Flowchart of Patient 1's diagnosis and treatment course. **(D)** Abdominal CT image of Patient 2. **(E)** Small bowel endoscopy image of Patient 2. **(F)** Flowchart of Patient 2's diagnosis and treatment course.

showed positive expression of CK7, along with negative expression of TTF-1. The final diagnosis was low-grade lung cancer in the right lung, staging as T4N2MO. The patient had a smoking history of more than 40 years, averagely smoking 20 cigarettes per day.

From April 29 to June 18, 2021, the patient received 2 cycles chemotherapy of EP regimen (Etoposide + Nedaplatin) and chest RT with a total dose of 60 GY/30 F. On July 21, 2021, a review of the enhanced CT scan of the brain showed abnormal signals in the left temporal and parietal lobes, suggesting metastasis. Therefore, the patient received brain RT with a total dose of 30 GY/3 F, and the treatment plan was changed to chemotherapy of TP regimen (Albumin paclitaxel + Nedaplatin) combined with immunotherapy

of Pembrolizumab for 4 cycles. The efficacy evaluation showed SD in the lungs and complete response (CR) in the brain. Starting from November 17, 2021, the patient received single-agent Pembrolizumab as maintenance for 19 cycles, and the efficacy evaluation showed SD.

On January 4, 2023, the patient returned for a abdominal CT due to hoarseness and abdominal distension, which showed multiple enlarged lymph nodes in the 2R region of the mediastinum, thickening of the wall of the jejunum in the upper left abdomen (Figure 1D). On January 13, 2023, a small intestine endoscopy revealed a luminal growth in the upper part of the jejunum, located approximately 35 cm from the pylorus (Figure 1E), and biopsy confirmed poorly differentiated cancer. Immunohistochemistry showed positive staining for CK7 and NapsinA, negative staining for TTF-1 and CK20. The possibility of poorly differentiated lung cancer with jejunal metastasis was considered.

From January 12, 2023, the patient received RT for swollen mediastinal lymph nodes in 2R region, administering a dose of 45 GY/15 F. Between February 8 and March 25, 2023, the patient underwent chemotherapy with Gemcitabine in combination with immunotherapy of Pembrolizumab for 3 cycles. Upon evaluation in April 2023, the therapeutic effect was classified as SD, which prompted a treatment regimen of immunotherapy maintenance with Pembrolizumab. The patient is currently receiving regular treatment (Figure 1F).

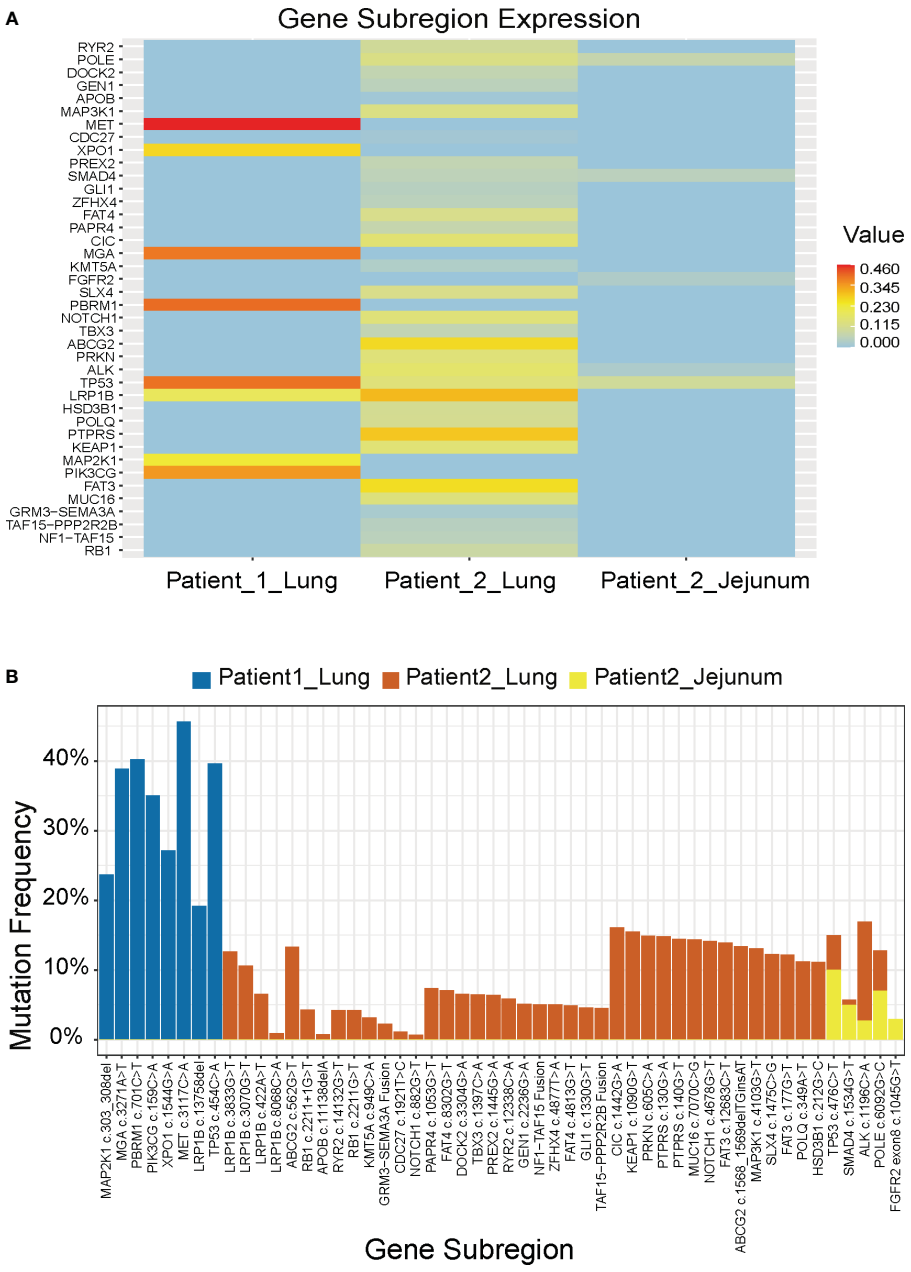


FIGURE 2
Molecular features of the 2 patients. **(A)** Mutation genes and their mutation frequencies. **(B)** Comparison of mutation sites and their mutation frequencies.

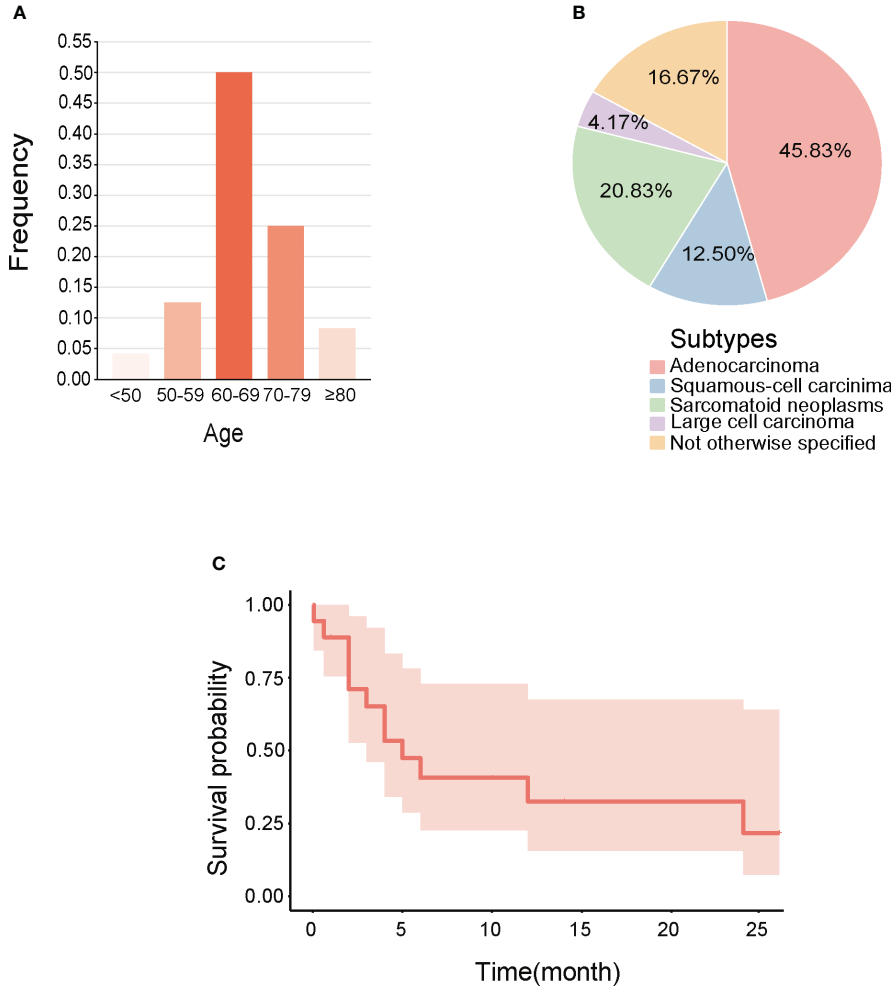


FIGURE 3
Clinical features of small Intestine metastasis from NSCLC reported in recent 5 years. **(A)** Age range. **(B)** Pathological types. **(C)** Survival curves made for integrating the prognosis.

TABLE 1 Reported cases of small intestine metastasis from primary NSCLC within the past 5 years.

Age Gender	Type of histological	Symptoms	Other metastasis	Treatment	Prognosis	Author
62 M	Adenocarcinoma	Abdominal pain, anemia	Lymph node	Chemotherapy	Died 3 months after metastasis	Ogasawara et al. (2022) (8)
86 M	Non-small cell cancer	Systemic syndrome, anemia	–	Palliative treatment	Died after 4 months	C. Saldaña-Dueñas et al. (2019) (9)
68 M	Adenocarcinoma	Anemia	Lymph node, liver	Chemotherapy	Survived at 5th month	C. Saldaña-Dueñas et al. (2019) (9)
59 F	Adenocarcinoma	Bloody stools, anaemia	Adrenal gland	Small intestine resection, immunotherapy	–	D. Kosciuszek et al. (2022) (10)
61 M	Squamous-cell carcinoma	Abdominal distention, weight loss	Stomach	Surgery, chemotherapy	Died 6 months after surgery	Li et al. (2018) (11)
60 M	Adenocarcinoma	Edema	Lymph nodes, adrenal glands, chest wall, bone	Surgery	Died at 19th day following diagnosis	Wang et al. (2018) (12)

(Continued)

TABLE 1 Continued

Age Gender	Type of histological	Symptoms	Other metastasis	Treatment	Prognosis	Author
64 M	Adenocarcinoma	Melena, weight loss, anaemia	Lymph node, liver	Palliative care	–	Kosasih et al. (2019) (13)
73 F	Adenocarcinoma	Weakness, melena, anaemia	Lymph node, liver	Chemo-radiotherapy	–	Kosasih et al. (2019) (13)
81 F	Poorly differentiated neoplasm	Abdominal pain, anaemia	Liver, sigmoid colon	Surgery	Death on the second day of surgery	P. Misiakos et al. (2019) (14)
62 M	Adenocarcinoma	Abdominal pain	Brain, bone, liver	Targeted treatment	Died 2 months later after metastasis	Zhu et al. (2021) (15)
70 M	Non-small cell cancer	Abdominal pain	Brain, liver	Surgery, chemotherapy	Aive after 1 month of follow-up	Sautto et al. (2022) (16)
41 M	Large cell carcinoma	Abdominal pain, vomiting	Lymph nodes, pleura, adrenal glands, bone	Surgery	–	Kim et al. (2020) (17)
65 M	Squamous cell carcinoma	Abdominal pain	Stomach, rectum	Surgical resection	Alive after 2 years of follow-up	Zhu et al. (2022) (18)
70 M	Pleomorphic carcinoma	Fecal occult blood, anaemia	Stomach, peritoneum	Surgery, immunotherapy	Cancer-free state at 26th month	Suzuki et al. (2022) (19)
61 M	Epidermoid carcinoma	Abdominal pain, vomiting, constipation	–	Surgery	Died 1 year later after progress	Rejab et al. (2022) (20)
69 M	Adenocarcinoma	Bloating, decreased flatus	–	Targeted treatment	Alive at 14th month of follow-up	Chen et al. (2018) (21)
75 M	Adenocarcinoma	Melena, anemia	Liver, bone, brain	Palliative care	Died 2 months later after diagnosis	Ohira et al. (2019) (22)
65 F	Adenocarcinoma	Intestinal obstruction	–	Surgery	Relapse after 2 years	Pierro et al. (2019) (23)
77 M	Adenocarcinoma	Poor oral intake, nausea, vomiting, diarrhea	–	Surgery	Died for postoperative complications	Nguyen et al. (2019) (24)
71 M	Undifferentiated pleomorphic sarcoma	Fever, upper abdominal pain	Brain	Surgery	Alive at 10th month of follow-up	Eto et al. (2021) (25)
55 M	Carcinoma with sarcomatoid differentiation	Epigastric pain, melena	Lymph nodes, adrenal gland	Surgery	Died 4 months after surgery	Xie et al. (2020) (26)
61 M	Sarcomatoid carcinoma	Melena, fever, anaemia	–	–	Died 5 months after diagnosis	Xie et al. (2020) (26)
66 M	Sarcomatoid carcinoma	Melena, anaemia	Epiglottis, lymph nodes, adrenal gland	Immunotherapy, targeted treatment	–	Gao et al. (2022) (27)
57 M	Undifferentiated sarcoma	Acute abdomen	–	Surgery	Died 2 months after surgery	Pleština et al. (2019) (28)

2.3 Genetic profiling of the 2 patients

We compiled the NGS results of both patients, including the lung tumor of patient 1 and the lung and small intestine masses of patient 2 (Figure 2A). In Patient 1, the top 5 genes with the highest mutation frequency were TP53, MGA, PBRM1, PIK3CG, and MET, whereas in Patient 2, they were TP53, ABCG2, LRP1B, FAT3, and PTPRS. Furthermore, we observed that both cases exhibited some shared genetic mutations, such as TP53 and LRP1B, while Patient 2 experienced a novel gene mutation, FGFR2, after small intestine metastasis (Figure 2B).

3 Discussion and literature review

Our analysis of reported cases small intestine metastasis from NSCLC over the past 5 years indicated that most patients were over 60 years old (Figure 3A), the primary pathological type was mostly adenocarcinoma (Figure 3B), and that the initial symptoms of

metastatic disease typically manifested as abdominal pain and anemia, with a poor prognosis that often resulted in death within 6 months (Figure 3C, Table 1).

We discovered that the mutations in these two patients were predominantly missense mutations. TP53 is the most commonly mutated gene in human cancer, and 70% of its mutations are missense in the DNA-binding domain (29). These mutations increase the incidence of distant metastasis, and are associated with reduced survival rates. TP53 serves as the major prognostic factor for early and advanced NSCLC, with an OS of 27 vs. 19 months ($p < 0.001$) (30). The co-occurring mutations with other genes such as EGFR, STK11, or KRAS have been shown to be independent biomarkers for immune checkpoint inhibitors. LRP1B was reported to reduce NSCLC cell proliferation and migration as a putative tumor suppressor, and its loss-of-function mutations might lead to oncogenesis (31). Compared with the NGS results before small intestine metastasis in patient 2, a newly emerging mutation of FGFR2 was found in the metastatic tumors. A retrospective analysis of 5,557 Chinese solid tumor patients showed that the frequency of FGFR2 alterations in lung cancer was 1%, and FGFR2 was highly expressed in NSCLC tissues, which was correlated with tumor metastasis and poor prognosis (32). Targeted therapy against FGFR2 has become a hot topic in lung cancer treatment, and the role of FGFR inhibitors including ponatinib, regorafenib, pazopanib, lenvatinib, and nintedanib in NSCLC is being intensively studied.

Between patients, there existed some distinct genetic subregions. The heterogeneity of either patient's tumor might serve as one contributing factor, whereby diverse subclonal populations or mutational patterns might even coexist within a single patient's sample. Tumor heterogeneity engenders high genetic, epigenetic, and phenotypic diversity among tumor cells, thus constituting a significant constraint on targeted therapies (33). In addition, tumor progression and evolution epitomize a dynamic process, wherein the course may vary among different patients, owing to disparities such as the tumor microenvironment. One aspect we can ascertain is that investigating shared mutated genes holds promise as a future avenue of researches.

Furthermore, the gene mutations in the primary lung tumors were essentially the same as those in the metastatic tumors. However, the mutation frequency decreased after the small intestine metastasis, leading us to speculate that the metastasis might cause selective pressure on the NSCLC cells, resulting in the survival and expansion of a specific subset of cells with lower mutation frequency. Only 20% of detected mutations were demonstrated to exhibit significant alterations following conventional therapy, including a tendency towards a decrease in mutation frequency.

NGS was not conducted on the small intestine metastasis in Patient 1 due to the inherently limited quantity of the endoscopically obtained tissue sample. The sample was utilized not only to ascertain the origin of the small intestine mass whether primary or metastatic, but also underwent additional immunohistochemical analyses. Subsequently, PD-L1 testing was performed to further establish immunotherapy indications. Determining the nature of the tumor and guiding immunotherapy decisions stand as paramount considerations for clinical practitioners. Furthermore, the incomparability of the patient's blood sample results led us to abandon the notion of utilizing blood samples.

The treatment of intestine metastasis of lung cancer is not standardized in clinical practice. For patients with multiple metastases of lung cancer, RT and chemotherapy are preferred, and surgery can be an effective palliative treatment for severe complications such as bleeding, obstruction, and perforation. RT can reduce the risk of distant metastasis in early-stage NSCLC patients. In NSCLC cases with small intestine metastasis, RT can further control the growth of metastatic lesions and alleviate abdominal pain and digestive dysfunction. The efficacy of RT is influenced by genomic molecular variations that determine the radiosensitivity of cancer cells. Moreover, RT can enhance the efficacy of systemic treatments through its abscopal effects. In cases of metastatic lung cancer, concurrent local RT and metastatic targeting therapy can strengthen the sustained control of local metastases (34). Further research is needed to determine the optimal combination and sequence of RT and other treatments, such as chemotherapy or targeted therapy.

More molecular studies are required to elucidate the underlying mechanisms driving NSCLC migration to the small intestine. The mutated genes TP53 and LRP1B, shared by the 2 patients, along with the newly mutated gene FGFR2 that has emerged after small bowel metastasis, need to be further verified as the driving force behind the spread of lung cancer to the small intestine. Targeting the mutated genes or their downstream signaling pathways represents a breakthrough for exploring novel therapeutic strategies.

4 Conclusions

Metastasis is a complex multi-step process that involves various factors such as tumor microenvironment, genetic instability, and immune response. It is essential to perform molecular marker detection in NSCLC patients and clarify the correlation between the pattern of small intestine metastasis and driver gene mutations. This not only enables the identification of tumor progression, but also serves as a biological predictor of radiation sensitivity and an identifying factor for potential therapeutic targets, thereby guiding personalized treatment plans.

5 Patient's perspective and informed consent

The emergence of small bowel metastasis signified disease progression for the patient. The patient exhibited high compliance during further examinations and treatments, and demonstrated an understanding of the importance of NGS testing. In the absence of therapeutic targets, the patient received immunotherapy. The efficacy assessments after every two cycles

showed stable disease, with an improvement in the patient's quality of life.

The patient in this case report had given consent for anonymous publication.

Data availability statement

The datasets presented in this article are not readily available because of ethical/privacy restrictions. Requests to access the datasets should be directed to the corresponding authors.

Ethics statement

The studies involving humans were approved by Zhongnan Hospital of Wuhan University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. 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

YG: Project administration, Supervision, Writing – review & editing. MW: Visualization, Writing – original draft. GC: Data curation, Writing – original draft. JL: Investigation, Writing – original draft. ZF: Investigation, Writing – original draft. YL: Investigation, Writing – original draft. CX: Project administration, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by National Natural Science Foundation of China (81800429 and 81972852), Young and Middle-Aged Medical Backbone Talents of Wuhan (WHQG201902), Translational Medicine and Interdisciplinary Research Joint Fund of Zhongnan Hospital of Wuhan University (ZNJC202007), Medical Science and Technology Innovation Platform Construction, Support Project of Zhongnan Hospital of Wuhan University (PTXM2022009), and Chinese Society of Clinical Oncology TopAlliance Tumor Immune Research Fund (Y-JS2019-036).

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

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. Sun Z, Aubry MC, Deschamps C, Marks RS, Okuno SH, Williams BA, et al. Histologic grade is an independent prognostic factor for survival in non-small cell lung cancer: an analysis of 5018 hospital- and 712 population-based cases. *J Thorac Cardiovasc Surg* (2006) 131(5):1014–20. doi: 10.1016/j.jtcvs.2005.12.057

3. Li W, Liu JB, Hou LK, Yu F, Zhang J, Wu W, et al. et al: Liquid biopsy in lung cancer: significance in diagnostics, prediction, and treatment monitoring. *Mol Cancer* (2022) 21(1):25. doi: 10.1186/s12943-021-01462-z

4. Yang CJ, Hwang JJ, Kang WY, Chong IW, Wang TH, Sheu CC, et al. Gastrointestinal metastasis of primary lung carcinoma: clinical presentations and outcome. *Lung Cancer (Amsterdam Netherlands)* (2006) 54(3):319–23. doi: 10.1016/j.lungcan.2006.08.007

5. Hu Y, Feit N, Huang Y, Xu W, Zheng S, Li X. Gastrointestinal metastasis of primary lung cancer: An analysis of 366 cases. *Oncol Lett* (2018) 15(6):9766–76. doi: 10.3892/ol.2018.8575

6. Yoshimoto A, Kasahara K, Kawashima A. Gastrointestinal metastases from primary lung cancer. *Eur J Cancer (Oxford Engl 1990)* (2006) 42(18):3157–60. doi: 10.1016/j.ejca.2006.08.030

7. Gao G, Deng L. [Association between EGFR, ALK and KRAS gene status and synchronous distant Organ metastasis in non-small cell lung cancer]. *Zhongguo fei ai za zhi = Chin J Lung Cancer* (2018) 21(7):536–42. doi: 10.3779/j.issn.1009-3419.2018.07.06

8. Ogasawara N, Ono S, Sugiyama T, Adachi K, Yamaguchi Y, Izawa S, et al. Small-intestinal metastasis from lung carcinoma. *Case Rep Gastroenterol* (2022) 16(1):195–200. doi: 10.1159/000523663

9. Saldaña-Dueñas C, Rullán-Iriarte M, Macías-Mendizábal E. Anemia como presentación inicial del cáncer de pulmón con afectación intestinal metastásica. *Rev Gastroenterología México* (2019) 84(2):259–62. doi: 10.1016/j.rgmx.2018.04.003

10. Kosciuszek ND, Noel P, Takabe K, Seitelman E, Datta R, Gunasekaran G, et al. Intraluminal small bowel metastasis from primary lung cancer. *World J Oncol* (2022) 13(6):409–16. doi: 10.14740/wjon1532

11. Li X, Li S, Ma Z, Zhao S, Wang X, Wen D. Multiple gastrointestinal metastases of squamous-cell lung cancer: A case report. *Medicine* (2018) 97(24):e11027. doi: 10.1097/MD.0000000000011027

12. Wang J, Chen Y, Zhang S, Chen Q. Perforation of small intestine secondary to metastatic lung adenocarcinoma: A case report. *Medicine* (2018) 97(49):e13469. doi: 10.1097/MD.0000000000013469

13. Kosasih S, Muhammad Nawawi KN, Wong Z, Chia Hsin DC, Ban AY, Raja Ali RA. Upper gastrointestinal bleed due to duodenal metastases of lung adenocarcinoma: Report of two cases and review of literature. *Case Rep Med* (2019) 2019:3437056. doi: 10.1155/2019/3437056

14. Misiakos EP, Gouloumi AR, Schizas D, Damaskou V, Tsapralis D, Farrugia FA, et al. Small bowel perforation with multiple intestinal metastases from lung carcinoma: A case report. *Oncol Lett* (2019) 17(4):3862–6. doi: 10.3892/ol.2019.10031

15. Zhu M, Shu J, Liu X, Leng L, Wen Q. Gastrointestinal hemorrhage caused by duodenal metastasis from a primary lung adenocarcinoma: A case report. *Mol Clin Oncol* (2021) 14(3):62. doi: 10.3892/mco.2021.2224

16. Sautto F, Tinsley S, Neychev V. Small bowel perforation due to rare metastasis from stage IV lung adenocarcinoma. *Cureus* (2022) 14(9):e29551. doi: 10.7759/cureus.29551

17. Kim MC, Jang MH, Ahn JH. Metastatic large cell carcinoma of the lung: A rare cause of acute small bowel obstruction. *Thorac Cancer* (2020) 11(11):3379–82. doi: 10.1111/1759-7714.13656

18. Zhu F, Lin J, Liao H, Wang Y, Pan J. Gene profiling of metastatic small intestinal squamous cell carcinoma after lung squamous cell carcinoma surgery: a case report. *Trans Cancer Res* (2022) 11(4):967–72. doi: 10.21037/tcr-22-481

19. Suzuki T, Noda M, Yamamura A, Ohishi H, Notsuda H, Eba S, et al. Persistent fecal occult blood due to the small intestinal metastasis of pleomorphic lung carcinoma. *J Surg Case Rep* (2022) 2022(2):rjac043. doi: 10.1093/jscr/rjac043

20. Rejab H, Fendri S, Trigui A, Abid B, Youssef M, Ketata S, et al. A rare case of small intestinal metastasis from lung cancer. *Clin Case Rep* (2022) 10(11):e6666. doi: 10.1002/ccr3.6666

21. Chen HF, Zhang QX, Zhu YC, Du KQ, Li XF, Wu LX, et al. Intestinal metastasis from primary ROS1-positive lung adenocarcinoma cancer patients responding to crizotinib. *OncoTargets Ther* (2018) 11:7821–5. doi: 10.2147/OTT.S178985

22. Ohira T, Iraha A, Kinjo T, Hokama A, Fujita J. Small intestinal metastasis from primary lung cancer. *Polish Arch Internal Med* (2019) 129(1):57–8. doi: 10.20452/pamw.4374

23. Pierro A, Cilla S, De Ninno M, Sallustio G. Recurrent jejunal metastasis from occult lung cancer. *Digestive liver Dis Off J Ital Soc Gastroenterol Ital Assoc Study Liver* (2019) 51(1):166. doi: 10.1016/j.dld.2018.07.008

24. Nguyen BD, Ryan MS. Gastrointestinal: Positron emission tomography–computed tomography imaging of multifocal intestinal metastasis from primary lung cancer. *J Gastroenterol Hepatol* (2019) 34(5):814–4. doi: 10.1111/jgh.14591

25. Eto R, Hoshino I, Takiguchi N, Tonooka T, Gunji H, Kuwayama N, et al. Surgical resection of undifferentiated pleomorphic sarcoma (UPS) with jejunal perforation, a suspected case of metastasis of lung cancer. *Clin J Gastroenterol* (2021) 14(5):1386–91. doi: 10.1007/s12328-021-01483-7

26. Xie X, Tu N, Wang Q, Cheng Z, Han X, Bu L. (18) F-FDG PET/CT imaging of small intestinal metastasis from pulmonary sarcomatoid carcinoma: brief report and review of the literature. *Thorac Cancer* (2020) 11(8):2325–30. doi: 10.1111/1759-7714.13468

27. Gao C, Zou Q, Liu H. Pulmonary sarcomatoid carcinoma with epiglottis and ileum metastasis detected by 18F-FDG PET/CT. *Clin Nucl Med* (2022) 47(3):231–3. doi: 10.1097/RLU.0000000000003936

28. Pleština S, Librenjak N, Marušić A, Batelja Vuletić L, Janevski Z, Jakopović M. An extremely rare primary sarcoma of the lung with peritoneal and small bowel metastases: a case report. *World J Surg Oncol* (2019) 17(1):147. doi: 10.1186/s12957-019-1691-8

29. Sabapathy K, Lane DP. Therapeutic targeting of p53: All mutants are equal, but some mutants are more equal than others. *Nat Rev Clin Oncol* (2018) 15(1):13–30. doi: 10.1038/nrclinonc.2017.151

30. Jiao XD, Qin BD, You P, Cai J, Zang YS. The prognostic value of TP53 and its correlation with EGFR mutation in advanced non-small cell lung cancer, an analysis based on cBioPortal data base. *Lung Cancer (Amsterdam Netherlands)* (2018) 123:70–5. doi: 10.1016/j.lungcan.2018.07.003

31. Li Y, Knisely JM, Lu W, McCormick LM, Wang J, Henkin J, et al. Low density lipoprotein (LDL) receptor-related protein 1B impairs urokinase receptor regeneration on the cell surface and inhibits cell migration. *J Biol Chem* (2002) 277(44):42366–71. doi: 10.1074/jbc.M207705200

32. Gu W, Yang J, Wang Y, Xu J, Wang X, Du F, et al. Comprehensive identification of FGFR1-4 alterations in 5 557 Chinese patients with solid tumors by next-generation sequencing. *Am J Cancer Res* (2021) 11(8):3893–906.

33. Maleki EH, Bahrami AR, Matin MM. Cancer cell cycle heterogeneity as a critical determinant of therapeutic resistance. *Genes Dis* (2024) 11(1):189–204. doi: 10.1016/j.gendis.2022.11.025

34. Kroeze SGC, Pavic M, Stellamans K, Lievens Y, Becherini C, Scorsetti M, et al. Metastases-directed stereotactic body radiotherapy in combination with targeted therapy or immunotherapy: systematic review and consensus recommendations by the EORTC-ESTRO OligoCare consortium. *Lancet Oncol* (2023) 24(3):e121–32. doi: 10.1016/S1470-2045(22)00752-5

frontiers | Frontiers in Oncology

TYPE Review
PUBLISHED 11 December 2023
DOI 10.3389/fonc.2023.1260374

Check for updates

OPEN ACCESS

EDITED BY
Hongda Liu,
Nanjing Medical University, China

REVIEWED BY
Xuemei Wen,
The University of Texas at Austin, United States
Yuwei Wang,
Shaanxi University of Chinese Medicine, China

*CORRESPONDENCE
Minal Çalışkan
mcaliskanba@dsi.com

RECEIVED 17 July 2023
ACCEPTED 16 November 2023
PUBLISHED 11 December 2023

CITATION
Çalışkan M and Tazaki K (2023) AI/ML advances in non-small cell lung cancer biomarker discovery. *Front. Oncol.* 13:1260374. doi: 10.3389/fonc.2023.1260374

COPYRIGHT
© 2023 Çalışkan and Tazaki. 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.

AI/ML advances in non-small cell lung cancer biomarker discovery

Minal Çalışkan[1]* and Koichi Tazaki[2]

[1]Translational Science Department, Precision Medicine Function, Daiichi Sankyo, Inc., Basking Ridge, NJ, United States, [2]Translational Science Department I, Precision Medicine Function, Daiichi Sankyo, Tokyo, Japan

Lung cancer is the leading cause of cancer deaths among both men and women, representing approximately 25% of cancer fatalities each year. The treatment landscape for non-small cell lung cancer (NSCLC) is rapidly evolving due to the progress made in biomarker-driven targeted therapies. While advancements in targeted treatments have improved survival rates for NSCLC patients with actionable biomarkers, long-term survival remains low, with an overall 5-year relative survival rate below 20%. Artificial intelligence/machine learning (AI/ML) algorithms have shown promise in biomarker discovery, yet NSCLC-specific studies capturing the clinical challenges targeted and emerging patterns identified using AI/ML approaches are lacking. Here, we employed a text-mining approach and identified 215 studies that reported potential biomarkers of NSCLC using AI/ML algorithms. We catalogued these studies with respect to BEST (Biomarkers, EndpointS, and other Tools) biomarker sub-types and summarized emerging patterns and trends in AI/ML-driven NSCLC biomarker discovery. We anticipate that our comprehensive review will contribute to the current understanding of AI/ML advances in NSCLC biomarker research and provide an important catalogue that may facilitate clinical adoption of AI/ML-derived biomarkers.

KEYWORDS

NSCLC, lung adenocarcacinoma, machine learning, biomarkers, non-small cell lung cancer

Introduction

Lung cancer is the leading cause of cancer deaths among both men and women (1), representing approximately 25% of cancer deaths each year (2). Lung cancer is divided into two main histological subtypes: small-cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). NSCLC constitutes approximately 85% of all lung cancer cases and is the focus of our study. The treatment landscape of NSCLC is rapidly evolving due to progress in biomarker-driven targeted therapies. Mutations in 11 genes (*EGFR*, *KRAS*, *ALK*, *ROS1*, *BRAF*, *NTRK1*, *NTRK2*, *NTRK3*, *MET*, *RET*, *ERBB2*) have been reported as FDA-recognized biomarkers predicting patients' response to targeted therapies. Similarly, IHC (Immunohistochemistry) quantified PD-L1 (*CD274*) expression, microsatellite instability,

and Tumor Mutation Burden (TMB) have been used in clinical settings to assess whether NSCLC patients could benefit from Immune Checkpoint Inhibitor (ICIs) (Table S1A).

Biomarkers are being used at an ever-increasing rate to predict disease risk, prognosis, and treatment response. Several national and international efforts have been established to standardize and catalogue disease biomarkers. BEST (Biomarkers, EndpointS, and other Tools), a joint task force between the FDA and NIH, was formed to standardize biomarker definitions in different contexts of clinical use (3). The EDRN (Early Detection Research Network) catalogues biomarkers that may improve detection of early-stage cancers (4). The Pharmacogenomics Knowledgebase (PharmGKB) curates the impact of genetic variation on drug response and catalogues pharmacogenetic biomarkers (5). The FDA regulates and catalogues pharmacogenomic biomarkers in drug labeling (6). Resources such as OncoKB, COSMIC, ClinVar, and ICGC (incorporating TCGA and Cancer Genome Project data) provide prevalence information and clinical significance assertions for genetic biomarkers in cancer (7–10). The My Cancer Genome from Vanderbilt University (11) offers an integrative database summarizing the potential clinical impact of genetic as well as protein expression and genomic instability biomarkers. Similarly, TCIA (The Cancer Imaging Archive) and IBSI (Imaging Biomarker Standardization Initiative) were formed to curate and standardize image biomarkers (12, 13). Professional societies such as the NCCN (National Comprehensive Cancer Network), ESMO (European Society for Medical Oncology), ASCO (American Society of Clinical Oncology), CAP (College of American Pathologists), IASLC (International Association for the Study of Lung Cancer), and AMP (Association for Molecular Pathology) provide clinical guideline recommendations for disease biomarker testing to help improve diagnosis and selection of targeted therapies.

These important efforts contribute to improving the delivery of personalized treatment decisions. Advancements in targeted treatments in the last 20 years have improved survival of NSCLC patients with actionable biomarkers (14). However, the long-term survival rate of NSCLC is still poor with an overall relative 5-year survival rate of less than 20% (15). Clinically utilized biomarkers for NSCLC were identified using traditional statistical approaches and are currently assumed to be mutually exclusive in therapeutic decision-making. However, there is growing evidence showing that actionable biomarkers of NSCLC can co-occur within the same patient's tumor (16–18) and it is crucial to evaluate both linear and non-linear effects of the disease biomarkers. In this regard, machine learning algorithms promise more flexible model building and the ability to recognize non-linear, complex patterns in high dimensional datasets. Notably, several AI/ML-enabled medical devices have been FDA-approved and are being used in clinical settings for automated tissue segmentation (i.e., the use of computer algorithms to identify and distinguish different structures within medical images) and feature extraction (i.e., identification of specific patterns from the medical images to aid in diagnosis) from lung CT (Computed Tomography) images (19). Similarly, several deep learning approaches have been developed to aid whole slide image analysis (20, 21) and promise to enable enhanced performance in digital pathology workflows.

Overall, machine learning algorithms have made noteworthy contributions to NSCLC diagnostic workflows and promise growing applications in biomarker discovery. In this study, we sought to evaluate trends in AI/ML applications in NSCLC biomarker research. Using a text-mining approach, we identified 215 studies that reported potential biomarkers of NSCLC using AI/ML algorithms. We catalogued these studies with respect to BEST (3) biomarker sub-types and summarized emerging patterns and trends in AI/ML-driven NSCLC biomarker discovery We emphasize that our focus in this study was to compile potential use-cases for AI/ML in NSCLC biomarker research. Therefore, we did not capture the model performance metrics of the studies we reviewed, nor did we appraise the validity of the prediction models. For quantitative sources on AI/ML model appraisal, we recommend the readers to refer to guidelines such as CHARMS (22), MLP-BIOM (23), TRIPOD (24), and PROBAST (25).

Methods

NSCLC terms

We downloaded the EMBL-EBI Experimental Factor Ontology (EFO) (26) obo file on April 28th, 2022. We extracted all disease IDs under the EFO:0003060; non-small cell lung cancer disease category. A total of 22 NSCLC sub-types (Figure 1A) and 85 additional synonymous disease IDs were present in the EFO dataset, which collectively formed the NSCLC terms category (Table S1B).

AI/ML terms

We used the following four terms to represent the AI/ML terms category: machine learning, artificial intelligence, deep learning, neural network.

Text mining strategy

We downloaded MEDLINE/PubMed abstracts in xml format from the National Library of Medicine on May 4th, 2022. We used custom Python scripts to extract PubMed IDs from abstracts that include at least one of the NSCLC and at least one of the AI/ML terms. This approach (Figure 1B) identified a total of 491 articles that likely contain the findings of AI/ML studies on NSCLC.

Literature review approach

We reviewed each of the 491 manuscripts and excluded 31 manuscripts that were not classified as a research article or were not written in English. We evaluated the remaining 460 manuscripts and identified 215 that reported AI/ML models that were developed to identify NSCLC biomarkers.

Specifically, we required that the models incorporated at least one data type aligning with the BEST Glossary (3) biomarker definition (i.e., a molecular, histologic, radiographic, or physiologic characteristic that is measured as an indicator of

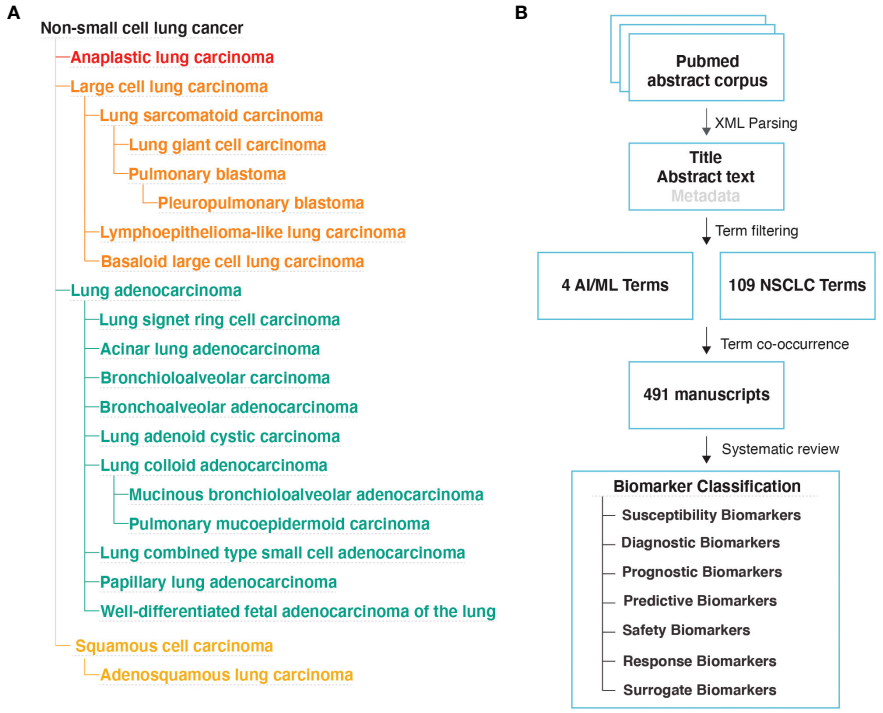


FIGURE 1
Literature Review Methodology **(A)** NSCLC disease subtypes based on the EMBL-EBI Experimental Factor Ontology (EFO) database. **(B)** Study design used to identify manuscripts that implemented AI/ML algorithms to discover new biomarkers for NSCLC.

normal biological processes, pathogenic processes, or biological responses to an exposure or intervention). Across the 215 studies, we were able to categorize biomarker data types into four broad groups: (i) Molecular Biomarkers (e.g., gene expression, genotype, DNA methylation), (ii) Histologic Biomarkers (e.g., Whole Slide Image, Cytology microphotographs), (iii) Radiologic Biomarkers (e.g., Computed Tomography (CT), Magnetic Resonance Imaging (MRI), PET/CT (Positron Emission Tomography/Computed Tomography)), and (iv) Multimodal Biomarkers (i.e., a combination of different modes or types of data).

We also required that the outcome of the AI/ML models can be categorized under one of the following seven biomarker categories: (i) Susceptibility/Risk, (ii) Diagnostic, (iii) Prognostic, (iv) Predictive, (v) Response, (vi) Safety, and (vii) Surrogate. The first six biomarker categories were defined based on the BEST Glossary (3) definitions. The last biomarker category, Surrogate Biomarkers, were defined as biomarkers that were not directly measured but were inferred using AI/ML applied to other, often less invasive, patient data (Figure 1B; Please see Box 1. Glossary, for Biomarker category definitions).

Results

AI/ML-derived susceptibility/risk biomarkers of NSCLC

The risk of developing a complex disease is explained by a combination of genetic and environmental factors. For NSCLC, cigarette smoking is the number one environmental risk factor with smokers being 15-30 times more likely to develop NSCLC than non-smokers (27). Among non-smokers, NSCLC is observed significantly more frequently in females than males, suggesting sex is a risk factor beyond cigarette smoking (28). Exposure to asbestos, radon, or other pollutants have also been reported as environmental risk factors of NSCLC (29). While NSCLC is considered a disease of the elderly with a median patient age of 70 at diagnosis, a subset of NSCLC patients (1-10%) are diagnosed at younger ages (<40 years) (30), indicating potential germline or distinct somatic driver mutations may be present in different patient age groups. Genome-wide association studies focusing on germline genetic variants have reported 16 independent loci associated with risk of developing NSCLC (31) (Table S1C). Polygenic risk score models based on the collective effect of these germline genetic variants were shown to successfully predict NSCLC risk beyond age and smoking years (32). While somatic genetic variants are important biomarkers used in selection of targeted therapies, they are not suitable for NSCLC risk assessment, as accessing lung tissue samples cannot be justified for routine risk assessment purposes. Similarly, using non-invasive genetic approaches such as circulating tumor DNA (ctDNA) sequencing is not suitable for risk prediction because ctDNA is at low concentrations even in early-stage cancers (33).

Susceptibility/Risk Biomarkers are defined as biomarkers that indicate the potential for developing a disease or medical condition in an individual who does not currently have clinically apparent disease or the medical condition (3). Using our approach

BOX 1 Glossary of biomarker types.

Descriptions of the Susceptibility/Risk, Diagnostic, Prognostic, Predictive, Response, Safety biomarkers were retrieved from the BEST (Biomarkers, EndpointS, and other Tools) Resource (3). BEST defines a biomarker as a defined characteristic that is measured as an indicator of normal biological processes, pathogenic processes, or responses to an exposure or intervention, including therapeutic interventions. Molecular, histologic, radiologic, or physiologic characteristics are types of biomarkers. *Biomarkers that predict histological disease subsets were included under the "Diagnostic Biomarkers" category. Biomarkers that predict molecular or potential molecular subsets were included under the "Surrogate Biomarkers" category.

Biomarker Type	Description
Susceptibility/ Risk Biomarkers	A biomarker that indicates the potential for developing a disease or medical condition in an individual who does not currently have clinically apparent disease or the medical condition.
Diagnostic Biomarkers*	A biomarker used to detect or confirm presence of a disease or condition of interest or to identify individuals with a subtype of the disease.
Prognostic Biomarkers	A biomarker used to identify likelihood of a clinical event, disease recurrence or progression in patients who have the disease or medical condition of interest.
Predictive Biomarkers	A biomarker used to identify individuals who are more likely than similar individuals without the biomarker to experience a favorable or unfavorable effect from exposure to a medical product or an environmental agent.
Safety Biomarkers	A biomarker measured before or after an exposure to a medical product or an environmental agent to indicate the likelihood, presence, or extent of toxicity as an adverse effect.
Response Biomarkers	A biomarker used to show that a biological response, potentially beneficial or harmful, has occurred in an individual who has been exposed to a medical product or an environmental agent.
Surrogate Biomarkers	A biomarker that was not directly measured but was inferred using AI/ML applied to other, often less invasive, patient data.

(Figure 1B), we found that machine learning studies focused primarily on integrating behavioral risk factors, family history, and environmental factors into NSCLC risk modeling. We found only one study that applied ML to identify biomarkers that could be used for NSCLC risk prediction (34) (Table S1D). In this study, Umu et al. reported that ML models of circulating serum RNA levels can predict NSCLC risk 6-8 years before manifestation of disease symptoms and provided evidence that feature selection approaches (i.e., selecting the most discriminative variables while eliminating the redundant or irrelevant variables; please see Pudjihartono et al. (35) for a summary on feature selection algorithms) and histology-specific data subsets may enhance model performance metrics [for model performance metrics including accuracy, recall, specificity, precision, F1-score, please see Hicks et al. (36)].

AI/ML-derived diagnostic biomarkers of NSCLC

Early symptoms of NSCLC including shortness of breath, fatigue, coughing, and loss of appetite are often mistaken for other conditions due to their non-specific nature. The US Preventive Services Task Force recommends annual risk screening using low-dose CT for high-risk individuals who are between 50 and 80 years old and have at least a 20-pack-year smoking history (37). However, despite these efforts approximately 55% of NSCLC patients present with locally advanced or metastatic disease at the time of diagnosis (38). When NSCLC is suspected, the initial evaluation is performed using imaging tools including chest X-ray, CT, or PET/CT scan. Diagnosis requires histological confirmation using tissue samples stained with Hematoxylin and Eosin (H&E). When tissue morphology is insufficient for proper classification, immunohistochemistry (e.g., TTF-1, Napsin A, CK7, P63, CK5/6) is recommended to aid differential diagnosis (39). While molecular testing of somatic mutations could contribute to diagnosis of NSCLC, current use cases of such testing are primarily limited to informing the treatment plans of already diagnosed patients.

Diagnostic biomarkers are defined as biomarkers that are used to detect or confirm the presence of a disease or condition of interest or to identify individuals with a subtype of the disease (3) (Figure 2A). Using the approach shown in Figure 1B, we identified 69 studies that used machine learning approaches to identify potential diagnostic biomarkers of NSCLC (Table S1E). Overall, most diagnostic efforts concentrated around building models that could be used to distinguish the two most common histological subtypes of NSCLC; lung adenocarcinoma (LAD) and squamous cell carcinoma (SCC) (41–61). Additionally, several studies have reported AI/ML models and proposed biomarkers that could be used to distinguish NSCLC or LAD from healthy control/non-malignant samples (43, 53, 59, 62–74), as well as for differential diagnosis of NSCLC and SCLC (75–78) (Figure 2B).

A body of literature has reported diagnostic AI/ML models leveraging CT or PET/CT radiologic datasets (Table S1E). These studies primarily used radiomics or CNN (Convolutional Neural Network)-based approaches to extract image features. Radiomics-based approaches are often criticized for having high variability due

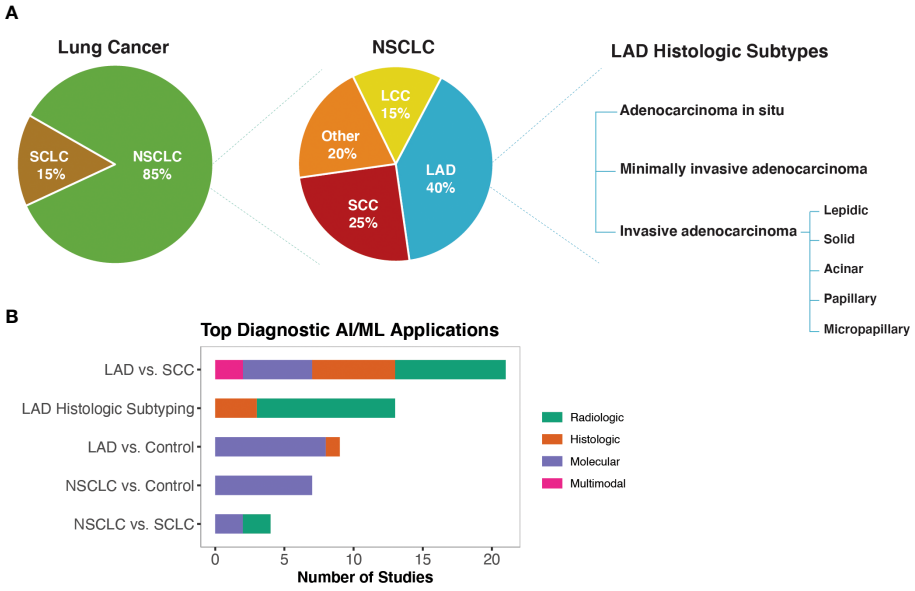


FIGURE 2
AI/ML Applications for Diagnostic Biomarker Discovery **(A)** Lung cancer, NSCLC, and LAD histologic subtypes, respectively. Subtype frequencies were retrieved from Schabath et al. (40) **(B)** Bar graph of the top five model outcomes/topics where AI/ML algorithms have been developed to identify potential diagnostic biomarkers for NSCLC. Color coding indicates the broad biomarker data type used in these studies (i.e., Molecular, Histologic, Radiologic, Multimodal data types). The acronyms are as follows: SCLC, Small Cell Lung Cancer; NSCLC, Non-Small Cell Lung Cancer; LAD, Lung Adenocarcinoma; SCC, Squamous Cell Carcinoma; LCC, Large Cell Carcinoma.

to use of manual/semi-automatic tumor segmentation techniques as well as for relying on pre-defined mathematical equations/hand-crafted features. Unlike radiomics-based approaches, CNN-based study designs often build end-to-end algorithms that automate the tissue segmentation, feature extraction, and classifier training steps. Although CNNs offer the potential to reduce human introduced bias, they require larger training datasets compared to radiomics-based approaches and offer less interpretability. In this regard, we identified several studies that integrated radiomics and CNN-based approaches to improve model prediction accuracy while providing clinical interpretability (52, 79–82).

Histology-based diagnosis of NSCLC subtypes can be complex as visual inspections by pathologists are prone to subjective assessments and may result in different interpretations. CNNs trained on H&E-stained Whole Slide Images (WSIs) have shown encouraging results for automated differential diagnosis of LAD vs. SCC as well as for histologic subtyping of LAD growth patterns (60, 83, 84). Despite these efforts, challenges related to the interpretability of CNN-based classifiers as well as computational constraints of high-resolution WSI datasets continue to be obstacles to their widespread clinical utility. Deep feature visualization (i.e., the process of generating visual representations of the features learned from deep neural networks) and resolution-based knowledge distillation (i.e., an approach to transfer knowledge from a high-resolution neural network to a smaller lower-resolution one) were among the emerging approaches to improve interpretability and computational feasibility of deep learning solutions for digital pathology (67, 85).

Molecular biomarkers, in particular somatic driver mutations, are increasingly being used to guide treatment plans for NSCLC patients. Molecular testing of tumor tissue biopsies is currently the gold standard practice to identify actionable molecular biomarkers, but the invasive nature of this process limits its use in routine diagnostic screening. Emerging non-invasive liquid biopsy tests also have limited applications for routine diagnostic screening, as ctDNA is at low concentrations in early-stage cancers (33). An ideal diagnostic biomarker requires low invasiveness and easy detection to allow early diagnosis. However, we found that ML studies that leveraged molecular biomarkers for diagnostic purposes have mainly used genome-wide gene expression data derived from lung tissue. A recurrent finding from these studies was that non-coding RNA expression signatures could differentiate NSCLC/LAD tissue from normal tissue (43, 59, 72, 73). Recapitulating known biology, ML algorithms that used lung gene expression levels to distinguish LAD vs. SCC have reported *TP63*, a known IHC marker for differentiating LAD vs. SCC, as well as several keratin-related genes (e.g., *KRT5*, *KRT6A*, *KRT14*, *SERPINB13*) among the top discriminative features (i.e., top attributes impacting model's ability to differentiate between different classes) (48, 56). In liquid biopsy-based diagnostic studies, gene-expression signatures from tumor-educated platelets and small extracellular vesicles as well as cfDNA (cell-free DNA) fragmentation patterns (DELFI score; proportion of short (100-150 bp) to long (86–155) cfDNA fragments) were reported as potential biomarkers for NSCLC diagnosis (63, 68, 76).

AI/ML-derived prognostic biomarkers of NSCLC

NSCLC prognosis has been correlated with several clinical and demographic parameters including but not limited to the histologic

subtype, disease stage, patient performance status, age, sex, blood hemoglobin and calcium levels, blood neutrophil-to-lymphocyte ratio, and serum lactate dehydrogenase and alkaline phosphatase levels (87, 88) (Figure 3A). Disease prognosis as well as the therapeutic options for NSCLC also depend on the molecular biology of the tumor (87) (Figure 3B). Similarly, Minimal Residual Disease (MRD) (i.e., small number of cancer cells that may remain in the body after cancer treatment and even when patient is in remission) levels have recently started being used in predicting NSCLC relapse risk (89).

Prognostic biomarkers are defined as biomarkers that are used to identify the likelihood of a clinical event, disease recurrence, or progression in patients who have the disease or medical condition of interest (3). Prognostic biomarkers are often confused with predictive biomarkers because predictive biomarkers are associated with prognostic outcomes in response to receiving a particular treatment. With NSCLC having FDA recognized predictive biomarkers (Table S1A), cataloguing prognostic biomarkers independent of predictive biomarkers can be misleading because biomarkers that were once associated with unfavorable outcomes can now be associated with favorable outcomes in response to targeted therapies. To identify a predictive biomarker, BEST recommends a comparison of a treatment to a control in patients with and without the biomarker (3). However, upon reviewing AI/ML studies of NSCLC biomarker research, we found that published prognostic and predictive biomarker studies are often confounded in single-arm evaluations. Acknowledging these issues, we used proxy definitions and catalogued studies as "Prognostic" when the prognostic outcomes were investigated regardless of the patients' treatment status and as "Predictive" when prognostic outcomes were investigated in patient cohorts that were exposed to a specific medical product or an environmental agent.

We identified 58 manuscripts that reported AI/ML models to identify potential prognostic biomarkers of NSCLC (Table S1F). The most frequently studied prognostic outcomes were LAD Survival (79, 87, 90–104), NSCLC Survival (105–117), LAD Lymph Node Metastasis (98, 118–124), NSCLC Staging (66, 125–130), and NSCLC Recurrence (131–133) (Figure 3C).

Time-to-event is the typical outcome variable when the metric of prognosis is a survival phenotype. However, native ML models cannot handle time-to-event data while accommodating censored observations. Reflecting this, we found that ML studies predicting NSCLC/LAD survival mainly formulated the survival analysis as a classification problem and transformed time-to-event data into dichotomized endpoints (90–94, 96, 100, 102, 103, 106, 108–111, 113, 116, 117, 134, 135). To this end, utilizing Random Survival Forests (RSF) for continuous time-to-event survival prediction and those aiming to identify optimal time-to-event ML models were emerging (98, 99, 101, 105), but further applications and research in this area are warranted.

Prognostic ML studies using CT and PET/CT datasets were primarily based on pre-treatment images (79, 93, 105, 109, 111, 115, 126, 127, 129, 130, 133, 136). These studies emphasized the need for improved multi-institution data integration and image harmonization approaches to help build robust prognostic models (105, 109, 126). Prognostic molecular and multimodal ML studies mainly leveraged tumor gene expression datasets (Table S1F). Tumor microenvironment (TME) gene expression signatures have been investigated frequently in the context of developing prognostic ML models for NSCLC (92, 96, 99, 100, 110, 137, 138). In addition to TME immune gene signatures, other components such as hypoxia, pyroptosis, and intercellular communication were prioritized to build prognostic gene models for NSCLC (92, 96, 137).

AI/ML-derived predictive biomarkers of NSCLC

Clinical response to drugs can be influenced by many factors, including patient age, sex, body mass index, concomitant therapies, genetic make-up, circadian and seasonal variations, and drug absorption, distribution, metabolism, excretion (ADME) profiles. Precision/Personalized Medicine aims to customize treatment regimens based on known variables that predict response to available therapies. Pharmacogenetics and Pharmacogenomics efforts currently are the major driving forces enabling Precision Medicine in NSCLC treatment. Mutations in 11 genes (*EGFR*, *KRAS*, *ALK*, *ROS1*, *BRAF*, *NTRK1*, *NTRK2*, *NTRK3*, *MET*, *RET*, *ERBB2*), IHC quantified PD-L1 (*CD274*) expression, microsatellite instability, and Tumor Mutation Burden (TMB) constitute the FDA recognized predictive biomarkers predicting response to NSCLC therapies (Table S1A; Figure 3B). The FDA requires that Companion Diagnostics (CDx) tests/devices are used when screening these predictive biomarkers to accurately identify patient cohorts who are likely to benefit from the therapeutic products.

BEST defines predictive biomarkers as biomarkers that are used to identify individuals who are more likely than similar individuals without the biomarker to experience a favorable or unfavorable effect from exposure to a medical product or an environmental agent (3). In this study, as described in the previous section, we used proxy definitions due to the single arm study designs of the published AI/ML-based prognostic and predictive biomarker studies. We catalogued studies as "Prognostic" when the prognostic outcomes were investigated regardless of the patients' treatment status and as "Predictive" when prognostic outcomes are investigated in patient cohorts that were exposed to a specific medical product or an environmental agent. We identified 34 manuscripts that used ML approaches to identify potential predictive biomarkers of NSCLC (Table S1G). The most frequently studied predictive outcomes were NSCLC/LAD Response to ICIs (96, 100, 139–152), NSCLC Response to Radiotherapy (86, 114, 153–156), NSCLC Response to Tyrosine Kinase Inhibitors (TKIs) (151, 157–159), and NSCLC Response to Chemoradiotherapy (160–162) (Figure 3D).

Three tumor-centric biomarkers; PD-L1 expression (≥ 1% or ≥ 50% of tumor cells), Microsatellite Instability (mutations in ≥30% of microsatellites/mismatch repair deficient), and Tumor Mutation Burden (TMB-H; ≥10 somatic mutations/Mb) are FDA-approved biomarkers for ICIs used to treat NSCLC (Figure 3B). However,

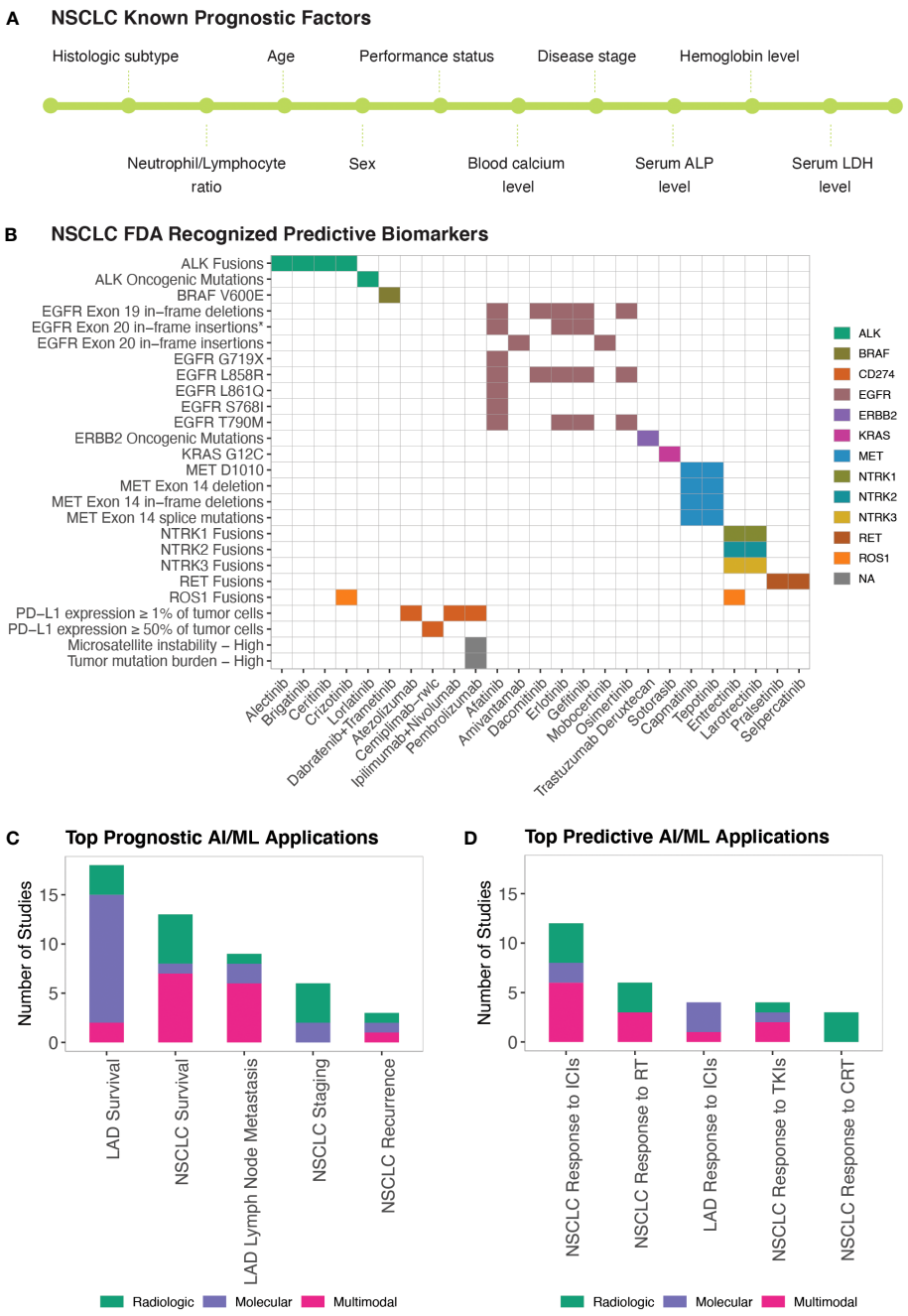


FIGURE 3
AI/ML Applications for Prognostic and Predictive Biomarker Discovery **(A)** Commonly studied prognostic factors of NSCLC. LDH and ALP stand for Lactate Dehydrogenase and Alkaline Phosphatase, respectively. **(B)** FDA-approved Predictive Biomarkers for Non-Small Cell Lung Cancer. Gene, drug names, and biomarkers were retrieved from the Table of Pharmacogenomic Biomarkers in Drug Labeling (6) in December, 2022. The list of ALK and ERBB2 oncogenic mutations is included in Table S1A. *EGFR Exon 20 in-frame insertions (excluding A763_Y764insFQEA) are drug resistance biomarkers for Erlotinib; Gefitinib; Afatinib. **(C)** Bar graph of the top five model outcomes/topics where AI/ML algorithms have been developed to identify potential prognostic biomarkers for NSCLC. Color coding indicates the broad biomarker data type used in these studies (i.e., Molecular, Histologic, Radiologic, Multimodal data types) **(D)** Bar graph of the top five model outcomes/topics where AI/ML algorithms have been developed to identify potential predictive biomarkers for NSCLC. Color coding indicates the broad biomarker data type used in these studies (i.e., Molecular, Histologic, Radiologic, Multimodal data types). ICIs, Immune Checkpoint Inhibitors; TKIs, Tyrosine Kinase Inhibitors; RT, Radiotherapy; CRT, Chemoradiotherapy.

only a fraction of the biomarker-positive NSCLC patients(20-30%) respond to ICI therapies (163). Among the ML studies we compiled, 16 reported predictive biomarkers for ICIs (96, 100, 139–152). In addition to the FDA-approved biomarkers, TMB and PD-L1 tumor proportion score (143, 145), TME-related immune gene signatures (141, 148), neutrophil-to-lymphocyte ratio (142, 143), and mutant allele tumor heterogeneity (MATH) (145) were reported as potential biomarkers predicting response to ICIs.

Mutations in eight genes (*EGFR*, *ALK*, *ROS1*, *NTRK1*, *NTRK2*, *NTRK3*, *MET*, *RET*) are FDA-approved biomarkers predicting

response to NSCLC TKIs (Figure 3B). For biomarker positive NSCLC patients, the overall response rate to TKIs is more than 60% (164, 165). Through our literature search, we identified four studies that leveraged AI/ML to identify predictive biomarkers for TKIs in NSCLC patients (151, 157–159) and two in LAD-specific cohorts (87, 166). These studies mainly reported radiomics-based predictive models (151, 157, 158, 166). There was a report of a liquid-biopsy based protein signature in patients with *ALK* rearrangements predicting response to crizotinib (159) and an OncoCast ML framework that revealed that mutations in *TP53* and *ARID1A* define a high-risk group with shorter survival in patients who received TKI therapies (87).

Stereotactic body radiation therapy (SBRT) is the standard of care treatment for early-stage NSCLC patients who are not candidates for surgery (153). For qualified patients, local control rate with SBRT treatment is around 90% but, as with surgical patients, distant failure is observed in about 20% of patients (167). Accurate prediction of response to SBRT in NSCLC patients can help identify patients who are more likely to benefit from upfront SBRT vs. systemic therapies. To this end, radiological image-only (114, 154, 155) and multimodal (86, 153, 156) classifiers have been reported, with consistent findings that CNNs demonstrate superior predictive power compared to pre-defined tumor image features (114, 154), and that inclusion of the biologically effective dose (BED) of SBRT improves predictive abilities of the models built (86, 153).

Concurrent chemoradiotherapy (CCRT) is a standard treatment option for stage II and stage III NSCLC patients with unresectable locally advanced cancer. The overall response rate to CCRT is around 80% (168). Identifying patient subsets who may benefit from intensified CCRT is important for better treatment planning. We identified three studies that reported predictive biomarkers for CRT (160–162) (Figure 3D). All three studies were based on PET/CT data and here *ad-hoc* consensus and fusion ML approaches were shown to increase the prediction accuracies of the resulting models (161, 162).

AI/ML-derived safety biomarkers of NSCLC

The FDA Adverse Event Reporting System (FAERS) is the primary surveillance tool that contains adverse events and safety concerns that are attributed to marketed drugs (169). Early detection of treatment-related adverse events is important for symptom management and successful treatment. Non-specific cancer treatments including radiotherapy and chemotherapy are mainly associated with hematologic (e.g., anemia, neutropenia, fatigue), gastrointestinal (e.g., nausea, vomiting), and dermatological (e.g., skin rashes, hair loss) toxicities. Pneumonitis is among the most severe of toxicities attributed to lung radiotherapy. Targeted therapies have a more variable spectrum of adverse events relative to non-specific cancer treatments (169, 170). TKIs are commonly associated with reversible symptoms including skin changes, vomiting, and diarrhea. However, more serious drug-specific adverse events such as interstitial lung disease and pericarditis have also been reported. Similarly, adverse events from ICIs range from reversible symptoms (e.g., dizziness, pyrexia) to more serious off-target inflammations that are referred to as immune-related adverse events (irAEs) (Figure 4A). While mild adverse effects can be managed symptomatically, moderate or severe toxicities necessitate dose reduction or treatment breaks, ultimately impacting drug efficacies. Similarly, many promising combination therapies for NSCLC face challenges due to increased toxicities of drug combinations (171), highlighting the need to identify safety biomarkers that can help tailor treatment approaches or guide the design of clinical trials.

Safety biomarkers are defined as biomarkers measured before or after an exposure to a medical product or an environmental agent to indicate the likelihood, presence, or extent of toxicity as an adverse effect (3). We found seven studies that used AI/ML to identify safety biomarkers that can be used to predict adverse events in response to NSCLC treatments. Potential safety biomarkers were reported for radiation pneumonitis (172–175), radiotherapy-induced lung fibrosis (176), ICI-induced cardiac toxicities (177), and to distinguish radiotherapy-induced vs. ICI-induced pneumonitis in patients who were treated both with radiotherapy and ICIs (178) (Figure 4B).

Radiation pneumonitis (RP) is a common (15-40%) complication of lung radiotherapy (179). The severity of RP is tracked using the National Cancer Institute Common Toxicity Criteria with radiation pneumonitis grade ≥ 2 (RP2) being symptomatic and limiting daily living activities. We identified four studies that developed ML models to predict RP2 outcome (172–175) and one study that built a classifier to distinguish radiation vs. ICI induced pneumonitis (178). Here, the concepts of integrating latent and hand-crafted variables (175), dosimetric and radiomic features (173), as well as clinical and baseline cytokine levels (174), were employed to improve the accuracy of RP2 risk prediction models.

Radiation induced lung fibrosis (RILF) is a severe side effect of radiotherapy that significantly impacts quality of life and can lead to non-cancer related death. RILF is classified from grade 0 to grade 5 depending on the clinical manifestation. Accumulating evidence suggests that genetic background as well as cytokines involved in tissue reorganization and immune response modulation are important factors contributing to RILF pathogenesis (180). We identified one ML study that built a classifier to predict RILF risk (176). This study highlighted baseline circulating CCL4 levels, along with dosimetric and clinical parameters as top discriminating features predicting grade ≥ 2 risk (176).

ICI-associated cardiotoxicity is rare but often fatal. Combination immune therapy has been shown to be a risk factor for ICI-associated cardiac events (181). Using our literature search approach, we identified one study that built ML models to predict cardiac events in patients receiving ICI therapy (177). In this study, Heilbroner et al. reported increased age, extremes of weight, presence of cardiac history, low percentage of lymphocytes, and high percentage of neutrophils among the top predictors of ICI-associated cardiotoxicity risk (177).

AI/ML-derived surrogate biomarkers of NSCLC

Identification of biomarkers from tissue biopsies is challenging due to their invasive nature of collection and small tissue volume,

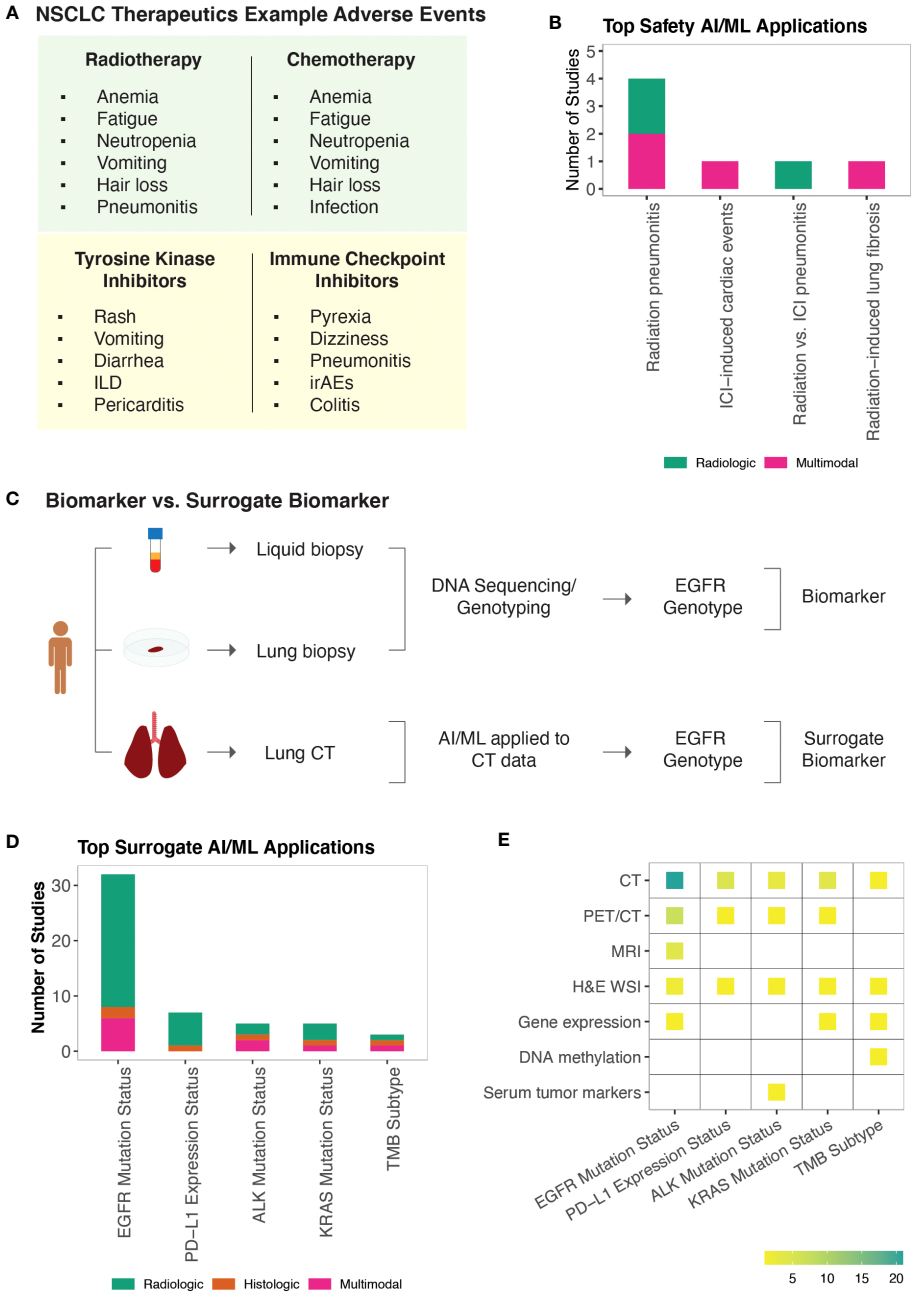


FIGURE 4
AI/ML Applications for Safety and Surrogate Biomarker Discovery **(A)** NSCLC non-specific and targeted therapies and examples of their known adverse events. Note that adverse events associated with KRAS and Serine/threonine kinase inhibitors are not included. Representative adverse events were pulled from Open Targets Pharmacovigilance tables, which are based on the FDA Adverse Event Reporting database (169, 170) **(B)** Bar graph of the model outcomes/topics where AI/ML algorithms have been developed to identify potential safety biomarkers for NSCLC. Color coding indicates the broad biomarker data type used in these studies (i.e., Molecular, Histologic, Radiologic, Multimodal data types). **(C)** An example illustration of "Biomarker Quantification" (i.e., Direct assay) vs. "Surrogate Biomarker" Prediction (i.e., AI/ML applied to indirect assay data) **(D)** Bar graph of the surrogate biomarker types where biomarker prediction was made through applying AI/ML to other indirect data types. Color coding indicates the broad biomarker data type used in these studies (i.e., Molecular, Histologic, Radiologic, Multimodal data types). **(E)** A heatmap of the indirect biomarker data types that were leveraged through AI/ML applications to predict corresponding surrogate biomarker types. Heatmap density indicates the number of ML studies that used the data types shown on the y-axis to infer NSCLC biomarkers displayed on the x-axis.

limiting their usefulness for performing repeated measurements, additional tests, and longitudinal monitoring. Non-invasive CDx assays (e.g., cobas EGFR Mutation Test v2, FoundationOne Liquid CDx, Guardant360 CDx) of ctDNA have been approved for certain NSCLC biomarkers (182). Alternative efforts continue to be pursued to detect approved and potential biomarkers of NSCLC. To this end, AI/ML algorithms have been applied to use relatively non-invasive patient data as a substitute to predict clinically

approved or other potential biomarker types, which are collectively referred as the surrogate biomarkers in this study (Figure 4C).

Under the surrogate biomarker category, we identified 60 ML studies (Table S1I). The top predicted biomarkers were *EGFR* mutation status (44, 58, 183–209), PD-L1 expression status (190, 210–215), *ALK* mutation status (94, 216–219), *KRAS* mutation status (44, 194, 197, 220), and TMB subtype (221–223) (Figures 4D, E).

In clinical practice, *EGFR* and *KRAS* mutations are routinely detected using DNA-based assays including real-time PCR and sequencing. The detection of PD-L1 expression is based on IHC assays and is considered suboptimal (224). *ALK* mutations or rearrangements can be detected through both DNA- and protein-based assays. There is currently one FDA-approved CDx test for TMB status, which is solid biopsy and sequencing based (182). As a complementary method to existing biomarker tests, we found several ML studies that have demonstrated the value of using CT or PET/CT datasets to predict *EGFR* (58, 188, 190–195, 197, 198, 200, 201, 203–209), PD-L1 (190, 211–215), *ALK* (216–219), *KRAS* (44, 190, 194, 197, 220), and to some extent TMB (223) status. Besides the promise of using radiological image data to predict surrogate molecular biomarkers, the proposed models were also shown to provide potential utility in understanding tumor heterogeneity in which biological inference from different image pixels were shown to reflect intra-tumor heterogeneity (207–209).

In addition to noninvasive radiological data, invasive yet potentially time-efficient and tissue saving alternatives were also reported to be useful in predicting surrogate biomarkers. For example, Sha et al. developed a deep learning model that could predict PD-L1 status from H&E stained WSIs in NSCLC patients (210). Similarly, Chen et al. developed ML models that were trained on H&E stained WSIs to predict multiple genetic aberrations (*ALK*, *BRAF*, *EGFR*, *ROS1* mutation status) and transcriptional subtypes (proximal-inflammatory, proximal-proliferative, terminal respiratory unit) of LAD (94), highlighting the potential of AI/ML approaches to infer different molecular characteristics through the repeated use of the same biological material.

Discussion

Biomarker discovery is a multifaceted process with many applications in healthcare such as identification of high-risk patients, improving diagnostic accuracies, as well as predicting prognostic outcomes and sensitivity to therapeutics. Despite the advancements in targeted therapies, approximately 30% of NSCLC patients do not harbor known driver mutations, and about 55% do not carry actionable mutations (225, 226). Additionally, even among patients who respond to targeted therapies, adverse events and acquired resistance may interrupt treatment plans, leading to disease progression. Expanding the repertoire of NSCLC biomarkers is critical for both the development of innovative treatments as well as for monitoring disease progression and adverse events. Here, to our knowledge, we report the first comprehensive review of AI/ML applications in the NSCLC biomarker space, catalogue the clinical challenges that are targeted by these studies, and summarize emerging patterns that could inform researchers and clinicians in this field.

Formally, ML is a sub-field of AI and the approaches used across the 215 manuscripts we catalogued could have fit under the ML category, however we used AI/ML interchangeably as this was the case in most published manuscripts. Similarly, the difference between ML and traditional statistics has been the subject of many controversies (227). ML can be built upon both statistical and algorithmic frameworks and common statistical methods can be used for both inferential and predictive modeling. We therefore relied on authors' self-declaration regarding the use of AI/ML methods in predicting potential biomarkers for NSCLC. Additionally, starting in 2014, guidelines such as CHARMS (22), MLP-BIOM (23), TRIPOD (24), and PROBAST (25) have been published to improve the reporting and appraisal of the prediction models used for diagnostic and prognostic purposes. However, we did not evaluate the manuscripts based on these checklists as our goal was to catalogue the ongoing AI/ML efforts in NSCLC biomarker research rather than assessing the immediate clinical utility of the proposed models or biomarkers.

We catalogued 215 studies identified with respect to the BEST (3) biomarker sub-types (Figure 1B). We did not find any AI/ML-derived biomarkers that could fit under the Response Biomarker category (i.e., a biomarker used to show that a biological response, potentially beneficial or harmful, has occurred in an individual who has been exposed to a medical product or an environmental agent). We instead included a new biomarker category, which we referred to as Surrogate Biomarkers, where AI/ML algorithms have been applied to relatively non-invasive patient data to predict the presence of clinically approved or other potential biomarkers of NSCLC. While biomarker discovery is often formulated as a feature selection problem (228), we also included studies that did not select features but had reported classification utility with respect to relevant organismal phenotypes under each biomarker category.

As expected, the models proposed have not been evaluated for their clinical utility. However, the clinical questions, computational challenges, and emerging solutions discussed here can serve as a reference for clinicians and data scientists leveraging biomarker datasets and AI/ML in medicine. Transfer learning methods to relax training set requirements, data harmonization algorithms to minimize technical variability in data generation, the contexts of feature selection and stability to allow interpretable models are among areas that are rapidly advancing. In addition to the stochastic nature of AI/ML models, tumor-specific temporal and spatial molecular heterogeneities, the dynamic composition of the TME, and limitations in tumor tissue access continue to further challenge the evolving landscape of biomarker modeling for NSCLC. Non-invasive approaches including liquid biopsy-based biomarkers and surrogate biomarkers inferred through the use of AI/ML, hold promise to navigate these limitations and advance our understanding of the dynamic nature of tumor progression. Longitudinal data generated through non-invasive means can, however, pose a new challenge; the data generated can be overwhelmingly large as well as complex to analyze and interpret efficiently using traditional methods. The use of automated AI/ML tools in clinical monitoring may thus be essential to facilitate

efficient analysis of the substantial amounts of longitudinal biomarker data.

Of note, while AI/ML models have been used (229) and show potential for numerous applications in clinical trials, including opportunities to enhance trial design, safety monitoring, and predictive analytics, there are currently no FDA-released guidelines or performance metrics specific to the use or evaluation of AI/ML algorithms in clinical trials (230). To this end, we anticipate that FDA guidelines for regulating AI/ML-based medical devices (231) and CONSORT-AI (232) recommendations for reporting AI-interventions in trials will facilitate the development of a formalized regulatory process, enabling effective and robust use of AI/ML in clinical trials.

Identification, cataloguing, and continuous updating of emerging biomarkers can expedite the clinical adoption of the innovative biomarkers and technologies. Here, we provided an overview of the fast-growing AI/ML applications in NSCLC biomarker discovery space and discussed the gaps and challenges in the field. By compiling relevant literature on NSCLC biomarker discovery, we revealed a comprehensive picture of the clinical challenges that are commonly targeted using AI/ML approaches and highlighted potential biomarkers and signatures that once adequately appraised may be translated into clinical decision support systems.

Author contributions

MÇ: Conceptualization, Data curation, Investigation, Methodology, Visualization, Writing – original draft, Formal analysis. KT: Methodology, Validation, Writing – review & editing, Investigation.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study received funding from Daiichi Sankyo, Inc. The funder approved publication of this article once it was conducted and prepared by the authors of this study.

Acknowledgments

The author MÇ gratefully acknowledges Daniel Baker for sharing his time and providing support during the writing of this manuscript.

Conflict of interest

Authors MÇ and KT were employed by the company Daiichi Sankyo.

The authors declare that this study received funding from Daiichi Sankyo, Inc. The funder had the following involvement in the study: approval for publication.

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.2023.1260374/full#supplementary-material

References

1. Duma N, Santana-Davila R, Molina JR. Non-small cell lung cancer: epidemiology, screening, diagnosis, and treatment. *Mayo Clin Proc* (2019) 94(8):1623–40. doi: 10.1016/j.mayocp.2019.01.013

2. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin* (2022) 72(1):7–33. doi: 10.3322/caac.21708

3. FDA-NIH Biomarker Working Group. *BEST (Biomarkers, EndpointS, and other Tools) Resource*. Silver Spring (MD) Bethesda (MD (2016).

4. Srivastava S, Wagner PD. The early detection research network: A national infrastructure to support the discovery, development, and validation of cancer biomarkers. *Cancer Epidemiol Biomarkers Prev* (2020) 29(12):2401–10. doi: 10.1158/1055-9965.EPI-20-0237

5. Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF, et al. Pharmacogenomics knowledge for personalized medicine. *Clin Pharmacol Ther* (2012) 92(4):414–7. doi: 10.1038/clpt.2012.96

6. *US Food and Drug Administration Table of Pharmacogenomic Biomarkers in Drug Labeling* Available at: https://www.fda.gov/drugs/science-and-research-drugs/table-pharmacogenomic-biomarkers-drug-labeling.

7. Chakravarty D, Gao J, Phillips SM, Kundra R, Zhang H, Wang J, et al. OncoKB: A precision oncology knowledge base. *JCO Precis Oncol* (2017) 2017. doi: 10.1200/PO.17.00011

8. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res* (2019) 47(D1):D941–D7. doi: 10.1093/nar/gky1015

9. Landrum MJ, Chitipiralla S, Brown GR, Chen C, Gu B, Hart J, et al. ClinVar: improvements to accessing data. *Nucleic Acids Res* (2020) 48(D1):D835–D44. doi: 10.1093/nar/gkz972

10. Zhang J, Bajari R, Andric D, Gerthoffert F, Lepsa A, Nahal-Bose H, et al. The international cancer genome consortium data portal. *Nat Biotechnol* (2019) 37(4):367–9. doi: 10.1038/s41587-019-0055-9

11. Jain N, Mittendorf KF, Holt M, Lenoue-Newton M, Maurer I, Miller C, et al. The My Cancer Genome clinical trial data model and trial curation workflow. *J Am Med Inform Assoc* (2020) 27(7):1057–66. doi: 10.1093/jamia/ocaa066

12. Zwanenburg A, Vallieres M, Abdalah MA, Aerts H, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. (2020) 295 (2):328–38. doi: 10.1148/radiol.2020191145

13. Prior FW, Clark K, Commean P, Freymann J, Jaffe C, Kirby J, et al. TCIA: An information resource to enable open science. *Annu Int Conf IEEE Eng Med Biol Soc* (2013) 2013:1282–5. doi: 10.1109/EMBC.2013.6609742

14. Benjamin DJ, Haslam A, Gill J, Prasad V. Targeted therapy in lung cancer: Are we closing the gap in years of life lost? *Cancer Med* (2022) 11(18):3417–24. doi: 10.1002/cam4.4703

15. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin* (2018) 68 (1):7–30. doi: 10.3322/caac.21442

16. Nagasaka M, Singh V, Baca Y, Sukari A, Kim C, Mamdani H, et al. The effects of HER2 alterations in EGFR mutant non-small cell lung cancer. *Clin Lung Cancer.* (2022) 23(1):52–9. doi: 10.1016/j.cllc.2021.08.012

17. Passaro A, Attili I, Rappa A, Vacirca D, Ranghiero A, Fumagalli C, et al. Genomic characterization of concurrent alterations in non-small cell lung cancer (NSCLC) harboring actionable mutations. *Cancers (Basel)* (2021) 13(9). doi: 10.3390/cancers13092172

18. Zhao Y, Wang S, Yang Z, Dong Y, Wang Y, Zhang L, et al. Co-occurring potentially actionable oncogenic drivers in non-small cell lung cancer. *Front Oncol* (2021) 11:665484. doi: 10.3389/fonc.2021.665484

19. *Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices* Available at: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices?utmsource=FDALinkedin#resources.

20. Senaras C, Niazi MKK, Lozanski G, Gurcan MN. DeepFocus: Detection of out-of-focus regions in whole slide digital images using deep learning. *PloS One* (2018) 13 (10):e0205387. doi: 10.1371/journal.pone.0205387

21. Wang S, Wang T, Yang L, Yang DM, Fujimoto J, Yi F, et al. ConvPath: A software tool for lung adenocarcinoma digital pathological image analysis aided by a convolutional neural network. *EBioMedicine.* (2019) 50:103–10. doi: 10.1016/j.ebiom.2019.10.033

22. Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PloS Med* (2014) 11(10):e1001744. doi: 10.1371/journal.pmed.1001744

23. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: A multidisciplinary view. *J Med Internet Res* (2016) 18(12):e323. doi: 10.2196/jmir.5870

24. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* (2015) 350:g7594. doi: 10.1161/CIRCULATIONAHA.114.014508

25. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med* (2019) 170(1):51–8. doi: 10.7326/M18-1376

26. Malone J, Holloway E, Adamusiak T, Kapushesky M, Zheng J, Kolesnikov N, et al. Modeling sample variables with an Experimental Factor Ontology. *Bioinformatics.* (2010) 26(8):1112–8. doi: 10.1093/bioinformatics/btq099

27. Shane B, Clark SA. Non small cell lung cancer. (2022).

28. Baiu I, Titan AL, Martin LW, Wolf A, Backhus L. The role of gender in non-small cell lung cancer: a narrative review. *J Thorac Dis* (2021) 13(6):3816–26. doi: 10.21037/jtd-20-3128

29. Zappa C, Mousa SA. Non-small cell lung cancer: current treatment and future advances. *Transl Lung Cancer Res* (2016) 5(3):288–300. doi: 10.21037/tlcr.2016.06.07

30. Thomas A, Chen Y, Yu T, Jakopovic M, Giaccone G. Trends and characteristics of young non-small cell lung cancer patients in the United States. *Front Oncol* (2015) 5:113. doi: 10.3389/fonc.2015.00113

31. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* (2019) 47(D1):D1005–D12. doi: 10.1093/nar/gky1120

32. Dai J, Lv J, Zhu M, Wang Y, Qin N, Ma H, et al. Identification of risk loci and a polygenic risk score for lung cancer: a large-scale prospective cohort study in Chinese populations. *Lancet Respir Med* (2019) 7(10):881–91. doi: 10.1016/S2213-2600(19)30144-4

33. Gao Q, Zeng Q, Wang Z, Li C, Xu Y, Cui P, et al. Circulating cell-free DNA for cancer early detection. *Innovation (Camb).* (2022) 3(4):100259. doi: 10.1016/j.xinn.2022.100259

34. Umu SU, Langseth H, Zuber V, Helland A, Lyle R, Rounge TB. Serum RNAs can predict lung cancer up to 10 years prior to diagnosis. *Elife.* (2022) 11. doi: 10.7554/eLife.71035

35. Pudjihartono N, Fadason T, Kempa-Liehr AW, O'Sullivan JM. A review of feature selection methods for machine learning-based disease risk prediction. *Front Bioinform* (2022) 2:927312. doi: 10.3389/fbinf.2022.927312

36. Hicks SA, Strumke I, Thambawita V, Hammou M, Riegler MA, Halvorsen P, et al. On evaluation metrics for medical applications of artificial intelligence. *Sci Rep* (2022) 12(1):5979. doi: 10.1038/s41598-022-09954-8

37. Force USPST, Krist AH, Davidson KW, Mangione CM, Barry MJ, Cabana M, et al. Screening for lung cancer: US preventive services task force recommendation statement. *JAMA.* (2021) 325(10):962–70. doi: 10.1001/jama.2021.1117

38. Zigman Suchsland M, Kowalski L, Burkhardt HA, Prado MG, Kessler LG, Yetisgen M, et al. How timely is diagnosis of lung cancer? Cohort study of individuals with lung cancer presenting in ambulatory care in the United States. *Cancers (Basel)* (2022) 14(23). doi: 10.3390/cancers14235756

39. Gurda GT, Zhang L, Wang Y, Chen L, Geddes S, Cho WC, et al. Utility of five commonly used immunohistochemical markers TTF-1, Napsin A, CK7, CK5/6 and P63 in primary and metastatic adenocarcinoma and squamous cell carcinoma of the lung: a retrospective study of 246 fine needle aspiration cases. *Clin Transl Med* (2015) 4:16. doi: 10.1186/s40169-015-0057-2

40. Schabath MB, Cote ML. Cancer progress and priorities: lung cancer. *Cancer Epidemiol Biomarkers Prev* (2019) 28(10):1563–79. doi: 10.1158/1055-9965.EPI-19-0221

41. Jain D, Nambirajan A, Chen G, Geisinger K, Hiroshima K, Layfield L, et al. NSCLC subtyping in conventional cytology: results of the international association for the study of lung cancer cytology working group survey to determine specific cytomorphologic criteria for adenocarcinoma and squamous cell carcinoma. *J Thorac Oncol* (2022) 17(6):793–805. doi: 10.1016/j.jtho.2022.02.013

42. Janssen C, Boskamp T, Hauberg-Lotte L, Behrmann J, Deininger SO, Kriegsmann M, et al. Robust subtyping of non-small cell lung cancer whole sections through MALDI mass spectrometry imaging. *Proteomics Clin Appl* (2022) 16(4): e2100068. doi: 10.1002/prca.202100068

43. Sulewska A, Niklinski J, Charkiewicz R, Karabowicz P, Biecek P, Baniecki H, et al. A signature of 14 long non-coding RNAs (lncRNAs) as a step towards precision diagnosis for NSCLC. *Cancers (Basel)* (2022) 14(2). doi: 10.3390/cancers14020439

44. Trivizakis E, Souglakos J, Karantanas A, Marias K. Deep radiotranscriptomics of non-small cell lung carcinoma for assessing molecular and histology subtypes with a data-driven analysis. *Diagnostics (Basel).* (2021) 11(12). doi: 10.3390/diagnostics11122383

45. Le Page AL, Ballot E, Truntzer C, Derangere V, Ilie A, Rageot D, et al. Using a convolutional neural network for classification of squamous and non-squamous non-small cell lung cancer based on diagnostic histopathology HES images. *Sci Rep* (2021) 11(1):23912. doi: 10.21203/rs.3.rs-646715/v1

46. Neumann JM, Freitag H, Hartmann JS, Niehaus K, Galanis M, Griesshammer M, et al. Subtyping non-small cell lung cancer by histology-guided spatial metabolomics. *J Cancer Res Clin Oncol* (2022) 148(2):351–60. doi: 10.1007/s00432-021-03834-w

47. Dehkharghanian T, Rahnamayan S, Riasatian A, Bidgoli AA, Kalra S, Zaveri M, et al. Selection, visualization, and interpretation of deep features in lung adenocarcinoma and squamous cell carcinoma. *Am J Pathol* (2021) 191(12):2172–83. doi: 10.1016/j.ajpath.2021.08.013

48. Chen JW, Dhahbi J. Lung adenocarcinoma and lung squamous cell carcinoma cancer classification, biomarker identification, and gene expression analysis using overlapping feature selection methods. *Sci Rep* (2021) 11(1):13323. doi: 10.1038/s41598-021-92725-8

49. Liu H, Jiao Z, Han W, Jing B. Identifying the histologic subtypes of non-small cell lung cancer with computed tomography imaging: a comparative study of capsule net, convolutional neural network, and radiomics. *Quant Imaging Med Surg* (2021) 11 (6):2756–65. doi: 10.21037/qims-20-734

50. Aydin N, Celik O, Aslan AF, Odabas A, Dundar E, Sahin MC. Detection of lung cancer on computed tomography using artificial intelligence applications developed by deep learning methods and the contribution of deep learning to the classification of lung carcinoma. *Curr Med Imaging.* (2021) 17(9):1137–41. doi: 10.2174/1573405617666210204210500

51. Yang F, Chen W, Wei H, Zhang X, Yuan S, Qiao X, et al. Machine learning for histologic subtype classification of non-small cell lung cancer: A retrospective multicenter radiomics study. *Front Oncol* (2020) 10:608598. doi: 10.3389/fonc.2020.608598

52. Marentakis P, Karaiskos P, Kouloulias V, Kelekis N, Argentos S, Oikonomopoulos N, et al. Lung cancer histology classification from CT images based on radiomics and deep learning models. *Med Biol Eng Comput* (2021) 59 (1):215–26. doi: 10.1007/s11517-020-02302-w

53. Wang J, Xie X, Shi J, He W, Chen Q, Chen L, et al. Denoising autoencoder, A deep learning algorithm, aids the identification of A novel molecular signature of lung adenocarcinoma. *Genomics Proteomics Bioinf* (2020) 18(4):468–80. doi: 10.1016/j.gpb.2019.02.003

54. Ren C, Zhang J, Qi M, Zhang J, Zhang Y, Song S, et al. Machine learning based on clinico-biological features integrated (18)F-FDG PET/CT radiomics for distinguishing squamous cell carcinoma from adenocarcinoma of lung. *Eur J Nucl Med Mol Imaging.* (2021) 48(5):1538–49. doi: 10.1007/s00259-020-05065-6

55. Han Y, Ma Y, Wu Z, Zhang F, Zheng D, Liu X, et al. Histologic subtype classification of non-small cell lung cancer using PET/CT images. *Eur J Nucl Med Mol Imaging.* (2021) 48(2):350–60. doi: 10.1007/s00259-020-04771-5

56. Yuan F, Lu L, Zou Q. Analysis of gene expression profiles of lung cancer subtypes with machine learning algorithms. *Biochim Biophys Acta Mol Basis Dis* (2020) 1866 (8):165822. doi: 10.1016/j.bbadis.2020.165822

57. Hyun SH, Ahn MS, Koh YW, Lee SJ. A machine-learning approach using PET-based radiomics to predict the histological subtypes of lung cancer. *Clin Nucl Med* (2019) 44(12):956–60. doi: 10.1097/RLU.0000000000002810

58. Koyasu S, Nishio M, Isoda H, Nakamoto Y, Togashi K. Usefulness of gradient tree boosting for predicting histological subtype and EGFR mutation status of non-small cell lung cancer on (18)F FDG-PET/CT. *Ann Nucl Med* (2020) 34(1):49–57. doi: 10.1007/s12149-019-01414-0

59. Sherafatian M, Arjmand F. Decision tree-based classifiers for lung cancer diagnosis and subtyping using TCGA miRNA expression data. *Oncol Lett* (2019) 18 (2):2125–31. doi: 10.3892/ol.2019.10462

60. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyo D, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med* (2018) 24(10):1559–67. doi: 10.1038/s41591-018-0177-5

61. Zhu X, Dong D, Chen Z, Fang M, Zhang L, Song J, et al. Radiomic signature as a diagnostic factor for histologic subtype classification of non-small cell lung cancer. *Eur Radiol* (2018) 28(7):2772–8. doi: 10.1007/s00330-017-5221-1

62. Wang S, Wang Q, Fan B, Gong J, Sun L, Hu B, et al. Machine learning-based screening of the diagnostic genes and their relationship with immune-cell infiltration in patients with lung adenocarcinoma. *J Thorac Dis* (2022) 14(3):699–711. doi: 10.21037/jtd-22-206

63. Li W, Zhu L, Li K, Ye S, Wang H, Wang Y, et al. Machine learning-assisted dual-marker detection in serum small extracellular vesicles for the diagnosis and prognosis prediction of non-small cell lung cancer. *Nanomaterials (Basel)* (2022) 12(5). doi: 10.3390/nano12050809

64. Li C, Tian C, Zeng Y, Liang J, Yang Q, Gu F, et al. Integrated analysis of MATH-based subtypes reveals a novel screening strategy for early-stage lung adenocarcinoma. *Front Cell Dev Biol* (2022) 10:769711. doi: 10.3389/fcell.2022.769711

65. Pedraz-Valdunciel C, Giannoukakos S, Potie N, Gimenez-Capitan A, Huang CY, Hackenberg M, et al. Digital multiplexed analysis of circular RNAs in FFPE and fresh non-small cell lung cancer specimens. *Mol Oncol* (2022) 16(12):2367–83. doi: 10.1002/1878-0261.13182

66. Lu H, Gao NL, Tong F, Wang J, Li H, Zhang R, et al. Alterations of the human lung and gut microbiomes in non-small cell lung carcinomas and distant metastasis. *Microbiol Spectr.* (2021) 9(3):e0080221. doi: 10.1128/Spectrum.00802-21

67. Lin CK, Chang J, Huang CC, Wen YF, Ho CC, Cheng YC. Effectiveness of convolutional neural networks in the interpretation of pulmonary cytologic images in endobronchial ultrasound procedures. *Cancer Med* (2021) 10(24):9047–57. doi: 10.1002/cam4.4383

68. Goswami C, Chawla S, Thakral D, Pant H, Verma P, Malik PS, et al. Molecular signature comprising 11 platelet-genes enables accurate blood-based diagnosis of NSCLC. *BMC Genomics* (2020) 21(1):744. doi: 10.1186/s12864-020-07147-z

69. Huang L, Wang L, Hu X, Chen S, Tao Y, Su H, et al. Machine learning of serum metabolic patterns encodes early-stage lung adenocarcinoma. *Nat Commun* (2020) 11 (1):3556. doi: 10.1038/s41467-019-14242-7

70. Noreldeen HAA, Du L, Li W, Liu X, Wang Y, Xu G. Serum lipidomic biomarkers for non-small cell lung cancer in nonsmoking female patients. *J Pharm BioMed Anal* (2020) 185:113220. doi: 10.1016/j.jpba.2020.113220

71. Shen N, Du J, Zhou H, Chen N, Pan Y, Hoheisel JD, et al. A diagnostic panel of DNA methylation biomarkers for lung adenocarcinoma. *Front Oncol* (2019) 9:1281. doi: 10.3389/fonc.2019.01281

72. Smolander J, Stupnikov A, Glazko G, Dehmer M, Emmert-Streib F. Comparing biological information contained in mRNA and non-coding RNAs for classification of lung cancer patients. *BMC Cancer.* (2019) 19(1):1176. doi: 10.1186/s12885-019-6338-1

73. Wang Y, Fu J, Wang Z, Lv Z, Fan Z, Lei T. Screening key lncRNAs for human lung adenocarcinoma based on machine learning and weighted gene co-expression network analysis. *Cancer biomark* (2019) 25(4):313–24. doi: 10.3233/CBM-190225

74. Xiao Y, Wu J, Lin Z, Zhao X. A deep learning-based multi-model ensemble method for cancer prediction. *Comput Methods Programs Biomed* (2018) 153:1–9. doi: 10.1016/j.cmpb.2017.09.005

75. Xie Z, Zhang H. Analysis of the diagnosis model of peripheral non-small-cell lung cancer under computed tomography images. *J Healthc Eng.* (2022) 2022:3107965. doi: 10.1155/2022/3107965

76. Mathios D, Johansen JS, Cristiano S, Medina JE, Phallen J, Larsen KR, et al. Detection and characterization of lung cancer using cell-free DNA fragmentomes. *Nat Commun* (2021) 12(1):5060. doi: 10.1038/s41467-021-24994-w

77. Chen BT, Chen Z, Ye N, Mambetsariev I, Fricke J, Daniel E, et al. Differentiating peripherally-located small cell lung cancer from non-small cell lung cancer using a CT radiomic approach. *Front Oncol* (2020) 10:593. doi: 10.3389/fonc.2020.00593

78. O'Shea K, Cameron SJ, Lewis KE, Lu C, Mur LA. Metabolomic-based biomarker discovery for non-invasive lung cancer screening: A case study. *Biochim Biophys Acta* (2016) 1860(11 Pt B):2682–7. doi: 10.1016/j.bbagen.2016.07.007

79. Wang C, Shao J, Lv J, Cao Y, Zhu C, Li J, et al. Deep learning for predicting subtype classification and survival of lung adenocarcinoma on computed tomography. *Transl Oncol* (2021) 14(8):101141. doi: 10.1016/j.tranon.2021.101141

80. Wang X, Li Q, Cai J, Wang W, Xu P, Zhang Y, et al. Predicting the invasiveness of lung adenocarcinomas appearing as ground-glass nodule on CT scan using multi-task learning and deep radiomics. *Transl Lung Cancer Res* (2020) 9(4):1397–406. doi: 10.21037/tlcr-20-370

81. Xia X, Gong J, Hao W, Yang T, Lin Y, Wang S, et al. Comparison and fusion of deep learning and radiomics features of ground-glass nodules to predict the invasiveness risk of stage-I lung adenocarcinomas in CT scan. *Front Oncol* (2020) 10:418. doi: 10.3389/fonc.2020.00418

82. Lu L, Wang D, Wang L, E L, Guo P, Li Z, et al. A quantitative imaging biomarker for predicting disease-free-survival-associated histologic subgroups in lung adenocarcinoma. *Eur Radiol* (2020) 30(7):3614–23. doi: 10.1007/s00330-020-06663-6

83. Wei JW, Tafe LJ, Linnik YA, Vaickus LJ, Tomita N, Hassanpour S. Pathologist-level classification of histologic patterns on resected lung adenocarcinoma slides with deep neural networks. *Sci Rep* (2019) 9(1):3358. doi: 10.1038/s41598-019-40041-7

84. Gertych A, Swiderska-Chadaj Z, Ma Z, Ing N, Markiewicz T, Cierniak S, et al. Convolutional neural networks can accurately distinguish four histologic growth patterns of lung adenocarcinoma in digital slides. *Sci Rep* (2019) 9(1):1483. doi: 10.1038/s41598-018-37638-9

85. DiPalma J, Suriawinata AA, Tafe LJ, Torresani L, Hassanpour S. Resolution-based distillation for efficient histology image classification. *Artif Intell Med* (2021) 119:102136. doi: 10.1016/j.artmed.2021.102136

86. Hindocha S, Charlton TG, Linton-Reid K, Hunter B, Chan C, Ahmed M, et al. A comparison of machine learning methods for predicting recurrence and death after curative-intent radiotherapy for non-small cell lung cancer: Development and validation of multivariable clinical prediction models. *EBioMedicine*. (2022) 77:103911. doi: 10.1016/j.ebiom.2022.103911

87. Shen R, Martin A, Ni A, Hellmann M, Arbour KC, Jordan E, et al. Harnessing clinical sequencing data for survival stratification of patients with metastatic lung adenocarcinomas. *JCO Precis Oncol* (2019) 3. doi: 10.1200/PO.18.00307

88. Shimizu K, Okita R, Saisho S, Maeda A, Nojima Y, Nakata M. Preoperative neutrophil/lymphocyte ratio and prognostic nutritional index predict survival in patients with non-small cell lung cancer. *World J Surg Oncol* (2015) 13:291. doi: 10.1186/s12957-015-0710-7

89. Frisone D, Friedlaender A, Addeo A. The role and impact of minimal residual disease in NSCLC. *Curr Oncol Rep* (2021) 23(12):136. doi: 10.1007/s11912-021-01131-w

90. Dessie EY, Chang JG, Chang YS. A nine-gene signature identification and prognostic risk prediction for patients with lung adenocarcinoma using novel machine learning approach. *Comput Biol Med* (2022) 145:105493. doi: 10.1016/j.compbiomed.2022.105493

91. Liu Y, Yang M, Sun W, Zhang M, Sun J, Wang W, et al. Developing prognostic gene panel of survival time in lung adenocarcinoma patients using machine learning. *Transl Cancer Res* (2020) 9(6):3860–9. doi: 10.21037/tcr-19-2739

92. Liu LP, Lu L, Zhao QQ, Kou QJ, Jiang ZZ, Gui R, et al. Identification and validation of the pyroptosis-related molecular subtypes of lung adenocarcinoma by bioinformatics and machine learning. *Front Cell Dev Biol* (2021) 9:756340. doi: 10.3389/fcell.2021.756340

93. Cho HH, Lee HY, Kim E, Lee G, Kim J, Kwon J, et al. Radiomics-guided deep neural networks stratify lung adenocarcinoma prognosis from CT scans. *Commun Biol* (2021) 4(1):1286. doi: 10.1038/s42003-021-02814-7

94. Chen L, Zeng H, Xiang Y, Huang Y, Luo Y, Ma X. Histopathological images and multi-omics integration predict molecular characteristics and survival in lung adenocarcinoma. *Front Cell Dev Biol* (2021) 9:720110. doi: 10.3389/fcell.2021.720110

95. Min KW, Kim DH, Noh YK, Son BK, Kwon MJ, Moon JY. Cancer-associated fibroblasts are associated with poor prognosis in solid type of lung adenocarcinoma in a machine learning analysis. *Sci Rep* (2021) 11(1):16779. doi: 10.1038/s41598-021-96344-1

96. Shi R, Bao X, Unger K, Sun J, Lu S, Manapov F, et al. Identification and validation of hypoxia-derived gene signatures to predict clinical outcomes and therapeutic responses in stage I lung adenocarcinoma patients. *Theranostics.* (2021) 11(10):5061–76. doi: 10.7150/thno.56202

97. Deng F, Shen L, Wang H, Zhang L. Classify multicategory outcome in patients with lung adenocarcinoma using clinical, transcriptomic and clinico-transcriptomic data: machine learning versus multinomial models. *Am J Cancer Res* (2020) 10 (12):4624–39.

98. Cai Q, He B, Zhang P, Zhao Z, Peng X, Zhang Y, et al. Exploration of predictive and prognostic alternative splicing signatures in lung adenocarcinoma using machine learning methods. *J Transl Med* (2020) 18(1):463. doi: 10.1186/s12967-020-02635-y

99. Xue L, Bi G, Zhan C, Zhang Y, Yuan Y, Fan H. Development and validation of a 12-gene immune relevant prognostic signature for lung adenocarcinoma through machine learning strategies. *Front Oncol* (2020) 10:835. doi: 10.3389/fonc.2020.00835

100. Bao X, Shi R, Zhao T, Wang Y. Mast cell-based molecular subtypes and signature associated with clinical outcome in early-stage lung adenocarcinoma. *Mol Oncol* (2020) 14(5):917–32. doi: 10.1002/1878-0261.12670

101. Ma B, Geng Y, Meng F, Yan G, Song F. Identification of a sixteen-gene prognostic biomarker for lung adenocarcinoma using a machine learning method. *J Cancer.* (2020) 11(5):1288–98. doi: 10.7150/jca.34585

102. Li Y, Ge D, Gu J, Xu F, Zhu Q, Lu C. A large cohort study identifying a novel prognosis prediction model for lung adenocarcinoma through machine learning strategies. *BMC Cancer.* (2019) 19(1):886. doi: 10.1186/s12885-019-6101-7

103. Shin B, Park S, Hong JH, An HJ, Chun SH, Kang K, et al. Cascaded wx: A novel prognosis-related feature selection framework in human lung adenocarcinoma transcriptomes. *Front Genet* (2019) 10:662. doi: 10.3389/fgene.2019.00662

104. Choi H, Na KJ. A risk stratification model for lung cancer based on gene coexpression network and deep learning. *BioMed Res Int* (2018) 2018:2914280. doi: 10.1155/2018/2914280

105. Amini M, Hajianfar G, Hadadi Avval A, Nazari M, Deevband MR, Oveisi M, et al. Overall survival prognostic modelling of non-small cell lung cancer patients using positron emission tomography/computed tomography harmonised radiomics features: the quest for the optimal machine learning algorithm. *Clin Oncol (R Coll Radiol).* (2022) 34(2):114–27. doi: 10.1016/j.clon.2021.11.014

106. Chen W, Hou X, Hu Y, Huang G, Ye X, Nie S. A deep learning- and CT image-based prognostic model for the prediction of survival in non-small cell lung cancer. *Med Phys* (2021) 48(12):7946–58. doi: 10.1002/mp.15302

107. Yuan Q, Cai T, Hong C, Du M, Johnson BE, Lanuti M, et al. Performance of a machine learning algorithm using electronic health record data to identify and estimate survival in a longitudinal cohort of patients with lung cancer. *JAMA Netw Open* (2021) 4(7):e2114723. doi: 10.1001/jamanetworkopen.2021.14723

108. Lin T, Mai J, Yan M, Li Z, Quan X, Chen X. A nomogram based on CT deep learning signature: A potential tool for the prediction of overall survival in resected non-small cell lung cancer patients. *Cancer Manag Res* (2021) 13:2897–906. doi: 10.2147/CMAR.S299020

109. Mukherjee P, Zhou M, Lee E, Schicht A, Balagurunathan Y, Napel S, et al. A shallow convolutional neural network predicts prognosis of lung cancer patients in multi-institutional CT-image data. *Nat Mach Intell* (2020) 2(5):274–82. doi: 10.1038/s42256-020-0173-6

110. Xu J, Nie H, He J, Wang X, Liao K, Tu L, et al. Using machine learning modeling to explore new immune-related prognostic markers in non-small cell lung cancer. *Front Oncol* (2020) 10:550002. doi: 10.3389/fonc.2020.550002

111. Wolsztynski E, O'Sullivan J, Hughes NM, Mou T, Murphy P, O'Sullivan F, et al. Combining structural and textural assessments of volumetric FDG-PET uptake in NSCLC. *IEEE Trans Radiat Plasma Med Sci* (2019) 3(4):421–33. doi: 10.1109/TRPMS.2019.2912433

112. Takahashi S, Asada K, Takasawa K, Shimoyama R, Sakai A, Bolatkan A, et al. Predicting deep learning based multi-omics parallel integration survival subtypes in lung cancer using reverse phase protein array data. *Biomolecules* (2020) 10(10). doi: 10.3390/biom10101460

113. Lai YH, Chen WN, Hsu TC, Lin C, Tsao Y, Wu S. Overall survival prediction of non-small cell lung cancer by integrating microarray and clinical data with deep learning. *Sci Rep* (2020) 10(1):4679. doi: 10.1038/s41598-020-61588-w

114. Baek S, He Y, Allen BG, Buatti JM, Smith BJ, Tong L, et al. Deep segmentation networks predict survival of non-small cell lung cancer. *Sci Rep* (2019) 9(1):17286. doi: 10.1038/s41598-019-53461-2

115. Sun W, Jiang M, Dang J, Chang P, Yin FF. Effect of machine learning methods on predicting NSCLC overall survival time based on Radiomics analysis. *Radiat Oncol* (2018) 13(1):197. doi: 10.1186/s13014-018-1140-9

116. Chatzimichail E, Matthaios D, Bouros D, Karakitsos P, Romanidis K, Kakolyris S, et al. gamma -H2AX: A novel prognostic marker in a prognosis prediction model of patients with early operable non-small cell lung cancer. *Int J Genomics* (2014) 2014:160236. doi: 10.1155/2014/160236

117. Hanai T, Yatabe Y, Nakayama Y, Takahashi T, Honda H, Mitsudomi T, et al. Prognostic models in patients with non-small-cell lung cancer using artificial neural networks in comparison with logistic regression. *Cancer Sci* (2003) 94(5):473–7. doi: 10.1111/j.1349-7006.2003.tb01467.x

118. Chen W, Xu M, Sun Y, Ji C, Chen L, Liu S, et al. Integrative predictive models of computed tomography texture parameters and hematological parameters for lymph node metastasis in lung adenocarcinomas. *J Comput Assist Tomogr.* (2022) 46(2):315–24. doi: 10.1097/RCT.0000000000001264

119. Lv J, Chen X, Liu X, Du D, Lv W, Lu L, et al. Imbalanced data correction based PET/CT radiomics model for predicting lymph node metastasis in clinical stage T1 lung adenocarcinoma. *Front Oncol* (2022) 12:788968. doi: 10.3389/fonc.2022.788968

120. Chong Y, Wu Y, Liu J, Han C, Gong L, Liu X, et al. Clinicopathological models for predicting lymph node metastasis in patients with early-stage lung adenocarcinoma: the application of machine learning algorithms. *J Thorac Dis* (2021) 13(7):4033–42. doi: 10.21037/jtd-21-98

121. Li Q, He XQ, Fan X, Zhu CN, Lv JW, Luo TY. Development and validation of a combined model for preoperative prediction of lymph node metastasis in peripheral lung adenocarcinoma. *Front Oncol* (2021) 11:675877. doi: 10.3389/fonc.2021.675877

122. Zhao X, Wang X, Xia W, Li Q, Zhou L, Li Q, et al. A cross-modal 3D deep learning for accurate lymph node metastasis prediction in clinical stage T1 lung adenocarcinoma. *Lung Cancer.* (2020) 145:10–7. doi: 10.1016/j.lungcan.2020.04.014

123. Liu C, Huang Q, Ma W, Qi L, Wang Y, Qu T, et al. A combination of tumor and molecular markers predicts a poor prognosis in lung adenocarcinoma. *Int J Clin Exp Pathol* (2019) 12(5):1690–701.

124. Wang Y, Deng H, Xin S, Zhang K, Shi R, Bao X. Prognostic and predictive value of three DNA methylation signatures in lung adenocarcinoma. *Front Genet* (2019) 10:349. doi: 10.3389/fgene.2019.00349

125. Zhao X, Wang X, Xia W, Zhang R, Jian J, Zhang J, et al. 3D multi-scale, multi-task, and multi-label deep learning for prediction of lymph node metastasis in T1 lung adenocarcinoma patients' CT images. *Comput Med Imaging Graph.* (2021) 93:101987. doi: 10.1016/j.compmedimag.2021.101987

126. Ubaldi L, Valenti V, Borgese RF, Collura G, Fantacci ME, Ferrera G, et al. Strategies to develop radiomics and machine learning models for lung cancer stage and histology prediction using small data samples. *Phys Med* (2021) 90:13–22. doi: 10.1016/j.ejmp.2021.08.015

127. Choi J, Cho HH, Kwon J, Lee HY, Park H. A cascaded neural network for staging in non-small cell lung cancer using pre-treatment CT. *Diagnostics (Basel)* (2021) 11(6). doi: 10.3390/diagnostics11061047

128. Miller HA, Yin X, Smith SA, Hu X, Zhang X, Yan J, et al. Evaluation of disease staging and chemotherapeutic response in non-small cell lung cancer from patient tumor-derived metabolomic data. *Lung Cancer.* (2021) 156:20–30. doi: 10.1016/j.lungcan.2021.04.012

129. Moitra D, Mandal RK. Automated AJCC (7th edition) staging of non-small cell lung cancer (NSCLC) using deep convolutional neural network (CNN) and recurrent neural network (RNN). *Health Inf Sci Syst* (2019) 7(1):14. doi: 10.1007/s13755-019-0077-1

130. Yu L, Tao G, Zhu L, Wang G, Li Z, Ye J, et al. Prediction of pathologic stage in non-small cell lung cancer using machine learning algorithm based on CT image feature analysis. *BMC Cancer.* (2019) 19(1):464. doi: 10.1186/s12885-019-5646-9

131. Kirienko M, Sollini M, Corbetta M, Voulaz E, Gozzi N, Interlenghi M, et al. Radiomics and gene expression profile to characterise the disease and predict outcome in patients with lung cancer. *Eur J Nucl Med Mol Imaging.* (2021) 48(11):3643–55. doi: 10.1007/s00259-021-05371-7

132. Luo R, Song J, Xiao X, Xie Z, Zhao Z, Zhang W, et al. Identifying CpG methylation signature as a promising biomarker for recurrence and immunotherapy in non-small-cell lung carcinoma. *Aging* (2020) 12(14):14649–76. doi: 10.18632/aging.103517

133. Ahn HK, Lee H, Kim SG, Hyun SH. Pre-treatment (18)F-FDG PET-based radiomics predict survival in resected non-small cell lung cancer. *Clin Radiol* (2019) 74(6):467–73. doi: 10.1016/j.crad.2019.02.008

134. Lee B, Chun SH, Hong JH, Woo IS, Kim S, Jeong JW, et al. DeepBTS: prediction of recurrence-free survival of non-small cell lung cancer using a time-binned deep neural network. *Sci Rep* (2020) 10(1):1952. doi: 10.1038/s41598-020-58722-z

135. Hu J, Yu H, Sun L, Yan Y, Zhang L, Jiang G, et al. Identification of an individualized metabolism prognostic signature and related therapy regimens in early stage lung adenocarcinoma. *Front Oncol* (2021) 11:650853. doi: 10.3389/fonc.2021.650853

136. Kim H, Goo JM, Lee KH, Kim YT, Park CM. Preoperative CT-based deep learning model for predicting disease-free survival in patients with lung adenocarcinomas. *Radiology.* (2020) 296(1):216–24. doi: 10.1148/radiol.2020192764

137. Zhao QY, Liu LP, Lu L, Gui R, Luo YW. A novel intercellular communication-associated gene signature for prognostic prediction and clinical value in patients with lung adenocarcinoma. *Front Genet* (2021) 12:702424. doi: 10.3389/fgene.2021.702424

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

139. Chen Y, Miao S, Zhao W. Identification and validation of significant gene mutations to predict clinical benefit of immune checkpoint inhibitors in lung adenocarcinoma. *Am J Transl Res* (2021) 13(3):1051–63.

140. Peng J, Zou D, Gong W, Kang S, Han L. Deep neural network classification based on somatic mutations potentially predicts clinical benefit of immune checkpoint blockade in lung adenocarcinoma. *Oncoimmunology.* (2020) 9(1):1734156. doi: 10.1080/2162402X.2020.1734156

141. Li C, Tian C, Zeng Y, Liang J, Yang Q, Gu F, et al. Machine learning and bioinformatics analysis revealed classification and potential treatment strategy in stage 3-4 NSCLC patients. *BMC Med Genomics* (2022) 15(1):33. doi: 10.1186/s12920-022-01184-1

142. Benzekry S, Grangeon M, Karlsen M, Alexa M, Bicalho-Frazeto I, Chaleat S, et al. Machine learning for prediction of immunotherapy efficacy in non-small cell lung cancer from simple clinical and biological data. *Cancers (Basel)* (2021) 13(24). doi: 10.3390/cancers13246210

143. Ahn BC, So JW, Synn CB, Kim TH, Kim JH, Byeon Y, et al. Clinical decision support algorithm based on machine learning to assess the clinical response to anti-programmed death-1 therapy in patients with non-small-cell lung cancer. *Eur J Cancer.* (2021) 153:179–89. doi: 10.1016/j.ejca.2021.05.019

144. Rounis K, Makrakis D, Papadaki C, Monastirioti A, Vamvakas L, Kalbakis K, et al. Prediction of outcome in patients with non-small cell lung cancer treated with second line PD-1/PDL-1 inhibitors based on clinical parameters: Results from a prospective, single institution study. *PloS One* (2021) 16(6):e0252537. doi: 10.1371/journal.pone.0252537

145. Jiang J, Jin Z, Zhang Y, Peng L, Zhang Y, Zhu Z, et al. Robust prediction of immune checkpoint inhibition therapy for non-small cell lung cancer. *Front Immunol* (2021) 12:646874. doi: 10.3389/fimmu.2021.646874

146. Trebeschi S, Bodalal Z, Boellaard TN, Tareco Bucho TM, Drago SG, Kurilova I, et al. Prognostic value of deep learning-mediated treatment monitoring in lung cancer patients receiving immunotherapy. *Front Oncol* (2021) 11:609054. doi: 10.3389/fonc.2021.609054

147. Yang Y, Yang J, Shen L, Chen J, Xia L, Ni B, et al. A multi-omics-based serial deep learning approach to predict clinical outcomes of single-agent anti-PD-1/PD-L1 immunotherapy in advanced stage non-small-cell lung cancer. *Am J Transl Res* (2021) 13(2):743–56.

148. Wiesweg M, Mairinger F, Reis H, Goetz M, Kollmeier J, Misch D, et al. Machine learning reveals a PD-L1-independent prediction of response to immunotherapy of non-small cell lung cancer by gene expression context. *Eur J Cancer.* (2020) 140:76–85. doi: 10.1016/j.ejca.2020.09.015

149. Arbour KC, Luu AT, Luo J, Rizvi H, Plodkowski AJ, Sakhi M, et al. Deep learning to estimate RECIST in patients with NSCLC treated with PD-1 blockade. *Cancer Discovery* (2021) 11(1):59–67. doi: 10.1158/2159-8290.CD-20-0419

150. Muller M, Hummelink K, Hurkmans DP, Niemeijer AN, Monkhorst K, Roder J, et al. A serum protein classifier identifying patients with advanced non-small cell lung cancer who derive clinical benefit from treatment with immune checkpoint inhibitors. *Clin Cancer Res* (2020) 26(19):5188–97. doi: 10.1158/1078-0432.CCR-20-0538

151. Dercle L, Fronheiser M, Lu L, Du S, Hayes W, Leung DK, et al. Identification of non-small cell lung cancer sensitive to systemic cancer therapies using radiomics. *Clin Cancer Res* (2020) 26(9):2151–62. doi: 10.1158/1078-0432.CCR-19-2942

152. Khorrami M, Prasanna P, Gupta A, Patil P, Velu PD, Thawani R, et al. Changes in CT radiomic features associated with lymphocyte distribution predict overall survival and response to immunotherapy in non-small cell lung cancer. *Cancer Immunol Res* (2020) 8(1):108–19. doi: 10.1158/2326-6066.CIR-19-0476

153. Avanzo M, Gagliardi V, Stancanello J, Blanck O, Pirrone G, El Naqa I, et al. Combining computed tomography and biologically effective dose in radiomics and deep learning improves prediction of tumor response to robotic lung stereotactic body radiation therapy. *Med Phys* (2021) 48(10):6257–69. doi: 10.1002/mp.15178

154. Hosny A, Parmar C, Coroller TP, Grossmann P, Zeleznik R, Kumar A, et al. Deep learning for lung cancer prognostication: A retrospective multi-cohort radiomics study. *PloS Med* (2018) 15(11):e1002711. doi: 10.1371/journal.pmed.1002711

155. Li S, Yang N, Li B, Zhou Z, Hao H, Folkert MR, et al. A pilot study using kernelled support tensor machine for distant failure prediction in lung SBRT. *Med Image Anal* (2018) 50:106–16. doi: 10.1016/j.media.2018.09.004

156. Jayasurya K, Fung G, Yu S, Dehing-Oberije C, De Ruysscher D, Hope A, et al. Comparison of Bayesian network and support vector machine models for two-year survival prediction in lung cancer patients treated with radiotherapy. *Med Phys* (2010) 37(4):1401–7. doi: 10.1118/1.3352709

157. Zhu JM, Sun L, Wang L, Zhou TC, Yuan Y, Zhen X, et al. Radiomics combined with clinical characteristics predicted the progression-free survival time in first-line targeted therapy for advanced non-small cell lung cancer with EGFR mutation. *BMC Res Notes.* (2022) 15(1):140. doi: 10.1186/s13104-022-06019-x

158. Tang X, Li Y, Yan WF, Qian WL, Pang T, Gong YL, et al. Machine learning-based CT radiomics analysis for prognostic prediction in metastatic non-small cell lung cancer patients with EGFR-T790M mutation receiving third-generation EGFR-TKI osimertinib treatment. *Front Oncol* (2021) 11:719919. doi: 10.3389/fonc.2021.719919

159. Couetoux du Tertre M, Marques M, McNamara S, Gambaro K, Hoffert C, Tremblay L, et al. Discovery of a putative blood-based protein signature associated with response to ALK tyrosine kinase inhibition. *Clin Proteomics.* (2020) 17:5. doi: 10.1186/s12014-020-9269-6

160. Yoo J, Lee J, Cheon M, Woo SK, Ahn MJ, Pyo HR, et al. Predictive value of (18) F-FDG PET/CT using machine learning for pathological response to neoadjuvant concurrent chemoradiotherapy in patients with stage III non-small cell lung cancer. *Cancers (Basel)* (2022) 14(8). doi: 10.3390/cancers14081987

161. Sepehri S, Tankyevych O, Iantsen A, Visvikis D, Hatt M, Cheze Le Rest C. Accurate Tumor Delineation vs. Rough Volume of Interest Analysis for (18)F-FDG PET/CT Radiomics-Based Prognostic Modeling inNon-Small Cell Lung Cancer. *Front Oncol* (2021) 11:726865. doi: 10.3389/fonc.2021.726865

162. Sepehri S, Tankyevych O, Upadhaya T, Visvikis D, Hatt M, Cheze Le Rest C. Comparison and fusion of machine learning algorithms for prospective validation of PET/CT radiomic features prognostic value in stage II-III non-small cell lung cancer. *Diagnostics (Basel).* (2021) 11(4). doi: 10.3390/diagnostics11040675

163. Kanwal B, Biswas S, Seminara RS, Jeet C. Immunotherapy in advanced non-small cell lung cancer patients: ushering chemotherapy through the checkpoint inhibitors? *Cureus* (2018) 10(9):e3254. doi: 10.7759/cureus.3254

164. Xu J, Zhang X, Yang H, Ding G, Jin B, Lou Y, et al. Comparison of outcomes of tyrosine kinase inhibitor in first- or second-line therapy for advanced non-small-cell lung cancer patients with sensitive EGFR mutations. *Oncotarget.* (2016) 7(42):68442–8. doi: 10.18632/oncotarget.12035

165. Gao G, Ren S, Li A, Xu J, Xu Q, Su C, et al. Epidermal growth factor receptor-tyrosine kinase inhibitor therapy is effective as first-line treatment of advanced non-small-cell lung cancer with mutated EGFR: A meta-analysis from six phase III randomized controlled trials. *Int J Cancer.* (2012) 131(5):E822–9. doi: 10.1002/ijc.27396

166. Ravanelli M, Agazzi GM, Ganeshan B, Roca E, Tononcelli E, Bettoni V, et al. CT texture analysis as predictive factor in metastatic lung adenocarcinoma treated with tyrosine kinase inhibitors (TKIs). *Eur J Radiol* (2018) 109:130–5. doi: 10.1016/j.ejrad.2018.10.016

167. Sebastian NT, Xu-Welliver M, Williams TM. Stereotactic body radiation therapy (SBRT) for early stage non-small cell lung cancer (NSCLC): contemporary insights and advances. *J Thorac Dis* (2018) 10(Suppl 21):S2451–S64. doi: 10.21037/jtd.2018.04.52

168. Conibear J, AstraZeneca UKL. Rationale for concurrent chemoradiotherapy for patients with stage III non-small-cell lung cancer. *Br J Cancer.* (2020) 123(Suppl 1):10–7. doi: 10.1038/s41416-020-01070-6

169. *FDA Adverse Event Reporting System (FAERS) Publich Dashboard* Available at: https://fis.fda.gov/sense/app/95239e26-e0be-42d9-a960-9a5f7f1c25ee/sheet/7a47a261-d58b-4203-a8aa-6d3021737452/state/analysis.

170. Ochoa D, Hercules A, Carmona M, Suveges D, Baker J, Malangone C, et al. The next-generation Open Targets Platform: reimagined, redesigned, rebuilt. *Nucleic Acids Res* (2023) 51(D1):D1353–D9. doi: 10.1093/nar/gkac1046

171. Kalra A, Rashdan S. The toxicity associated with combining immune check point inhibitors with tyrosine kinase inhibitors in patients with non-small cell lung cancer. *Front Oncol* (2023) 13:1158417. doi: 10.3389/fonc.2023.1158417

172. Katsuta Y, Kadoya N, Mouri S, Tanaka S, Kanai T, Takeda K, et al. Prediction of radiation pneumonitis with machine learning using 4D-CT based dose-function features. *J Radiat Res* (2022) 63(1):71–9. doi: 10.1093/jrr/rrab097

173. Kawahara D, Imano N, Nishioka R, Ogawa K, Kimura T, Nakashima T, et al. Prediction of radiation pneumonitis after definitive radiotherapy for locally advanced non-small cell lung cancer using multi-region radiomics analysis. *Sci Rep* (2021) 11 (1):16232. doi: 10.1038/s41598-021-95643-x

174. Yu H, Wu H, Wang W, Jolly S, Jin JY, Hu C, et al. Machine learning to build and validate a model for radiation pneumonitis prediction in patients with non-small cell lung cancer. *Clin Cancer Res* (2019) 25(14):4343–50. doi: 10.1158/1078-0432.CCR-18-1084

175. Cui S, Luo Y, Tseng HH, Ten Haken RK, El Naqa I. Combining handcrafted features with latent variables in machine learning for prediction of radiation-induced lung damage. *Med Phys* (2019) 46(5):2497–511. doi: 10.1002/mp.13497

176. Yu H, Lam KO, Wu H, Green M, Wang W, Jin JY, et al. Weighted-support vector machine learning classifier of circulating cytokine biomarkers to predict radiation-induced lung fibrosis in non-small-cell lung cancer patients. *Front Oncol* (2020) 10:601979. doi: 10.3389/fonc.2020.601979

177. Heilbroner SP, Few R, Mueller J, Chalwa J, Charest F, Suryadevara S, et al. Predicting cardiac adverse events in patients receiving immune checkpoint inhibitors: a machine learning approach. *J Immunother Cancer* (2021) 9(10). doi: 10.1136/jitc-2021-002545

178. Chen X, Sheikh K, Nakajima E, Lin CT, Lee J, Hu C, et al. Radiation versus immune checkpoint inhibitor associated pneumonitis: distinct radiologic morphologies. *Oncologist.* (2021) 26(10):e1822–e32. doi: 10.1002/onco.13900

179. Jain V, Berman AT. Radiation pneumonitis: old problem, new tricks. *Cancers (Basel).* (2018) 10(7). doi: 10.3390/cancers10070222

180. Jarzebska N, Karetnikova ES, Markov AG, Kasper M, Rodionov RN, Spieth PM. Scarred lung. An update on radiation-induced pulmonary fibrosis. *Front Med (Lausanne).* (2020) 7:585756. doi: 10.3389/fmed.2020.585756

181. Mahmood SS, Fradley MG, Cohen JV, Nohria A, Reynolds KL, Heinzerling LM, et al. Myocarditis in patients treated with immune checkpoint inhibitors. *J Am Coll Cardiol* (2018) 71(16):1755–64. doi: 10.1016/j.jacc.2018.02.037

182. *List of Cleared or Approved Companion Diagnostic Devices (In vitro and Imaging Tools)* Available at: https://www.fda.gov/medical-devices/in-vitro-diagnostics/list-cleared-or-approved-companion-diagnostic-devices-in-vitro-and-imaging-tools.

183. Huang X, Sun Y, Tan M, Ma W, Gao P, Qi L, et al. Three-dimensional convolutional neural network-based prediction of epidermal growth factor receptor expression status in patients with non-small cell lung cancer. *Front Oncol* (2022) 12:772770. doi: 10.3389/fonc.2022.772770

184. Li S, Luo T, Ding C, Huang Q, Guan Z, Zhang H. Detailed identification of epidermal growth factor receptor mutations in lung adenocarcinoma: Combining radiomics with machine learning. *Med Phys* (2020) 47(8):3458–66. doi: 10.1002/mp.14238

185. Hong D, Xu K, Zhang L, Wan X, Guo Y. Radiomics signature as a predictive factor for EGFR mutations in advanced lung adenocarcinoma. *Front Oncol* (2020) 10:28. doi: 10.3389/fonc.2020.00028

186. Xiong JF, Jia TY, Li XY, Yu W, Xu ZY, Cai XW, et al. Identifying epidermal growth factor receptor mutation status in patients with lung adenocarcinoma by three-dimensional convolutional neural networks. *Br J Radiol* (2018) 91(1092):20180334. doi: 10.1259/bjr.20180334

187. Rios Velazquez E, Parmar C, Liu Y, Coroller TP, Cruz G, Stringfield O, et al. Somatic mutations drive distinct imaging phenotypes in lung cancer. *Cancer Res* (2017) 77(14):3922–30. doi: 10.1158/0008-5472.CAN-17-0122

188. Liu Y, Zhou J, Wu J, Wang W, Wang X, Guo J, et al. Development and validation of machine learning models to predict epidermal growth factor receptor mutation in non-small cell lung cancer: A multi-center retrospective radiomics study. *Cancer Control.* (2022) 29:10732748221092926. doi: 10.1177/10732748221092926

189. Haim O, Abramov S, Shofty B, Fanizzi C, DiMeco F, Avisdris N, et al. Predicting EGFR mutation status by a deep learning approach in patients with non-small cell lung cancer brain metastases. *J Neurooncol.* (2022) 157(1):63–9. doi: 10.1007/s11060-022-03946-4

190. Wang C, Xu X, Shao J, Zhou K, Zhao K, He Y, et al. Deep learning to predict EGFR mutation and PD-L1 expression status in non-small-cell lung cancer on computed tomography images. *J Oncol* (2021) 2021:5499385. doi: 10.1155/2021/5499385

191. Gui D, Song Q, Song B, Li H, Wang M, Min X, et al. AIR-Net: A novel multi-task learning method with auxiliary image reconstruction for predicting EGFR mutation status on CT images of NSCLC patients. *Comput Biol Med* (2022) 141:105157. doi: 10.1016/j.compbiomed.2021.105157

192. Yang X, Liu M, Ren Y, Chen H, Yu P, Wang S, et al. Using contrast-enhanced CT and non-contrast-enhanced CT to predict EGFR mutation status in NSCLC patients-a radiomics nomogram analysis. *Eur Radiol* (2022) 32(4):2693–703. doi: 10.1007/s00330-021-08366-y

193. Song J, Ding C, Huang Q, Luo T, Xu X, Chen Z, et al. Deep learning predicts epidermal growth factor receptor mutation subtypes in lung adenocarcinoma. *Med Phys* (2021) 48(12):7891–9. doi: 10.1002/mp.15307

194. Le NQK, Kha QH, Nguyen VH, Chen YC, Cheng SJ, Chen CY. Machine learning-based radiomics signatures for EGFR and KRAS mutations prediction in non-small-cell lung cancer. *Int J Mol Sci* (2021) 22(17). doi: 10.3390/ijms22179254

195. Yin G, Wang Z, Song Y, Li X, Chen Y, Zhu L, et al. Prediction of EGFR mutation status based on (18)F-FDG PET/CT imaging using deep learning-based model in lung adenocarcinoma. *Front Oncol* (2021) 11:709137. doi: 10.3389/fonc.2021.709137

196. Ren M, Yang H, Lai Q, Shi D, Liu G, Shuang X, et al. MRI-based radiomics analysis for predicting the EGFR mutation based on thoracic spinal metastases in lung adenocarcinoma patients. *Med Phys* (2021) 48(9):5142–51. doi: 10.1002/mp.15137

197. Dong Y, Hou L, Yang W, Han J, Wang J, Qiang Y, et al. Multi-channel multi-task deep learning for predicting EGFR and KRAS mutations of non-small cell lung cancer on CT images. *Quant Imaging Med Surg* (2021) 11(6):2354–75. doi: 10.21037/qims-20-600

198. Zhang B, Qi S, Pan X, Li C, Yao Y, Qian W, et al. Deep CNN model using CT radiomics feature mapping recognizes EGFR gene mutation status of lung adenocarcinoma. *Front Oncol* (2020) 10:598721. doi: 10.3389/fonc.2020.598721

199. Jiang X, Ren M, Shuang X, Yang H, Shi D, Lai Q, et al. Multiparametric MRI-based radiomics approaches for preoperative prediction of EGFR mutation status in spinal bone metastases in patients with lung adenocarcinoma. *J Magn Reson Imaging.* (2021) 54(2):497–507. doi: 10.1002/jmri.27579

200. Rossi G, Barabino E, Fedeli A, Ficarra G, Coco S, Russo A, et al. Radiomic detection of EGFR mutations in NSCLC. *Cancer Res* (2021) 81(3):724–31. doi: 10.1158/0008-5472.CAN-20-0999

201. Mu W, Jiang L, Zhang J, Shi Y, Gray JE, Tunali I, et al. Non-invasive decision support for NSCLC treatment using PET/CT radiomics. *Nat Commun* (2020) 11 (1):5228. doi: 10.1038/s41467-020-19116-x

202. Park YW, An C, Lee J, Han K, Choi D, Ahn SS, et al. Diffusion tensor and postcontrast T1-weighted imaging radiomics to differentiate the epidermal growth factor receptor mutation status of brain metastases from non-small cell lung cancer. *Neuroradiology.* (2021) 63(3):343–52. doi: 10.1007/s00234-020-02529-2

203. Liu Q, Sun D, Li N, Kim J, Feng D, Huang G, et al. Predicting EGFR mutation subtypes in lung adenocarcinoma using (18)F-FDG PET/CT radiomic features. *Transl Lung Cancer Res* (2020) 9(3):549–62. doi: 10.21037/tlcr.2020.04.17

204. Nair JKR, Saeed UA, McDougall CC, Sabri A, Kovacina B, Raidu BVS, et al. Radiogenomic models using machine learning techniques to predict EGFR mutations in non-small cell lung cancer. *Can Assoc Radiol J* (2021) 72(1):109–19. doi: 10.1177/0846537119899526

205. Li X, Yin G, Zhang Y, Dai D, Liu J, Chen P, et al. Predictive power of a radiomic signature based on (18)F-FDG PET/CT images for EGFR mutational status in NSCLC. *Front Oncol* (2019) 9:1062. doi: 10.3389/fonc.2019.01062

206. Zhao W, Yang J, Ni B, Bi D, Sun Y, Xu M, et al. Toward automatic prediction of EGFR mutation status in pulmonary adenocarcinoma with 3D deep learning. *Cancer Med* (2019) 8(7):3532–43. doi: 10.1002/cam4.2233

207. Li XY, Xiong JF, Jia TY, Shen TL, Hou RP, Zhao J, et al. Detection of epithelial growth factor receptor (EGFR) mutations on CT images of patients with lung adenocarcinoma using radiomics and/or multi-level residual convolutionary neural networks. *J Thorac Dis* (2018) 10(12):6624–35. doi: 10.21037/jtd.2018.11.03

208. Wang S, Shi J, Ye Z, Dong D, Yu D, Zhou M, et al. Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning. *Eur Respir J* (2019) 53(3). doi: 10.1183/13993003.00986-2018

209. Li Y, Lu L, Xiao M, Dercle L, Huang Y, Zhang Z, et al. CT slice thickness and convolution kernel affect performance of a radiomic model for predicting EGFR status in non-small cell lung cancer: A preliminary study. *Sci Rep* (2018) 8(1):17913. doi: 10.1038/s41598-018-36421-0

210. Sha L, Osinski BL, Ho IY, Tan TL, Willis C, Weiss H, et al. Multi-field-of-view deep learning model predicts nonsmall cell lung cancer programmed death-ligand 1 status from whole-slide hematoxylin and eosin images. *J Pathol Inform.* (2019) 10:24. doi: 10.4103/jpi.jpi_24_19

211. Wang C, Ma J, Shao J, Zhang S, Li J, Yan J, et al. Non-invasive measurement using deep learning algorithm based on multi-source features fusion to predict PD-L1 expression and survival in NSCLC. *Front Immunol* (2022) 13:828560. doi: 10.3389/fimmu.2022.828560

212. Jiang Z, Dong Y, Yang L, Lv Y, Dong S, Yuan S, et al. CT-based hand-crafted radiomic signatures can predict PD-L1 expression levels in non-small cell lung cancer: a two-center study. *J Digit Imaging.* (2021) 34(5):1073–85. doi: 10.1007/s10278-021-00484-9

213. Mu W, Jiang L, Shi Y, Tunali I, Gray JE, Katsoulakis E, et al. Non-invasive measurement of PD-L1 status and prediction of immunotherapy response using deep learning of PET/CT images. *J Immunother Cancer* (2021) 9(6). doi: 10.1136/jitc-2020-002118

214. Tian P, He B, Mu W, Liu K, Liu L, Zeng H, et al. Assessing PD-L1 expression in non-small cell lung cancer and predicting responses to immune checkpoint inhibitors using deep learning on computed tomography images. *Theranostics.* (2021) 11 (5):2098–107. doi: 10.7150/thno.48027

215. Zhu Y, Liu YL, Feng Y, Yang XY, Zhang J, Chang DD, et al. A CT-derived deep neural network predicts for programmed death ligand-1 expression status in advanced lung adenocarcinomas. *Ann Transl Med* (2020) 8(15):930. doi: 10.21037/atm-19-4690

216. Chang C, Sun X, Wang G, Yu H, Zhao W, Ge Y, et al. A machine learning model based on PET/CT radiomics and clinical characteristics predicts ALK rearrangement status in lung adenocarcinoma. *Front Oncol* (2021) 11:603882. doi: 10.3389/fonc.2021.603882

217. Song Z, Liu T, Shi L, Yu Z, Shen Q, Xu M, et al. The deep learning model combining CT image and clinicopathological information for predicting ALK fusion status and response to ALK-TKI therapy in non-small cell lung cancer patients. *Eur J Nucl Med Mol Imaging.* (2021) 48(2):361–71. doi: 10.1007/s00259-020-04986-6

218. Ma DN, Gao XY, Dan YB, Zhang AN, Wang WJ, Yang G, et al. Evaluating solid lung adenocarcinoma anaplastic lymphoma kinase gene rearrangement using noninvasive radiomics biomarkers. *Onco Targets Ther* (2020) 13:6927–35. doi: 10.2147/OTT.S257798

219. Song L, Zhu Z, Mao L, Li X, Han W, Du H, et al. Clinical, Conventional CT and radiomic feature-based machine learning models for predicting ALK rearrangement status in lung adenocarcinoma patients. *Front Oncol* (2020) 10:369. doi: 10.3389/fonc.2020.00369

220. Zhang T, Xu Z, Liu G, Jiang B, de Bock GH, Groen HJM, et al. Simultaneous identification of EGFR,KRAS,ERBB2, and TP53 mutations in patients with non-small cell lung cancer by machine learning-derived three-dimensional radiomics. *Cancers (Basel)* (2021) 13(8). doi: 10.3390/cancers13081814

221. Sadhwani A, Chang HW, Behrooz A, Brown T, Auvigne-Flament I, Patel H, et al. Comparative analysis of machine learning approaches to classify tumor mutation burden in lung adenocarcinoma using histopathology images. *Sci Rep* (2021) 11 (1):16605. doi: 10.1038/s41598-021-95747-4

222. Wang J, Chen P, Su M, Zhong G, Zhang S, Gou D. Integrative modeling of multiomics data for predicting tumor mutation burden in patients with lung cancer. *BioMed Res Int* (2022) 2022:2698190. doi: 10.1155/2022/2698190

223. He B, Dong D, She Y, Zhou C, Fang M, Zhu Y, et al. Predicting response to immunotherapy in advanced non-small-cell lung cancer using tumor mutational burden radiomic biomarker. *J Immunother Cancer* (2020) 8(2). doi: 10.1136/jitc-2020-000550

224. Wang M, Wang S, Trapani JA, Neeson PJ. Challenges of PD-L1 testing in non-small cell lung cancer and beyond. *J Thorac Dis* (2020) 12(8):4541–8. doi: 10.21037/jtd-2019-itm-010

225. Taha T, Khoury R, Brenner R, Nasrallah H, Shofaniyeh I, Yousef S, et al. Treatment of rare mutations in patients with lung cancer. *Biomedicines* (2021) 9(5). doi: 10.3390/biomedicines9050534

226. Simarro J, Murria R, Perez-Simo G, Llop M, Mancheno N, Ramos D, et al. Development, implementation and assessment of molecular diagnostics by next generation sequencing in personalized treatment of cancer: experience of a public reference healthcare hospital. *Cancers (Basel)* (2019) 11(8). doi: 10.3390/cancers11081196

227. Breiman L. Statistical modeling: the two cultures. *Stat Science.* (2001) 16(3):199–215.

228. Climente-Gonzalez H, Azencott CA, Kaski S, Yamada M. Block HSIC Lasso: model-free biomarker detection for ultra-high dimensional data. *Bioinformatics* (2019) 35(14):i427–i35. doi: 10.1093/bioinformatics/btz333

229. Plana D, Shung DL, Grimshaw AA, Saraf A, Sung JJY, Kann BH. Randomized clinical trials of machine learning interventions in health care: A systematic review. *JAMA Netw Open* (2022) 5(9):e2233946. doi: 10.1001/jamanetworkopen.2022.33946

230. Weissler EH, Naumann T, Andersson T, Ranganath R, Elemento O, Luo Y, et al. The role of machine learning in clinical research: transforming the future of evidence generation. *Trials.* (2021) 22(1):537. doi: 10.1186/s13063-021-05489-x

231. *Artificial Intelligence and Machine Learning in Software as a Medical Device* Available at: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device.

232. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, Spirit AI, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med* (2020) 26(9):1364–74. doi: 10.1038/s41591-020-1034-x

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 19 March 2024
DOI 10.3389/fonc.2024.1277206

Check for updates

OPEN ACCESS

EDITED BY
Sharon R. Pine,
University of Colorado Anschutz Medical Campus, United States

REVIEWED BY
Liang Zhao,
Johns Hopkins Medicine, United States
Masa Zdralevic,
University of Montenegro, Montenegro

*CORRESPONDENCE
Jiye Aa
jiyea@cpu.edu.cn
Wei Zhu
zhuwei@njmu.edu.cn

†These authors have contributed equally to this work

RECEIVED 14 August 2023
ACCEPTED 04 March 2024
PUBLISHED 19 March 2024

CITATION
Yu M, Wen W, Wang Y, Shan X, Yi X, Zhu W, Aa J and Wang G (2024) Plasma metabolomics reveals risk factors for lung adenocarcinoma.
Front. Oncol. 14:1277206.
doi: 10.3389/fonc.2024.1277206

COPYRIGHT
© 2024 Yu, Wen, Wang, Shan, Yi, Zhu, Aa 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.

Plasma metabolomics reveals risk factors for lung adenocarcinoma

Mengjie Yu[1†], Wei Wen[2†], Yue Wang[1†], Xia Shan[3], Xin Yi[1], Wei Zhu[4*], Jiye Aa[1*] and Guangji Wang[1]

[1]Key Laboratory of Drug Metabolism & Pharmacokinetics, China Pharmaceutical University, Nanjing, Jiangsu, China, [2]Department of Thoracic Surgery, First Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu, China, [3]Department of Respiration, The Affiliated Jiangning Hospital of Nanjing Medical University, Nanjing, Jiangsu, China, [4]Department of Oncology, First Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu, China

Background: Metabolic reprogramming plays a significant role in the advancement of lung adenocarcinoma (LUAD), yet the precise metabolic changes remain incompletely understood. This study aims to uncover metabolic indicators associated with the progression of LUAD.

Methods: A total of 1083 subjects were recruited, including 670 LUAD, 135 benign lung nodules (BLN) and 278 healthy controls (HC). Gas chromatography-mass spectrometry (GC/MS) was used to identify and quantify plasma metabolites. Odds ratios (ORs) were calculated to determine LUAD risk factors, and machine learning algorithms were utilized to differentiate LUAD from BLN.

Results: High levels of oxalate, glycolate, glycine, glyceric acid, aminomalonic acid, and creatinine were identified as risk factors for LUAD (adjusted ORs>1.2, P<0.03). Remarkably, oxalate emerged as a distinctive metabolic risk factor exhibiting a strong correlation with the progression of LUAD (adjusted OR=5.107, P<0.001; advanced-stage vs. early-stage). The Random Forest (RF) model demonstrated a high degree of efficacy in distinguishing between LUAD and BLN (accuracy = 1.00 and 0.73, F1-score= 1.00 and 0.79, and AUC = 1.00 and 0.76 in the training and validation sets, respectively). TCGA and GTEx gene expression data have shown that lactate dehydrogenase A (LDHA), a crucial enzyme involved in oxalate metabolism, is increasingly expressed in the progression of LUAD. High LDHA expression levels in LUAD patients are also linked to poor prognoses (HR=1.66, 95% CI=1.34-2.07, P<0.001).

Conclusions: This study reveals risk factors associated with LUAD.

KEYWORDS

LUAD, metabolomics, risk factor, oxalate, LDHA

Abbreviations: LUAD, lung adenocarcinoma; BLN, benign lung nodules; HC, healthy controls; GC/MS, gas chromatography-mass spectrometry; OR, odds ratios; TCGA, the Cancer Genome Atlas; GTEx, the Genotype-Tissue Expression; ROC, the receiver operating characteristic (ROC) curve; AUC, the area under the ROC curve; LDHA, the lactate dehydrogenase A; QC, quality control; PCA, principal component analysis; PLS-DA, partial least square to latent structure discriminant analysis; RF, Random Forest; GNB, Gaussian Naive Bayes; MLP, Multilayer Perceptron; SVM, Support Vector Machine; KNN, k-Nearest Neighbors; DCA, decision curve analysis.

Introduction

Lung cancer is one of the most common and lethal types of malignant tumors worldwide. According to the statistics, there were approximately 2.2 million new cases of lung cancer and 1.8 million deaths worldwide in 2020 (1). Lung adenocarcinoma (LUAD) is the most common histological subtype of lung cancer, accounting for approximately 50% of all lung cancer cases (2). With the development of society, the incidence of various lung nodules, including lung adenocarcinoma, remains high due to environmental pollution, smoking, and unhealthy diet (3, 4). Although the clinical application of low-dose computed tomography (LDCT) has dramatically increased lung nodules' detection rate and reduced lung adenocarcinoma's mortality rate, the false-positive rate of lung adenocarcinoma detected by LDCT is high (5). Therefore, there is an unmet need for diagnosing LUAD and accurately classifying lung nodules.

Metabolic reprogramming has been recognized as one of the 10 hallmarks of cancer contributing to tumorigenesis and tumor progression (6–8). Studying the metabolic preferences, physiological dependencies, and molecular mechanisms that underlie LUAD is essential for its diagnosis, progression, and prognosis. For example, a large-scale metabolomic analysis of LUAD showed that combining serum metabolic fingerprints with protein tumor markers by deep learning can be used for early LUAD detection (9). Targeted metabolomic studies of resected lesions deciphered the metabolic trajectory from atypical adenomatous hyperplasia to adenocarcinoma *in situ* and invasive adenocarcinoma, revealing that metabolic perturbations occur in the precancerous lesions of LUAD (10). Extensive research has shown that glucose metabolic pathways, fatty acid metabolism, and glutamine metabolic pathways are associated with the prognosis of LUAD (11–13). Few studies have been performed on how LUAD progresses over time, so characterizing LUAD's metabolic evolution is necessary from different stages.

This study focused on detecting and screening metabolic markers associated with LUAD. Plasma metabolites from patients with lung nodules (including LUAD and BLN) and their controls were analyzed using a metabolomics platform with GC/MS instrumentation. Metabolic patterns were evaluated, and metabolic markers were screened and described based on semiquantitative data, receiver operating characteristic (ROC) curve analysis and odds ratio (OR). Then, endogenous metabolites and machine learning algorithms were used to construct and evaluate classification models for LUAD. In addition, we found that as the pTNM stages advanced and tumor metastasis, oxalate and its metabolic key enzyme LDHA changed in LUAD. Considering the interaction between oxalate and LDHA, the linkage and functions between the two are briefly explained in the discussion.

Materials and methods

Human plasma collection

The experimental samples were collected from the First Affiliated Hospital of Nanjing Medical University from June 2015 to June 2021. Blood samples were collected from 6:00 am to 8:30 am after overnight fasting and kept under 4°C before stored at –80°C within 6 hours after plasma isolation (14). Subjects included in this study were free of metabolic abnormalities such as hypoproteinemia, weight loss, and negative nitrogen balance. This prospective study was approved by the Ethics Committee of the First Affiliated Hospital of Nanjing Medical University (No. 2016-SRFA-149), and informed consent was obtained from all subjects.

Chemicals and reagents

1, 2-^{13}C2-Myristic acid, methyl myristate, methoxamine hydrochloride (purity 98%) and pyridine (≥99.8% GC) were purchased from Sigma-Aldrich (St. Louis, MO, USA). N-methyl-trimethylsilyltrifluoroacetamide (MSTFA) and 1% trimethylchlorosilane (TMCS) were provided by Pierce Chemical (Rockford, IL, USA). Methanol and *n*-heptane were HPLC grade and obtained from Merck (Darmstadt, Germany).

GC/MS analysis, instrumental setting, and parameters

Plasma samples were processed, extracted, and derived in accordance with our previously developed methods (15). 50.0 µL plasma was added into 200.0 µL methanol (containing 1, 2-$^{13}C_2$-Myristic acid, 5.0 µg/mL). The specimens were vigorously extracted for 5.0 min and centrifuged at 20000×g for 10.0 min at 4°C. A 100.0 µL aliquot of the resulting supernatant was transferred to a GC vial and evaporated to dryness in a Speed-Vac concentrator (Thermo Fisher Scientific, Savant^{TM} SC250EXP, Holbrook, USA). The dried plasma samples were then methoxymated, where 30.0 µL of methoxyamine pyridine solution (10.0 mg/mL) was added to the residue and incubated for 16 h at room temperature. Then the samples were trimethylsilylated for another 1.0 h by adding 30.0 µL of MSTFA with 1% TMCS as a catalyst. At last, 30.0 µL n-heptane with methyl myristate (15.0 µg/mL) was added to each GC vial as an external standard to monitor the stability of GC/MS.

A 0.5 µL sample aliquot was injected into gas chromatography coupled to a mass spectrometer (Shimadzu GCMS-QP2010 Ultra, Kyoto, Japan) in split mode (split ratio 8:1). It was equipped with an Rtx-5MS capillary column (0.25 mm × 30 m × 0.25µm, Restek, PA, USA). The injector temperature was set at 250°C. Helium was used as the carrier gas at a 1.5 mL/min flow rate. The column temperature was initially kept at 80°C for 3.0 min, then raised to 300°C at a rate of 20°C/min, and held for 5.0 min. The mass spectrometer ion source temperature and interface temperature were both 220°C, and ions were generated by a 70-eV electron beam at a current of 3.2 mA. The mass spectra were acquired over the mass range of 50-700 m/z in a full scan mode, with each run of 19.0 min. Following the abovementioned procedure, the quality control (QC) samples were prepared for the plasma pool. All samples were randomly selected for GC-MS analysis to diminish the systematic variations.

The metabolites were by comparing the mass spectrum and retention indexes for the analyte with the corresponding values from the literature and various libraries [e.g., Mainlib and Public in

the National Institute of Standards and Technology (NIST) library 2.0 (2008) and Wiley 9 (Wiley-VCH Verlag GmbH & Co KGaA, Weinheim, Germany)]. Some standard compounds were also utilized to identify the metabolites.

Statistical analysis

After normalization against the internal standard, all the semiquantitative data from GC/MS were $\log_{10}$-transformed. The transformed data were imported into SIMCA-P 14.1 software (Germany, Sartorius, Goettingen) and pre-processed for multivariate statistical analysis using par scaling. Principal component analysis (PCA) and partial least square to latent structure discriminant analysis (PLS-DA) models were built and plotted to show the clustering or separation of samples from different groups. PLS-DA models were constructed and plotted to show the clustering or separation of samples from different groups. The goodness of fit for the PLS-DA models was evaluated using three quantitative parameters: R^2X, R^2Y and Q^2. R^2X and R^2Y are the explained variations, and Q^2 is the predicted variation, with a higher level of R^2Y and Q^2Y indicating the model's better fit and predictive performance. To avoid the classification obtained by supervised learning methods being chance and to test whether the model reproduces well and whether the data in the model are over-fitted, the validity of the built model was examined by a 7-fold cross-check and replacement test (200 times, cross-validation). The intercept of the R^2 and Q^2 regression lines to the axes was used to measure overfitting, and the model was valid when the intercept of Q^2 was negative.

To assess the differences among the groups, analysis of variance (ANOVA) and multiple comparisons (LSD) were used. The independent-sample t-test and Mann-Whitney U test were used to analyze normally and non-normally distributed data. OR calculations and ROC curve analysis were performed using SPSS 26.0 (SPSS Inc., Chicago, IL, USA), bar graphs were produced using GraphPad Prism 8.0, and heatmap and pathway analysis were performed using the online software MetaboAnalyst (https://www.metaboanalyst.ca/).

Model development and evaluation

All individuals were randomly divided into train and validation datasets in a ratio of seven to three. Hyperparameters of seven machine learning models [including XGBoost, AdaBoost, Random Forest (RF), Gaussian Naive Bayes (GNB), Multilayer Perceptron (MLP), Support Vector Machine (SVM), and k-Nearest Neighbors (KNN)] were optimized using 10-fold cross-validation. Ten groups were created randomly from the training set. In each iteration of the 10-fold cross-validation method, nine groups were randomly selected for training and the remaining groups were used as test sets. Thus, the test sets for each group were selected sequentially, which ensured that the evaluation results did not overlap. Then, to minimize errors due to unreasonable test set selection, the results of the 10 evaluations were averaged.

The ROC curve analysis, calibration curve, decision curve analysis (DCA), accuracy, F1-score, sensitivity and specificity were used to assess the model's performance. Model discrimination was assessed with ROC analysis, and the accuracy of its predictions was assessed with AUC. The calibration curve showed the calibration and the extent to which the model's predictions deviated from actual events. Clinical utility and net benefit were assessed with DCA, which allows estimating the net benefit by calculating the difference between the true positive rate and the false positive rate, weighted by the odds ratio of the selected risk threshold probabilities.

All analyses were performed with R software (version 4.0) and Python version 3.7.

Transcriptomics database

RNA-sequencing expression (level 3) profiles and corresponding clinical information for lung adenocarcinoma were downloaded from the TCGA dataset (https://portal.gdc.com). The current-release (V8) GTEx datasets were obtained from the GTEx data portal website (https://www.gtexportal.org/home/datasets). Statistical analyses were performed using R software v4.0.3 (R Foundation for Statistical Computing, Vienna, Austria). P-value <0.05 was considered statistically significant. All analyses were performed using the online website HOME for Researchers (https://www.home-for-researchers.com/static/index.html#/).

Cell culture

A549 and BEAS-2B cell lines were purchased from the Type Culture Center, Chinese Academy of Sciences (Shanghai, China). These cell lines were grown in RPMI-1640 medium supplemented with 10% (v/v) fetal bovine serum and 100 U/mL penicillin and streptomycin at 37°C and 5% CO2. GC/MS and metabolomics analysis methods were the same as previously reported (16).

Results

Metabolic phenotypes of LUAD and the controls

A total of 1083 subjects were included in this study, including 670 lung adenocarcinoma (LUAD), 135 benign lung nodules (BLN) and 278 healthy controls (HC). Benign lung nodules mainly include pulmonary hamartomas, hemangiomas, and inflammatory pseudotumors. LUAD and BLN participants were newly diagnosed and had not undergone anti-cancer treatment, including radiation therapy, chemotherapy, surgical intervention, or medication administration. All patients included in this study were diagnosed by pathological examination. The distribution of subjects is shown in Figure 1A and Supplementary Table 1.

GC/MS analysis of the plasma samples aligned the metabolites in typical chromatograms (Supplementary Figure 1).

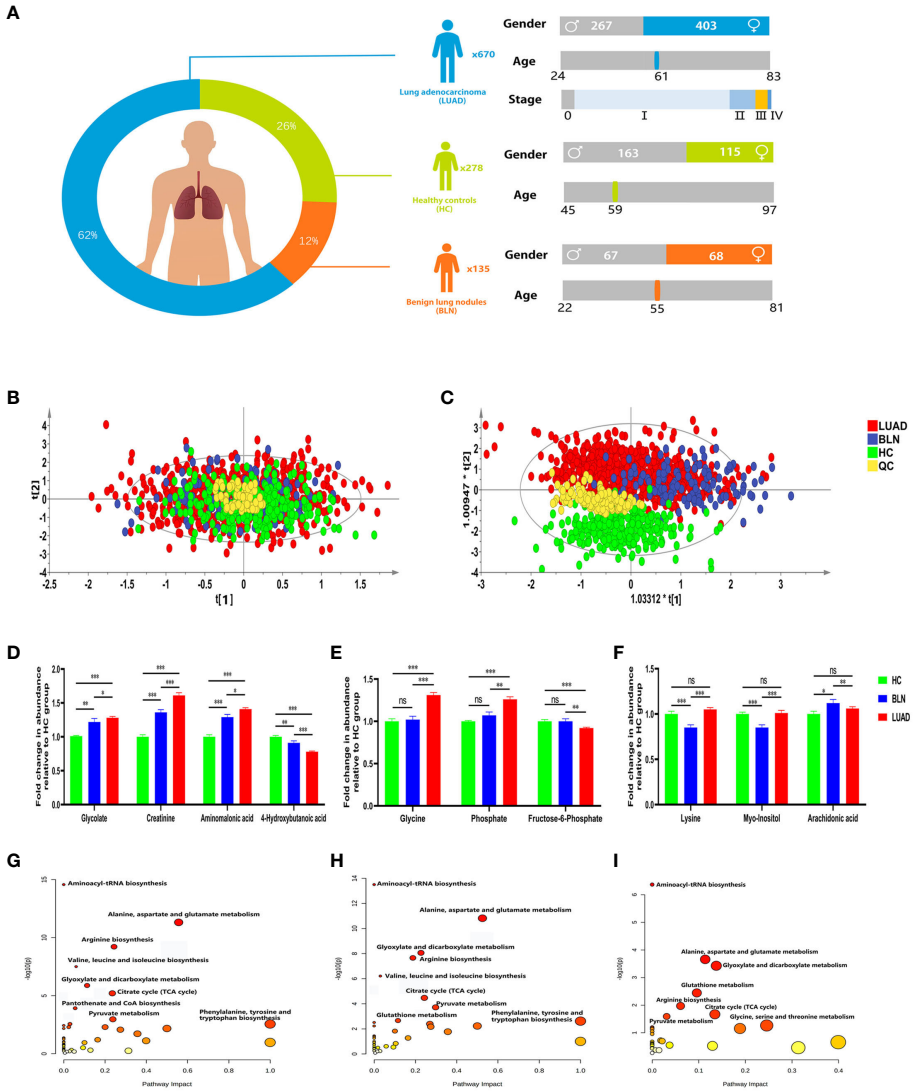


FIGURE 1
Distribution of subjects in this study and analysis of metabolic differences. **(A)** Distribution of subjects included in this study. **(B)** PCA score plot. **(C)** PLS-DA score plot. **(D)** Relative abundance of plasma glycolate, creatinine, aminomalonic acid and 4-hydroxybutanoic acid. **(E)** Relative abundance of plasma glycine, phosphate and fructose-6-phosphate. **(F)** Relative abundance of plasma lysine, myo-inositol and arachidonic acid (*, **and *** to denote $0.01 \leq p < 0.05$, $0.001 \leq p < 0.01$, and $p < 0.001$, respectively). **(G)** Metabolic pathway analysis of identified differential metabolites in BLN and HC groups; **(H)** Metabolic pathway analysis of identified differential metabolites in LUAD and HC groups; **(I)** Metabolic pathway analysis of identified differential metabolites in LUAD and BLN groups.

Deconvolution of the GC/MS chromatograms produced 103 independent peaks from the plasma samples, 54 of which were authentically identified as metabolites (Supplementary Table 2). Quantitative data were acquired for each metabolite in the plasma samples of the HC, BLN and LUAD cases.

The PCA and PLS-DA score plots (Figures 1B, C) demonstrated good clustering of pooled QC samples, indicating reproducibility of the assay and consistent instrument performance throughout the experiment. The supervised PLS-DA model showed that the samples in the HC and lung nodule groups (including LUAD and BLN) were distributed in different quadrants (Figure 1C), indicating significant metabolic differences between the two groups. Similarly, the overlap between LUAD and BLN indicates similar metabolic models. R^2X, R^2Y and Q^2 of the PLS-DA model were 0.628, 0.552 and 0.527, respectively, which suggested that the PLS-DA model had good adaptability and predictability. The permutation plot (Supplementary Figure 2) demonstrated that the PLS-DA models were valid: the Q^2 regression line had a negative intercept, and all of the permuted R^2 values to the left of the intercept were lower than the original point to the right. These results suggest that both variability and similarity in metabolic patterns are present in the three groups.

Metabolic features of LUAD

Based on statistical analysis (Table 1), 43 and 42 discriminant metabolites were found to differentiate LUAD and BLN patients

10.3389/fonc.2024.1277206

TABLE 1 List of discriminant metabolites: BLN vs. HC, LUAD vs. HC and LUAD vs. BLN.

Metabolites	HC (n=278)	BLN (n=135)	LUAD (n=670)	P1	P2	P3
Pyruvate	4.66 ± 0.16	4.58 ± 0.37	4.65 ± 0.42	<0.001	0.038	ns
Lactate	6.39 ± 0.21	6.32 ± 0.23	6.37 ± 0.36	<0.001	0.002	<0.001
Glycolate	3.86 ± 0.10	3.91 ± 0.17	3.93 ± 0.15	0.009	<0.001	0.020
Alanine	6.34 ± 0.20	6.26 ± 0.18	6.32 ± 0.17	<0.001	<0.001	0.004
Oxalate	5.38 ± 0.15	5.53 ± 0.17	5.42 ± 0.21	<0.001	<0.001	<0.001
3-Hydroxybutyric acid	4.82 ± 0.32	4.65 ± 0.33	4.67 ± 0.30	<0.001	<0.001	ns
Monomethylphosphate	4.53 ± 0.13	4.53 ± 0.15	4.54 ± 0.16	ns	0.017	ns
Valine	5.32 ± 0.15	5.25 ± 0.13	5.27 ± 0.14	<0.001	<0.001	ns
4-Hydroxybutanoic acid	4.84 ± 0.19	4.78 ± 0.18	4.71 ± 0.20	0.001	<0.001	<0.001
Phosphate	6.09 ± 0.06	6.09 ± 0.14	6.14 ± 0.18	ns	<0.001	0.004
Leucine	5.91 ± 0.15	5.86 ± 0.13	5.87 ± 0.13	<0.001	<0.001	ns
Isoleucine	5.51 ± 0.27	5.47 ± 0.21	5.47 ± 0.24	<0.001	<0.001	ns
Proline	5.63 ± 0.34	5.52 ± 0.33	5.59 ± 0.37	<0.001	0.032	0.012
Glycine	4.90 ± 0.20	4.92 ± 0.18	5.00 ± 0.22	ns	<0.001	<0.001
Succinate	3.80 ± 0.13	3.65 ± 0.18	3.64 ± 0.24	<0.001	<0.001	ns
Glyceric acid	3.92 ± 0.08	3.97 ± 0.17	4.00 ± 0.17	0.005	<0.001	ns
Fumarate	4.12 ± 0.17	4.07 ± 0.16	4.07 ± 0.24	0.007	<0.001	ns
Serine	5.52 ± 0.18	5.43 ± 0.14	5.47 ± 0.14	<0.001	<0.001	0.004
Threonine	5.15 ± 0.17	5.09 ± 0.12	5.11 ± 0.13	<0.001	<0.001	0.025
β-Alanine	4.97 ± 0.03	4.98 ± 0.03	4.97 ± 0.02	0.021	ns	0.017
Aminomalonic acid	5.42 ± 0.21	5.55 ± 0.17	5.59 ± 0.16	<0.001	<0.001	0.018
Malate	3.99 ± 0.10	3.80 ± 0.23	3.85 ± 0.25	<0.001	<0.001	0.018
Aspartate	4.36 ± 0.15	4.25 ± 0.22	4.29 ± 0.24	<0.001	<0.001	ns
Methionine	4.72 ± 0.14	4.65 ± 0.13	4.64 ± 0.14	<0.001	<0.001	ns
Pyroglutamate	5.63 ± 0.07	5.53 ± 0.24	5.61 ± 0.30	<0.001	<0.001	0.012
Cysteine	4.44 ± 0.10	4.42 ± 0.12	4.43 ± 0.12	0.036	ns	ns
Creatinine	4.45 ± 0.21	4.60 ± 0.17	4.67 ± 0.17	<0.001	<0.001	<0.001
Glutamate	5.29 ± 0.20	5.24 ± 0.23	5.26 ± 0.29	0.014	0.009	ns
Phenylalanine	5.22 ± 0.13	5.17 ± 0.11	5.17 ± 0.12	<0.001	<0.001	ns
Pyrophosphoric acid	5.10 ± 0.11	4.99 ± 0.13	5.03 ± 0.18	<0.001	<0.001	<0.001
Asparagine	4.41 ± 0.17	4.27 ± 0.15	4.32 ± 0.17	<0.001	<0.001	0.002
Glutamine	6.05 ± 0.13	5.91 ± 0.12	5.94 ± 0.13	<0.001	<0.001	0.003
Citrate	5.69 ± 0.10	5.58 ± 0.10	5.61 ± 0.10	<0.001	<0.001	0.008
Ornithine	5.00 ± 0.22	4.83 ± 0.25	4.95 ± 0.30	<0.001	0.006	<0.001
Glucose	5.89 ± 0.25	5.93 ± 0.55	5.84 ± 0.52	ns	0.040	ns
Lysine	4.53 ± 0.20	4.46 ± 0.18	4.54 ± 0.20	<0.001	ns	<0.001
Tyrosine	5.58 ± 0.13	5.52 ± 0.13	5.52 ± 0.12	<0.001	<0.001	ns
Palmitelaidic acid	4.08 ± 0.27	4.03 ± 0.31	4.04 ± 0.30	0.045	0.008	ns
Palmitic acid	5.71 ± 0.10	5.63 ± 0.12	5.66 ± 0.13	<0.001	<0.001	0.027

(Continued)

TABLE 1 Continued

Metabolites	HC (n=278)	BLN (n=135)	LUAD (n=670)	P1	P2	P3
Uric acid	5.80 ± 0.15	5.73 ± 0.20	5.72 ± 0.18	<0.001	<0.001	ns
Myo-Inositol	4.74 ± 0.12	4.67 ± 0.12	4.73 ± 0.15	<0.001	ns	<0.001
Linoleic acid	5.15 ± 0.13	5.01 ± 0.17	5.04 ± 0.18	<0.001	<0.001	ns
Oleic acid	5.57 ± 0.15	5.57 ± 0.14	5.56 ± 0.15	ns	0.046	ns
Cystine	3.93 ± 0.23	4.00 ± 0.30	3.96 ± 0.27	0.028	ns	ns
Fructose-6-Phosphate	3.88 ± 0.14	3.88 ± 0.14	3.84 ± 0.15	ns	<0.001	0.005
Arachidonic acid	4.23 ± 0.18	4.27 ± 0.19	4.23 ± 0.21	0.016	ns	0.003
Monopalmitin	4.32 ± 0.11	4.16 ± 0.11	4.21 ± 0.12	<0.001	<0.001	<0.001
Alpha-Tocopherol	5.12 ± 0.12	4.90 ± 0.26	4.89 ± 0.35	<0.001	<0.001	ns
Cholesterol	5.86 ± 0.06	5.84 ± 0.08	5.83 ± 0.14	0.013	0.003	ns

The data were log10 transformed and expressed as mean ± SD. P1, BLN vs. HC; P2, LUAD vs. HC; P3, LUAD vs. BLN; ns, P>0.05.

from healthy controls, respectively. Similarly, LUAD cases primarily showed different metabolomic patterns from BLN patients. According to the statistical analysis, 26 distinct metabolites were identified between LUAD and BLN.

Of the 42 metabolites differentiating BLN from HC, the levels of glycolate, creatinine and aminomalonic acid were higher in BLN, while 4-hydroxybutanoic acid was lower, and all the above metabolites deviated further in LUAD (Figure 1D). These findings indicate that the above metabolites are involved in the development of lung nodules (from averagely minimal damage to lung cancer). Although glycine, phosphate and fructose-6-phosphate showed significant differences between the LUAD and BLN groups, they had no significant difference between the BLN and HC groups (Figure 1E). It is, therefore, suggested that these markers are associated with LUAD. In addition, BLN cases showed deviations in lysine, myo-inositol, and arachidonic acid levels, whereas the LUAD and HC groups did not show any significant differences (Figure 1F), suggesting that they are markers associated with BLN.

Metabolic pathway analysis showed that patients with lung nodules were more affected by the following metabolic pathways: aminoacyl-tRNA biosynthesis, alanine, aspartate and glutamate metabolism, glyoxylate and dicarboxylate metabolism, citrate cycle, arginine biosynthesis (Figures 1G, H). The differential metabolic pathways between LUAD patients and BLN were: aminoacyl-tRNA biosynthesis, alanine, aspartate and glutamate metabolism, glyoxylate and dicarboxylate metabolism, glutathione metabolism, arginine biosynthesis (Figure 1I).

Potential risk factors for LUAD

The ROC analysis showed monopalmitin, succinate, glutamine, alpha-tocopherol, malate, asparagine and pyroglutamate performed well for HC and BLN differentiation (AUC ≥ 0.80) (Supplementary Table 3). Meanwhile, succinate and creatinine performed well for the differentiation of HC and LUAD (AUC ≥ 0.80) (Supplementary Table 4). However, each metabolite performed poorly in differentiating LUAD and BLN (AUC < 0.65) (Supplementary Table 5).

ORs values were calculated to assess the role of the above metabolites as risk factors for predicting BLN/LUAD occurrence (Table 2). Glycolate, oxalate, glyceric acid, β-alanine, aminomalonic acid, creatinine, cystine and arachidonic acid were found to be risk factors for BLN (ORs>1.0, P<0.05). In the meantime, glycolate, oxalate, glycine, glyceric acid, aminomalonic acid and creatinine were risk factors for LUAD (ORs>1.3, P<0.05). These results suggest that high levels of glycolate, oxalate, glyceric acid, aminomalonic acid and creatinine were independent risk factors of lung nodules. Whereas glycine may be a LUAD-specific risk factor, similarly, β-alanine, cystine and arachidonic acid may be BLN-specific risk factors.

Analyzing the differential metabolites between LUAD and BLN (Table 2), we found that alanine, phosphate, proline, glycine, serine, threonine, aminomalonic acid, malate, pyroglutamate, creatinine, pyrophosphoric acid, asparagine, glutamine, ornithine, lysine, palmitic acid, myo-inositol, monopalmitin 18 substances may be risk factors for the development of BLN to LUAD (ORs>1.1, P<0.05).

Oxalate characterizes the progression of LUAD

The occurrence and progression of lung adenocarcinoma are progressive processes over time. Identifying the metabolic changes during lung adenocarcinoma progression is highly important for diagnosing, treating, and prognosis. The pTNM staging is the most commonly used method of tumor staging, which mainly consists of three parts: T-primary tumor size, N-lymph node metastasis and M-distant metastasis. We set to stage 0, stage I and stage II as early-stage, stage III and stage IV as advanced-stage, considering the progression of LUAD and pTNM staging.

Statistical analysis showed that the metabolism in lung adenocarcinoma changes with tumorigenesis and progression (Table 3). Thirty-nine differential metabolites were found between early-stage LUAD and HC groups, and thirteen between advanced-stage LUAD and early-stage LUAD (Figure 2A). Notably, plasma asparagine, myo-inositol, ornithine, pyrophosphoric acid, threonine, and glutamine levels gradually decreased with increasing pTNM

10.3389/fonc.2024.1277206

TABLE 2 List of risk factors: BLN vs. HC, LUAD vs. HC and LUAD vs. BLN.

Cohorts	Risk factors	OR	95%CI	P-value	OR (adjusted)	95%CI (adjusted)	P-value (adjusted)
BLN vs. HC	Glycolate	2.283	[1.386, 3.761]	0.001	2.43	[1.464, 4.034]	0.001
	Oxalate	10.37	[5.911, 18.193]	<0.001	11.437	[6.356, 20.579]	<0.001
	Glyceric acid	3.081	[1.762, 5.390]	<0.001	3.139	[1.786, 5.518]	<0.001
	β-Alanine	13.991	[1.307, 149.738]	0.029	12.874	[1.171, 141.589]	0.037
	Aminomalonic acid	2.855	[1.991, 4.092]	<0.001	2.952	[2.033, 4.284]	<0.001
	Creatinine	3.659	[2.465, 5.430]	<0.001	3.861	[2.570, 5.802]	<0.001
	Cystine	1.045	[0.998, 1.095]	0.061	1.042	[0.994, 1.093]	0.09
	Arachidonic acid	1.533	[1.081, 2.174]	0.017	1.547	[1.083, 2.210]	0.016
LUAD vs. HC	Creatinine	2.892	[2.089, 4.003]	<0.001	3.408	[2.413, 4.813]	<0.001
	Oxalate	1.394	[1.107, 1.756]	0.005	1.287	[1.031, 1.607]	0.026
	Glycine	1.985	[1.610, 2.448]	<0.001	2.055	[1.656, 2.551]	<0.001
	Glyceric acid	3.211	[2.302, 4.479]	<0.001	3.204	[2.281, 4.502]	<0.001
	Aminomalonic acid	4.62	[3.511, 6.079]	<0.001	4.371	[3.307, 5.778]	<0.001
	Creatinine	6.319	[4.768, 8.373]	<0.001	6.473	[4.850, 8.639]	<0.001
LUAD vs. BLN	Alanine	1.771	[1.267, 2.475]	0.001	1.71	[1.222, 2.393]	0.002
	Phosphate	1.824	[1.222, 2.722]	0.003	1.935	[1.283, 2.919]	0.002
	Proline	1.158	[1.007, 1.331]	0.04	1.174	[1.018, 1.353]	0.027
	Glycine	1.768	[1.346, 2.322]	<0.001	1.812	[1.371, 2.395]	<0.001
	Serine	2.052	[1.339, 3.146]	0.001	2.337	[1.510, 3.616]	<0.001
	Threonine	1.597	[1.031, 2.474]	0.036	1.858	[1.182, 2.922]	0.007
	Aminomalonic acid	1.535	[1.094, 2.152]	0.013	1.47	[1.040, 2.080]	0.029
	Malate	1.316	[1.034, 1.676]	0.026	1.296	[1.019, 1.647]	0.034
	Pyroglutamate	1.356	[1.077, 1.708]	0.01	1.406	[1.113, 1.777]	0.004
	Creatinine	1.956	[1.412, 2.710]	<0.001	1.993	[1.425, 2.787]	<0.001
	Pyrophosphoric acid	1.35	[1.006, 1.811]	0.045	1.474	[1.091, 1.993]	0.012
	Asparagine	1.773	[1.240, 2.537]	0.002	2.033	[1.401, 2.951]	<0.001
	Glutamine	1.935	[1.246, 3.007]	0.003	2.193	[1.393, 3.454]	0.001
	Ornithine	1.596	[1.299, 1.960]	<0.001	1.64	[1.326, 2.028]	<0.001
	Lysine	1.857	[1.382, 2.494]	<0.001	1.982	[1.463, 2.686]	<0.001
	Palmitic acid	1.752	[1.119, 2.742]	0.014	1.69	[1.069, 2.672]	0.025
	Myo-Inositol	2.871	[1.802, 4.573]	<0.001	2.429	[1.495, 3.945]	<0.001
	Monopalmitin	2.803	[1.781, 4.411]	<0.001	2.954	[1.860, 4.693]	<0.001

All ORs were adjusted for age and gender.

TABLE 3 List of discriminant metabolites: early-stage vs. HC and advanced-stage vs. early-stage.

Metabolites	HC (n=278)	Early-stage (n=475)	Advanced-stage (n=39)	P1	P2
Glycolate	3.86 ± 0.10	3.92 ± 0.15	3.84 ± 0.17	<0.001	0.007
Oxalate	5.38 ± 0.15	5.45 ± 0.21	5.56 ± 0.14	<0.001	<0.001
Glycine	4.90 ± 0.20	4.97 ± 0.20	4.85 ± 0.21	<0.001	<0.001

(Continued)

TABLE 3 Continued

Metabolites	HC (n=278)	Early-stage (n=475)	Advanced-stage (n=39)	P1	P2
Fumarate	4.12 ± 0.17	4.04 ± 0.21	4.10 ± 0.18	<0.001	0.017
Threonine	5.15 ± 0.17	5.11 ± 0.13	5.07 ± 0.11	<0.001	0.042
Aminomalonic acid	5.42 ± 0.21	5.58 ± 0.16	5.51 ± 0.16	<0.001	0.006
Pyrophosphoric acid	5.10 ± 0.11	5.03 ± 0.20	4.92 ± 0.17	<0.001	<0.001
Asparagine	4.41 ± 0.17	4.30 ± 0.16	4.25 ± 0.17	<0.001	0.026
Glutamine	6.05 ± 0.13	5.94 ± 0.13	5.86 ± 0.13	<0.001	<0.001
Ornithine	5.00 ± 0.22	4.92 ± 0.28	4.80 ± 0.26	<0.001	0.014
Myo-Inositol	4.74 ± 0.12	4.71 ± 0.12	4.66 ± 0.18	0.011	0.005
β-Alanine	4.97 ± 0.03	4.97 ± 0.02	4.96 ± 0.02	ns	0.003
Arachidonic acid	4.23 ± 0.18	4.24 ± 0.22	4.36 ± 0.24	ns	0.004
Pyruvate	4.66 ± 0.16	4.60 ± 0.40	4.67 ± 0.37	<0.001	ns
Lactate	6.39 ± 0.21	6.33 ± 0.39	6.35 ± 0.21	<0.001	ns
Alanine	6.34 ± 0.20	6.30 ± 0.15	6.28 ± 0.15	<0.001	ns
3-Hydroxybutyric acid	4.82 ± 0.32	4.67 ± 0.31	4.71 ± 0.32	<0.001	ns
Valine	5.32 ± 0.15	5.25 ± 0.15	5.23 ± 0.10	<0.001	ns
4-Hydroxybutanoic acid	4.84 ± 0.19	4.71 ± 0.21	4.76 ± 0.27	<0.001	ns
Leucine	5.91 ± 0.15	5.86 ± 0.12	5.84 ± 0.11	<0.001	ns
Isoleucine	5.51 ± 0.27	5.47 ± 0.21	5.32 ± 0.47	<0.001	ns
Proline	5.63 ± 0.34	5.58 ± 0.35	5.42 ± 0.53	0.005	ns
Succinate	3.80 ± 0.13	3.61 ± 0.21	3.64 ± 0.28	<0.001	ns
Glyceric acid	3.92 ± 0.08	4.00 ± 0.16	4.05 ± 0.15	<0.001	ns
Serine	5.52 ± 0.18	5.46 ± 0.13	5.43 ± 0.11	<0.001	ns
Malate	3.99 ± 0.10	3.82 ± 0.22	3.88 ± 0.28	<0.001	ns
Aspartate	4.36 ± 0.15	4.27 ± 0.23	4.24 ± 0.23	<0.001	ns
Methionine	4.72 ± 0.14	4.65 ± 0.12	4.63 ± 0.12	<0.001	ns
Pyroglutamate	5.63 ± 0.07	5.56 ± 0.24	5.59 ± 0.23	<0.001	ns
Creatinine	4.45 ± 0.21	4.65 ± 0.16	4.58 ± 0.20	<0.001	ns
Glutamate	5.29 ± 0.20	5.22 ± 0.28	5.27 ± 0.22	<0.001	ns
Phenylalanine	5.22 ± 0.13	5.16 ± 0.11	5.18 ± 0.11	<0.001	ns
Citrate	5.69 ± 0.10	5.62 ± 0.10	5.60 ± 0.11	<0.001	ns
Tyrosine	5.58 ± 0.13	5.51 ± 0.11	5.52 ± 0.11	<0.001	ns
Palmitelaidic acid	4.08 ± 0.27	4.02 ± 0.30	3.96 ± 0.35	<0.001	ns
Palmitic acid	5.71 ± 0.10	5.65 ± 0.13	5.66 ± 0.12	<0.001	ns
Uric acid	5.80 ± 0.15	5.71 ± 0.18	5.67 ± 0.22	<0.001	ns
Linoleic acid	5.15 ± 0.13	5.04 ± 0.18	5.07 ± 0.17	<0.001	ns
Fructose 6-phosphate	3.88 ± 0.14	3.85 ± 0.14	3.86 ± 0.12	0.009	ns
Monopalmitin	4.32 ± 0.11	4.19 ± 0.11	4.21 ± 0.11	<0.001	ns
Alpha-Tocopherol	5.12 ± 0.12	4.89 ± 0.37	4.87 ± 0.35	<0.001	ns

The data were log10 transformed and expressed as mean ± SD. P1, Early-stage vs. HC; P2, Advanced-stage vs. Early-stage; ns, P>0.05.

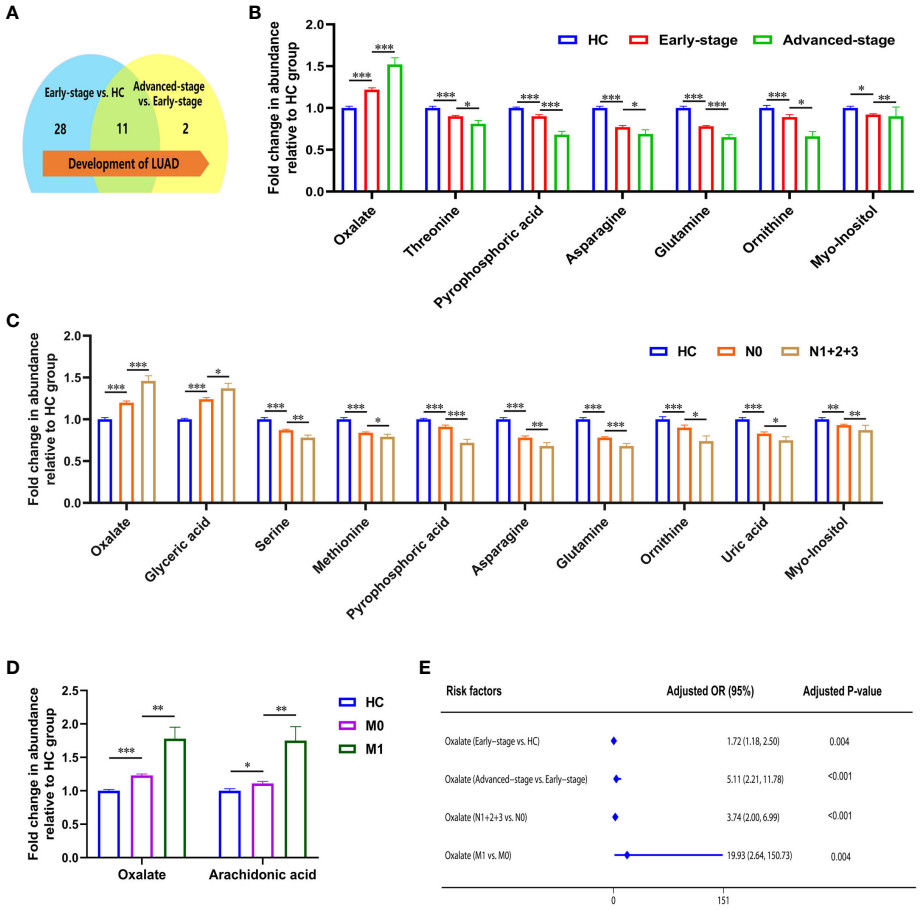


FIGURE 2
Metabolic features associated with the progression of lung adenocarcinoma. **(A)** Changes in the number of differential metabolites during lung adenocarcinoma development. **(B)** Relative abundance of asparagine, myo-inositol, ornithine, pyrophosphoric acid, threonine, and glutamine decreased, and oxalate increased in HC, early-stage, and advanced-stage patients. **(C)** Relative abundance of oxalate, glyceric acid, nonanoic acid, and arachidonic acid were elevated in patients with lymphatic metastasis, and levels of phosphate, proline, glycine, serine, methionine, creatinine, pyrophosphoric acid, asparagine, glutamine, ornithine, lysine, uric acid, and myo-inositol were decreased (N1 + 2+3, patients with lymphatic metastasis; N0, patients without lymphatic metastasis). **(D)** Relative abundance of oxalate and arachidonic acid in patients with distant metastatic (M1, patients with distant metastatic; M0, patients without distant metastatic). *, **and *** to denote $0.01 \leq p < 0.05$, $0.001 \leq p < 0.01$, and $p < 0.001$, respectively. **(E)** Odds ratio values of oxalate.

staging, while oxalate was elevated (Figure 2B). OR values suggested that oxalate may be a risk factor for progressive exacerbation of LUAD disease (Supplementary Table 6). These findings suggest that the above metabolites may be implicated in the development and progression of LUAD (from early to the advanced stage).

The correlation between metabolites and tumor size was analyzed with Spearman. Ten metabolites (alanine, valine, phosphate, leucine, asparagine, glutamine, ornithine, lysine, palmitic acid, and linoleic acid) were negatively correlated with tumor size (Supplementary Figure 3). With the appearance of lymphatic metastases, plasma levels of oxalate, glyceric acid, nonanoic acid, and arachidonic acid increased in LUAD, and phosphate, proline, glycine, serine, methionine, creatinine, pyrophosphoric acid, asparagine, glutamine, ornithine, lysine, uric acid, and myo-inositol were decreased (Figure 2C; Supplementary Table 7). In addition, high plasma levels of oxalate and glyceric acid may be risk factors for lymphatic metastasis in lung adenocarcinoma (Supplementary Table 7). Significant differences were observed in the glycolate, oxalate, glycine, and arachidonic acid levels between LUAD with distant metastasis and non-distant metastases (Supplementary Table 8). High levels of oxalate and arachidonic acid may be prognostic biomarkers for distant metastasis in lung adenocarcinoma (Figure 2D; Supplementary Table 8).

In summarizing the changes in metabolites during lung adenocarcinoma progression, it is clear that patients' metabolic profiles are changing. As the pTNM stages advanced and tumor metastasis, plasma oxalate levels increased, indicating that oxalate is closely associated with lung adenocarcinoma progression. Notably, OR values suggested that plasma oxalate may be a risk marker of the progression in LUAD (Figure 2E).

Development and evaluation of machine learning predictive models

To improve the differential diagnostic performance between HC, BLN and LUAD, we developed AI-based prediction models using seven machine learning algorithms (including XGBoost, RF,

AdaBoost, MLP, SVM, KNN and GNB). Glycolate, oxalate, glyceric acid, aminomalonic acid and creatinine were used to build predictive models for lung nodules and healthy controls. The results are shown in Figure 3; Supplementary Table 9. Compared with others, the RF model performs best in both the training and validation set (accuracy =1.00 and 0.85; F1-score =1.00 and 0.87; AUC=1.00 and 0.89, respectively). The calibration curves showed high consistency between the validation cohorts' predicted and observed survival probability. The DCA results showed that the RF model had an excellent net benefit across the whole range of threshold probabilities. Meanwhile, the RF models also showed excellent differentiation effects for distinguishing BLN or LUAD from HC, respectively (Supplementary Figures 4, 5). These results suggested that the RF model has the best diagnostic accuracy and applicability for distinguishing lung nodules from healthy controls.

4-hydroxybutanoic acid, monopalmitin, myo-inositol, β-alanine, oxalate, alanine, fructose-6-phosphate, glycolate, phosphate, and aminomalonic acid were screened out to construct prediction models for LUAD and BLN (Supplementary Figure 6). The results are shown in Figure 4; Supplementary Table 10. Compared with others, the RF model performs best in both the training and validation set (accuracy =1.00 and 0.73; F1-score =1.00 and 0.79; AUC=1.00 and 0.76, respectively). The calibration curves showed high consistency between the validation cohorts' predicted and observed survival probability. The DCA results showed that the RF model had an excellent net benefit across the whole range of threshold probabilities. These results suggested that the RF model can distinguish LUAD from BLN with high performance and accuracy.

RF modeling and plasma oxalate in an external validation set

A sample set was obtained from the Affiliated Hospital of Nanjing University Medical School in Nanjing, China for the purpose of external validation. This sample set comprised of 33 healthy controls, 23 patients with benign lung nodules, and 77 patients diagnosed with LUAD. The RF model previously developed was assessed on this distinct sample set, with the outcomes presented in Figure 5. The model exhibited high accuracy in predicting healthy controls and benign lung nodules, achieving an AUC of 0.99 (Figure 5A). The calibration curve and decision curve analysis further validated the model's precision and substantial net benefit (Figures 5B, C). In patients with benign lung nodules and lung adenocarcinoma, the model demonstrated an AUC of 0.866,

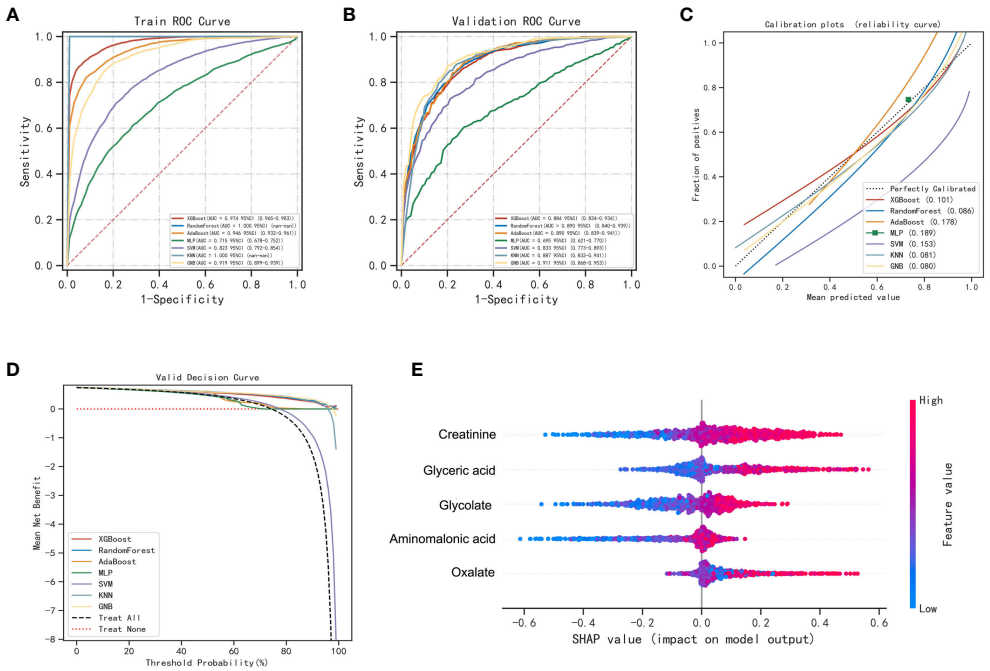


FIGURE 3
Performance of various machine learning models in pulmonary nodules and healthy controls. **(A)** The ROC curves of the training set. **(B)** The ROC curves of the validation set. **(C)** The calibration plots of the validation set. Each graph's 45° straight line perfectly matches the observed (y-axis) and predicted (x-axis) survival probabilities. A closer distance between two curves indicates greater accuracy. **(D)** Decision curve analysis graph showing the net benefit against threshold probabilities based on decisions from model outputs. **(E)** Shapley additive explanation (SHAP) summary plot of 5 feature clusters, derived by aggregating related values of a particular feature (e.g., the average, minimum, and maximum). Each dot corresponds to the SHAP value of the feature cluster for the lung cancer risk score of a given case patient or control subject at a certain point in time. A feature's SHAP value (x-axis) represents the contribution of the specific feature to the risk score, with positive values indicating a contribution that increases the risk score and negative values indicating a contribution that lowers the score. The location of the dot on the x-axis represents its SHAP value, whereas its color represents the cluster's value (the actual value of the feature that is represented in the cluster), with red representing higher values (for features measured along a continuum) or affirmative responses (for binary features). The dots are piled up vertically to show their density. The feature clusters are sorted by their mean absolute SHAP values.

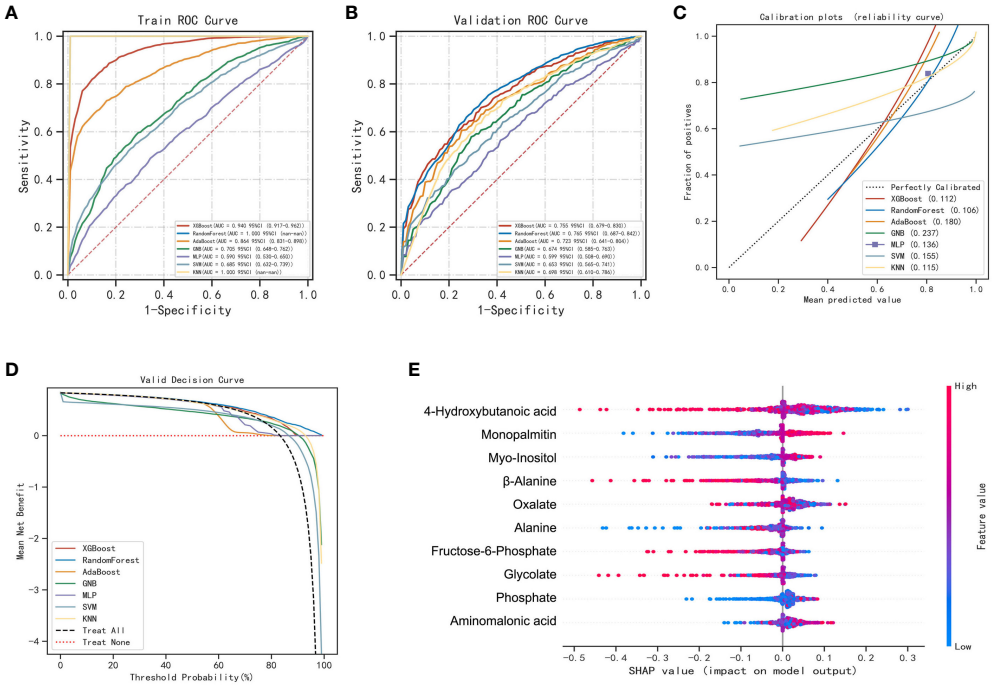


FIGURE 4
Performance of various machine learning models in LUAD and BLN groups. **(A)** The ROC curves of the training set. **(B)** The ROC curves of the validation set. **(C)** The calibration plots of the validation set. **(D)** Decision curve analysis graph showing the net benefit against threshold probabilities based on decisions from model outputs. **(E)** SHAP summary plot of 10 feature clusters.

indicating its relevance to these specific subgroups (Figures 5D, F). A noteworthy finding of the study was the correlation between elevated plasma oxalate levels and advanced pTNM stage and tumor metastasis in LUAD patients, as evidenced by an independent external validation set, underscoring the significant association between oxalate and LUAD progression (Figures 5G–I).

Alterations in pivotal enzymes involved in the metabolic pathway of oxalate

The primary sources of oxalate in humans are exogenous dietary intake (20-50%) and endogenous synthesis (50-80%) (17). The subjects collected in this study were mainly from Jiangsu Province, China, and there was slight dietary variation between subjects. Although ascorbic acid has been shown to have an essential effect on oxalate production, the process proceeds mainly through a non-enzymatic reaction, and the mechanism of action is unknown (18). In addition, we have ruled out the possibility of additional ascorbic acid supplementation in LUAD patients during treatment.

Glyoxalate is the primary precursor of endogenous oxalate synthesis (19). TCGA and GTEx database results showed significant differences in the expression of enzymes and transporters related to oxalate metabolism (LDHA, GRHPR, AGT, DAO, HAO2 and SLC26A1) in tumor tissues of lung adenocarcinoma patients (Figures 6A, B). Univariate and multivariate Cox regression analyses showed (Figure 6C) that LDHA, a key enzyme in the biosynthesis of oxalate, was significantly associated with the overall survival (OS) of LUAD patients and was a separately available prognostic indicator (HR = 1.66, 95% CI = 1.34 - 2.07, P < 0.001). A risk score was then calculated for each patient based on LDHA expression levels and risk factors, and patients were classified into low-risk and high-risk (Figure 6D). The heatmap revealed that high-risk patients tended to express LDHA genes at high levels, and low-risk patients tended to express LDHA genes at low levels (Figure 6D). Survival curves showed that patients with low-risk scores significantly had longer survival times than those with high-risk scores (Figure 6E). ROC analysis showed that LDHA could predict prognosis (Figure 6F).

The expression of LDHA in tumor tissues of lung adenocarcinoma increases with the advancement of the pTNM stage and lymphatic metastasis, suggesting that LDHA may be associated with abnormal immune function (Figures 7A–C). The European prospective investigation into cancer and nutrition (EPIC) array was used to analyze the correlation between immune cells and LDHA in the TCGA dataset. LDHA was negatively correlated with the expression of B cells, T cell CD8+, and endothelial cells (Figures 7D–F) and positively correlated with the expression of NK cells and uncharacterized cells (Figures 7G, H). In conclusion, LDHA is positively associated with the progression and prognosis of lung adenocarcinoma. The performance of LDHA further demonstrates the strong potential of oxalate as a risk factor of tumor progression in lung adenocarcinoma.

In order to assess the LDHA expression levels in LUAD, the protein expression of the LDHA gene was analyzed through

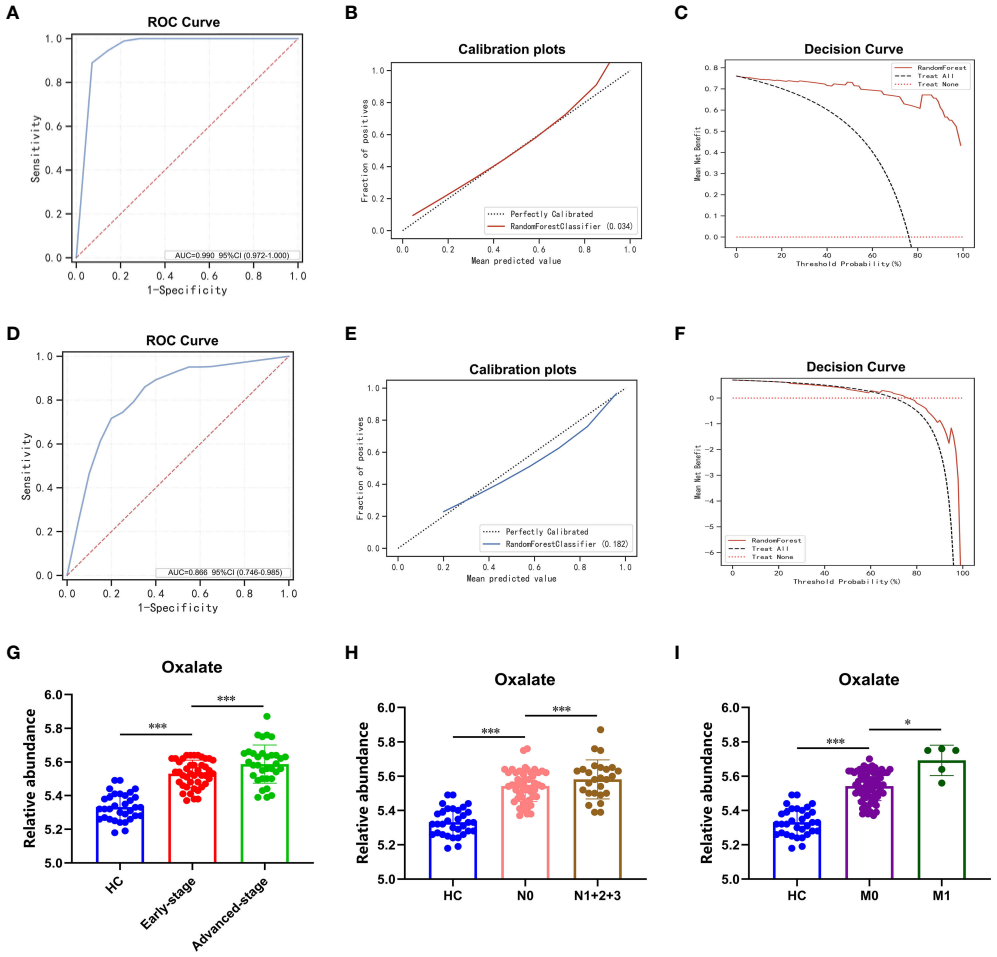


FIGURE 5
RF modeling and plasma oxalate in the external validation set. **(A)** The ROC curve of the external validation set comparing pulmonary nodules and healthy controls. **(B)** The calibration plots of the external validation set comparing pulmonary nodules and healthy controls. **(C)** The DCA plots of the external validation set comparing pulmonary nodules and healthy controls. **(D)** The ROC curve of the external validation set comparing LUAD and BLN. **(E)** The calibration plots of the external validation set comparing LUAD and BLN. **(F)** The DCA plots of the external validation set comparing LUAD and BLN. **(G)** Plasma oxalate increased in HC, early-stage, and advanced-stage LUAD. **(H)** Plasma oxalate was elevated in LUAD with lymphatic metastasis (N1 + 2+3, patients with lymphatic metastasis; N0, patients without lymphatic metastasis). **(I)** Plasma oxalate in LUAD with distant metastatic (M1, patients with distant metastatic; M0, patients without distant metastatic). * and *** to denote $0.01 \leq p < 0.05$ and $p < 0.001$, respectively.

immunohistochemistry, utilizing data from the Human Protein Atlas (HPA) database (https://www.proteinatlas.org/). The results indicated a significantly higher expression of LDHA in lung adenocarcinoma tissue compared to normal lung tissue (Figure 7I).

Oxalate and LDHA in lung adenocarcinoma cells

The relative content of oxalate and the expression of LDHA were further examined and analyzed in A549 and BEAS-2B cells. The results, depicted in Figures 7J, K, indicated a significantly higher oxalate content in A549 cells compared to BEAS-2B cells, accompanied by elevated LDHA expression. These findings suggest that increased LDHA and oxalate levels may represent a specific alteration in LUAD.

Discussion

Researchers have reported that genetic and microenvironmental factors drive clonal evolution within tumors, leading to metabolic liabilities and facilitating cancer progression (20, 21). Identifying the metabolic profile of lung adenocarcinoma is essential to understanding its progression and diagnosis. This study screened risk factors for LUAD by metabolomic analysis of plasma samples. We applied the machine learning algorithm to metabolite data to develop prediction models for lung nodules and LUAD. We confirmed that plasma oxalate is a risk factor associated with LUAD progression and further explored the interaction between oxalate and LDHA.

We analyzed plasma metabolites and confirmed the metabolic differences and similarities between lung adenocarcinoma and benign lung nodules. For metabolic differences, glycine and arachidonic acid may be LUAD and BLN-specific risk factors.

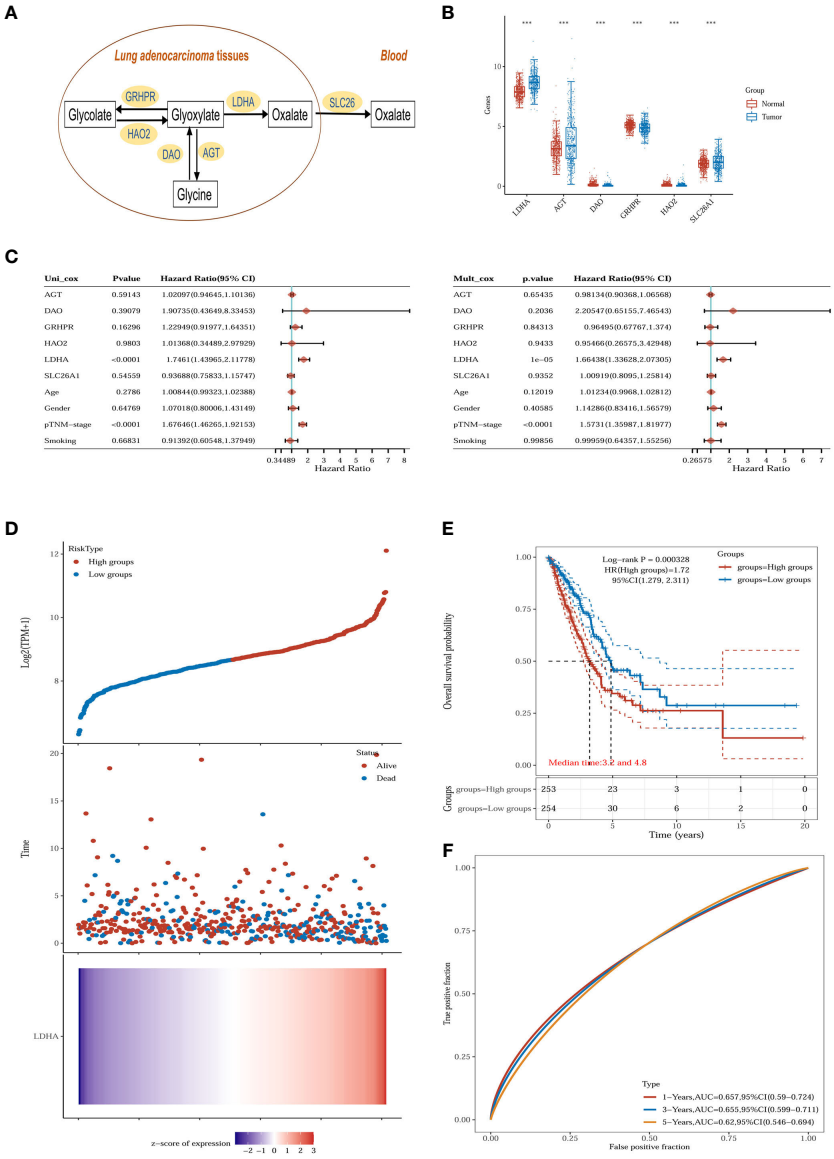


FIGURE 6
Oxalate metabolism in lung adenocarcinoma. **(A)** Oxalate-related metabolism in LUAD. **(B)** The oxalate-related enzymes [lactate dehydrogenase A (LDHA), glyoxylate reductase/hydroxypyruvate reductase (GRHPR), alanine–glyoxylate aminotransferase (AGT), d-amino acid oxidase (DAO), hydroxy acid oxidase 2 (HAO2) and solute carrier family 26 member 1 (SLC26A1)] are significantly changed in the tumor (n = 516) compared with normal tissue (n = 637). ***p < 0.001. **(C)** The p-value, risk coefficient (HR) and confidence interval are analyzed by univariate and multivariate Cox regression. **(D)** The gene expression, survival time and survival status of the TCGA dataset. The top scatterplot represents the gene expression from low to high. Different colors represent different groups. The scatter plot distribution represents the gene expression of different samples correspond to the survival time and survival status. The bottom Figure is the gene expression heatmap. **(E)** Kaplan-Meier survival analysis of the gene signature from the TCGA dataset, comparison among different groups was made by log-rank test. HR (High exp) represents the hazard ratio of the low-expression sample relatives to the high-expression sample. HR> 1 indicates the gene is a risk factor, and HR<1 indicates the gene is a protective factor.HR(95%CI), the median survival time (LT50) for different groups. **(F)** The ROC curve of the gene. The higher values of AUC correspond to higher predictive power.

The abnormal glycine metabolism has been widely reported in cancer (22, 23), so elevated plasma glycine in LUAD may be a common manifestation of cancer. Significantly, elevated plasma glycine may be an important phenomenon that distinguishes LUAD from benign nodules/healthy individuals. Surprisingly, arachidonic acid did not change significantly in the early-stage LUAD but was significantly elevated in BLN and advanced-stage LUAD (Tables 1, 3; Supplementary Tables 9, 10). Thus, plasma arachidonic acid may be important in distinguishing early-stage lung adenocarcinoma from benign lung nodules. This variable profile may be related to the degree of inflammatory response and abnormal immune response function. Based on the differences in metabolic patterns, we identified two panels of metabolites to distinguish LUAD from BLN/HC. Surprisingly, the machine learning models constructed with these two panels showed good prediction accuracy and diagnostic performance.

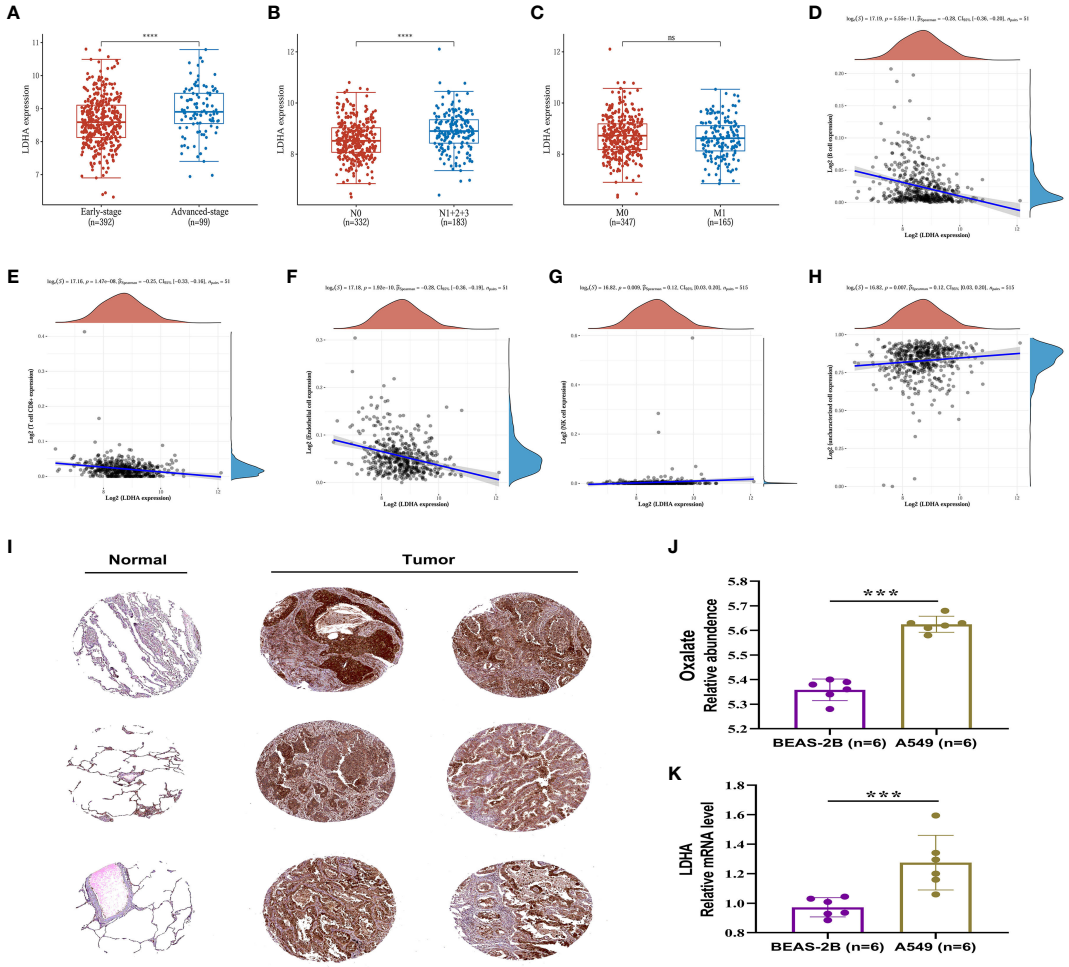


FIGURE 7
Expression and immune correlation analysis of LDHA in lung adenocarcinoma. **(A)** Expression distribution of LDHA gene in early-stage and advanced-stage LUAD tumor tissues. **(B)** Expression distribution of LDHA gene in LUAD tumor tissues of lymphatic metastasis and non-lymphoid metastasis groups. **(C)** Expression distribution of LDHA gene in LUAD tumor tissues of distant and non-distant metastasis groups. ****$p < 0.0001$, asterisks (*) stand for significance levels. **(D–H)** The correlations between LDHA gene expression and the immune score were analyzed with Spearman. The abscissa represents the distribution of the LDHA gene expression, and the ordinate represents the distribution of the immune score. The density curve on the right represents the trend in the distribution of the immune score, and the upper-density curve represents the trend in the distribution of the gene expression or the score. The value on the top represents the correlation p-value, correlation coefficient and correlation calculation method. **(I)** (IHC) The Immunohistochemistry (IHC) labelling images of normal lung tissue and tumor tissue were obtained from The Human Protein Atlas (http://www.proteinatlas.org/). **(J)** The relative content of oxalate in BESA-2B and A549 cells. **(K)** Expression of LDHA in BESA-2B and A549 cells. ***$p < 0.001$; 'ns' stands for non-significant, indicating that $P > 0.05$.

For metabolic similarity, we found that glycolate, oxalate, glyceric acid, aminomalonic acid and creatinine were elevated in both LUAD and BLN, showing potential as risk factors for lung nodules (with LUAD and BLN). Furthermore, a panel composed of these five metabolites showed excellent diagnostic performance for differentiating lung nodules from healthy controls. Interestingly, except for oxalate, the levels of the other four substances did not consistently increase with the exacerbation of LUAD. These results suggested that changes in glycolate, glyceric acid, aminomalonic acid, and creatinine may be co-occurring changes associated with the development of lung nodules. Machine learning models built from these metabolites could be used for the early screening of lung nodules.

With the improvement of technology and the availability of various kinds of big data, combining omics with machine learning algorithms has attracted increasing attention from the scientific community (24). Recently, machine learning algorithms (such as deep learning, random forest, and XGBoost) were applied to lung cancer research because of their advantages, which include high precision, robustness, simple operation, and fast response (9, 25–27). These machine learning methods have their merits and demerits, so choosing the best methods that best suit our research

is crucial. Seven commonly used machine learning algorithms were applied to this study. We found that the random forest model performed best by evaluating various metrics such as prediction accuracy, AUCs, positive predictive values, and negative predictive values. Machine learning models are considered 'black boxes,' which can be considered a limitation of the study. In this study, calibration curves, DCA analysis curves, and SHAP values were constructed to maximize model credibility and transparency. Generally, we were very rigorous in model screening, evaluation, and validation. Finally, we constructed an early screening model for lung nodules and a diagnostic model for lung adenocarcinoma using machine learning algorithms.

The present study is the first to identify oxalate as a risk factor closely associated with lung adenocarcinoma progression. Combining plasma metabolites and gene expression results from the TCGA and GTEx databases, and we found that elevated oxalate may be associated with LDHA overexpression. Many studies have documented elevated oxalate in tumor tissues and blood of lung cancer, and inhibition of LDHA expression can effectively reduce oxalate production (28–31). Therefore, it can be concluded that the overexpression of LDHA may cause an increase in plasma oxalate in lung adenocarcinoma patients, and inhibition of LDHA may reduce oxalate production.

LDHA is a promising therapeutic target for various malignancies, which plays important roles in tumorigenesis, progression, invasion and metastasis (32). Surprisingly, although LDHA promotes oxalate production, oxalate is a competitive inhibitor of LDHA (33). Several studies have shown that oxalate has anticancer effects on various cancer cell lines, including liver, breast, colorectal, lymphoma, medulloblastoma and ovarian cancer (33–38). In addition, oxalate has been shown to inhibit cell proliferation and migration and promote oxidative phosphorylation and epithelial-to-mesenchymal transition (35, 39). However, the oxalate concentrations required for significant therapeutic effects are too high for clinical use (40, 41). Because oxalate is an ionized conjugate base that readily complexes with divalent cations such as Mg^{2+} and Ca^{2+} (42). Hence, it could be hypothesized that free oxalate may have a potential protective effect and oxalate elevation may be a negative feedback-like 'protective response' against LDHA overexpression. However, high oxalate concentrations can lead to the formation of oxalate crystals, limiting its inhibitory effect on LDHA. Therefore, increasing the concentration of free oxalate and reducing the formation of oxalate crystals may be a new idea to inhibit LDHA expression and improve the prognosis of LUAD.

Limitations

This study analyzed the metabolic profile of patients with lung adenocarcinoma, but some limitations remain. The inclusion of solely Jiangsu Province, China subjects has resulted in a relatively small sample size, and it is possible that these results will not be reproducible on a wide scale across regions and populations. The mechanism of action between oxalate and LDHA was superficially investigated in this study, and further studies are needed to follow.

Conclusion

We analyzed the plasma metabolic profile of lung adenocarcinoma and provided an understanding of the progress of LUAD. This study has important implications for further translating basic research into more accurate diagnosis and treatment for clinical purposes.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: dbGAP database ID at TCGA (phs000178.v11.p8) and GTEx (phs000424.v9.p2).

Ethics statement

The studies involving humans were approved by the Ethics Committee of the First Affiliated Hospital of Nanjing Medical University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article. This prospective study was approved by the Ethics Committee of the First Affiliated Hospital of Nanjing Medical University (No. 2016-SRFA-149), and informed consent was obtained from all subjects.

Author contributions

MY: Writing – original draft, Data curation. WW: Writing – review & editing, Validation, Project administration. YW: Writing – review & editing, Methodology, Data curation. XS: Writing – review & editing, Investigation, Conceptualization. XY: Writing – review & editing, Software. WZ: Writing – review & editing, Resources, Project administration. JA: Writing – review & editing, Visualization, Resources, Funding acquisition. GW: Writing – review & editing, Visualization, Resources, Funding acquisition.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was financially supported by the leading technology foundation research project of Jiangsu province (BK20192005); the National Natural Science Foundation of the People's Republic of China (82173890); the CAMS Innovation Fund for Medical Sciences (CIFMS2021-I2M-5-011); the Fundamental Research Funds for the Central Universities (3012300076); the Postdoctoral Foundation of Jiangsu Province (1412300075).

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.2024.1277206/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. Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong KK. Non-small-cell lung cancers: a heterogeneous set of diseases. *Nat Rev Cancer*. (2014) 14:535–46. doi: 10.1038/nrc3775

3. Bode AM, Dong Z, Wang H. Cancer prevention and control: alarming challenges in China. *Natl Sci Rev*. (2016) 3:117–27. doi: 10.1093/nsr/nwv054

4. Ott S, Geiser T. [Epidemiology of lung tumors]. *Ther Umsch*. (2012) 69:381–8. doi: 10.1024/0040-5930/a000303

5. National Lung Screening Trial Research T, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. (2011) 365:395–409. doi: 10.1056/NEJMoa1102873

6. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. (2011) 144:646–74. doi: 10.1016/j.cell.2011.02.013

7. Keibler MA, Wasylenko TM, Kelleher JK, Iliopoulos O, Vander Heiden MG, Stephanopoulos G. Metabolic requirements for cancer cell proliferation. *Cancer Metab*. (2016) 4:16. doi: 10.1186/s40170-016-0156-6

8. Faubert B, Solmonson A, DeBerardinis RJ. Metabolic reprogramming and cancer progression. *Science*. (2020) 368:eaaw5473. doi: 10.1126/science.aaw5473

9. Wang L, Zhang M, Pan X, Zhao M, Huang L, Hu X, et al. Integrative serum metabolic fingerprints based multi-modal platforms for lung adenocarcinoma early detection and pulmonary nodule classification. *Adv Sci (Weinh)*. (2022) 9:e2203786. doi: 10.1002/advs.202203786

10. Nie M, Yao K, Zhu X, Chen N, Xiao N, Wang Y, et al. Evolutionary metabolic landscape from preneoplasia to invasive lung adenocarcinoma. *Nat Commun*. (2021) 12:6479. doi: 10.1038/s41467-021-26685-y

11. Wang H, Cui J, Yu J, Huang J, Li M. Identification of fatty acid metabolism-related lncRNAs as biomarkers for clinical prognosis and immunotherapy response in patients with lung adenocarcinoma. *Front Genet*. (2022) 13:855940. doi: 10.3389/fgene.2022.855940

12. Liu C, Wang Y. Identification of two subtypes and prognostic characteristics of lung adenocarcinoma based on pentose phosphate metabolic pathway-related long non-coding RNAs. *Front Public Health*. (2022) 10:902445. doi: 10.3389/fpubh.2022.902445

13. Liu J, Shen H, Gu W, Zheng H, Wang Y, Ma G, et al. Prediction of prognosis, immunogenicity and efficacy of immunotherapy based on glutamine metabolism in lung adenocarcinoma. *Front Immunol*. (2022) 13:960738. doi: 10.3389/fimmu.2022.960738

14. Yin P, Peter A, Franken H, Zhao X, Neukamm SS, Rosenbaum L, et al. Preanalytical aspects and sample quality assessment in metabolomics studies of human blood. *Clin Chem*. (2013) 59:833–45. doi: 10.1373/clinchem.2012.199257

15. Yu M, Wen W, Yi X, Zhu W, Aa J, Wang G. Plasma metabolomics reveals diagnostic biomarkers and risk factors for esophageal squamous cell carcinoma. *Front Oncol*. (2022) 12:829350. doi: 10.3389/fonc.2022.829350

16. Cao B, Li M, Zha W, Zhao Q, Gu R, Liu L, et al. Metabolomic approach to evaluating adriamycin pharmacodynamics and resistance in breast cancer cells. *Metabolomics*. (2013) 9:960–73. doi: 10.1007/s11306-013-0517-x

17. Williams HE, Wandzilak TR. Oxalate synthesis, transport and the hyperoxaluric syndromes. *J Urol*. (1989) 141:742–9. doi: 10.1016/s0022-5347(17)40999-2

18. Knight J, Madduma-Liyanage K, Mobley JA, Assimos DG, Holmes RP. Ascorbic acid intake and oxalate synthesis. *Urolithiasis*. (2016) 44:289–97. doi: 10.1007/s00240-016-0868-7

19. Robertson DS. The function of oxalic acid in the human metabolism. *Clin Chem Lab Med*. (2011) 49:1405–12. doi: 10.1515/CCLM.2011.238

20. Yuneva MO, Fan TW, Allen TD, Higashi RM, Ferraris DV, Tsukamoto T, et al. The metabolic profile of tumors depends on both the responsible genetic lesion and tissue type. *Cell Metab*. (2012) 15:157–70. doi: 10.1016/j.cmet.2011.12.015

21. Hensley CT, Faubert B, Yuan Q, Lev-Cohain N, Jin E, Kim J, et al. Metabolic heterogeneity in human lung tumors. *Cell*. (2016) 164:681–94. doi: 10.1016/j.cell.2015.12.034

22. Amelio I, Cutruzzola F, Antonov A, Agostini M, Melino G. Serine and glycine metabolism in cancer. *Trends Biochem Sci*. (2014) 39:191–8. doi: 10.1016/j.tibs.2014.02.004

23. Tajan M, Hennequart M, Cheung EC, Zani F, Hock AK, Legrave N, et al. Serine synthesis pathway inhibition cooperates with dietary serine and glycine limitation for cancer therapy. *Nat Commun*. (2021) 12:366. doi: 10.1038/s41467-020-20223-y

24. Li R, Li L, Xu Y, Yang J. Machine learning meets omics: applications and perspectives. *Brief Bioinform*. (2022) 23:bbab460. doi: 10.1093/bib/bbab460

25. Liu QX, Zhou D, Han TC, Lu X, Hou B, Li MY, et al. A noninvasive multianalytical approach for lung cancer diagnosis of patients with pulmonary nodules. *Adv Sci (Weinh)*. (2021) 8:2100104. doi: 10.1002/advs.202100104

26. Gould MK, Huang BZ, Tammemagi MC, Kinar Y, Shiff R. Machine learning for early lung cancer identification using routine clinical and laboratory data. *Am J Respir Crit Care Med*. (2021) 204:445–53. doi: 10.1164/rccm.202007-2791OC

27. Huang L, Wang L, Hu X, Chen S, Tao Y, Su H, et al. Machine learning of serum metabolic patterns encodes early-stage lung adenocarcinoma. *Nat Commun*. (2020) 11:3556. doi: 10.1038/s41467-020-17347-6

28. Ermer T, Nazzal L, Tio MC, Waikar S, Aronson PS, Knauf F. Oxalate homeostasis. *Nat Rev Nephrol*. (2022) 19:123–38. doi: 10.1038/s41581-022-00643-3

29. Shee K, Ahn J, Hamouche F, Mena J, Chi T, Stoller ML. Nedosiran dramatically reduces serum oxalate in dialysis-dependent primary hyperoxaluria 1: A compassionate use case report. *Urology*. (2021) 156:e147–9. doi: 10.1016/j.urology.2021.03.014

30. Mu Y, Zhou Y, Wang Y, Li W, Zhou L, Lu X, et al. Serum metabolomics study of nonsmoking female patients with non-small cell lung cancer using gas chromatography-mass spectrometry. *J Proteome Res*. (2019) 18:2175–84. doi: 10.1021/acs.jproteome.9b00069

31. Zheng R, Fang X, Chen X, Huang Y, Xu G, He L, et al. Knockdown of lactate dehydrogenase by adeno-associated virus-delivered CRISPR/Cas9 system alleviates primary hyperoxaluria type 1. *Clin Transl Med*. (2020) 10:e261. doi: 10.1002/ctm2.261

32. Claps G, Faouzi S, Quidville V, Chehade F, Shen S, Vagner S, et al. The multiple roles of LDH in cancer. *Nat Rev Clin Oncol*. (2022) 19:749–62. doi: 10.1038/s41571-022-00686-2

33. Qiao T, Xiong Y, Feng Y, Guo W, Zhou Y, Zhao J, et al. Inhibition of LDH-A by oxamate enhances the efficacy of anti-PD-1 treatment in an NSCLC humanized mouse model. *Front Oncol*. (2021) 11:632364. doi: 10.3389/fonc.2021.632364

34. Manerba M, Di Ianni L, Fiume L, Roberti M, Recanatini M, Di Stefano G. Lactate dehydrogenase inhibitors sensitize lymphoma cells to cisplatin without enhancing the drug effects on immortalized normal lymphocytes. *Eur J Pharm Sci*. (2015) 74:95–102. doi: 10.1016/j.ejps.2015.04.022

35. Valvona CJ, Fillmore HL. Oxamate, but not selective targeting of LDH-A, inhibits medulloblastoma cell glycolysis, growth and motility. *Brain Sci*. (2018) 8:56. doi: 10.3390/brainsci8040056

36. Zhou M, Zhao Y, Ding Y, Liu H, Liu Z, Fodstad O, et al. Warburg effect in chemosensitivity: targeting lactate dehydrogenase-A re-sensitizes taxol-resistant cancer cells to taxol. *Mol Cancer*. (2010) 9:33. doi: 10.1186/1476-4598-9-33

37. Zhao Y, Liu H, Liu Z, Ding Y, Ledoux SP, Wilson GL, et al. Overcoming trastuzumab resistance in breast cancer by targeting dysregulated glucose metabolism. *Cancer Res*. (2011) 71:4585–97. doi: 10.1158/0008-5472.CAN-11-0127

38. Xiang J, Zhou L, He Y, Wu S. LDH-A inhibitors as remedies to enhance the anticancer effects of PARP inhibitors in ovarian cancer cells. *Aging (Albany NY)*. (2021) 13:25920–30. doi: 10.18632/aging.203780

39. Convento MB, Pessoa EA, Cruz E, da Gloria MA, Schor N, Borges FT. Calcium oxalate crystals and oxalate induce an epithelial-to-mesenchymal transition in the proximal tubular epithelial cells: Contribution to oxalate kidney injury. *Sci Rep*. (2017) 7:45740. doi: 10.1038/srep45740

40. Cheng CS, Tan HY, Wang N, Chen L, Meng Z, Chen Z, et al. Functional inhibition of lactate dehydrogenase suppresses pancreatic adenocarcinoma progression. *Clin Transl Med*. (2021) 11:e467. doi: 10.1002/ctm2.467

41. Lea MA, Guzman Y, Desbordes C. Inhibition of growth by combined treatment with inhibitors of lactate dehydrogenase and either phenformin or inhibitors of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 3. *Anticancer Res*. (2016) 36:1479–88. doi: 10.1158/1538-7445.AM2016-32

42. Ermer T, Eckardt KU, Aronson PS, Knauf F. Oxalate, inflammasome, and progression of kidney disease. *Curr Opin Nephrol Hypertens*. (2016) 25:363–71. doi: 10.1097/MNH.0000000000000229

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 20 March 2024
DOI 10.3389/fonc.2024.1305262

Check for updates

OPEN ACCESS

EDITED BY
Miao Liu,
Harvard Medical School, United States

REVIEWED BY
Cheng Huang,
Fudan University, China
Hem Chandra Jha,
Indian Institute of Technology Indore, India

*CORRESPONDENCE
Yi-li Fu
fuyili1225@163.com
Bin Hu
hubin807@126.com

†These authors have contributed equally to this work

RECEIVED 10 October 2023
ACCEPTED 23 February 2024
PUBLISHED 20 March 2024

CITATION
Wang J, Cui S-p, Zhao Q, Gao Y, Ji Y, Liu Y, Miao J-b, Fu Y-l and Hu B (2024) Preoperative systemic immune-inflammation index-based nomogram for lung carcinoma following microwave ablation -a real world single center study.
Front. Oncol. 14:1305262.
doi: 10.3389/fonc.2024.1305262

COPYRIGHT
© 2024 Wang, Cui, Zhao, Gao, Ji, Liu, Miao, Fu and Hu. 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.

Preoperative systemic immune-inflammation index-based nomogram for lung carcinoma following microwave ablation -a real world single center study

Jing Wang†, Song-ping Cui†, Qing Zhao†, Ya Gao, Ying Ji, Yi Liu, Jin-bai Miao, Yi-li Fu* and Bin Hu*

Department of Thoracic Surgery, Beijing Institute of Respiratory Medicine and Beijing Chao-Yang Hospital, Capital Medical University, Beijing, China

Background: The preoperative inflammatory condition significantly influences the prognosis of malignancies. We aimed to investigate the potential significance of preoperative inflammatory biomarkers in forecasting the long-term results of lung carcinoma after microwave ablation (MWA).

Method: This study included patients who received MWA treatment for lung carcinoma from Jan. 2012 to Dec. 2020. We collected demographic, clinical, laboratory, and outcome information. To assess the predictive capacity of inflammatory biomarkers, we utilized the area under the receiver operating characteristic curve (AUC-ROC) and assessed the predictive potential of inflammatory biomarkers in forecasting outcomes through both univariate and multivariate Cox proportional hazard analyses.

Results: A total of 354 individuals underwent MWA treatment, of which 265 cases were included in this study, whose average age was 69.1 ± 9.7 years. The AUC values for the Systemic Inflammatory Response Index (SIRI) to overall survival (OS) and disease-free survival (DFS) were 0.796 and 0.716, respectively. The Cox proportional hazards model demonstrated a significant independent association between a high SIRI and a decreased overall survival (hazard ratio [HR]=2.583, P<0.001). Furthermore, a high SIRI independently correlated with a lower DFS (HR=2.391, P<0.001). We developed nomograms utilizing various independent factors to forecast the extended prognosis of patients. These nomograms exhibited AUC of 0.900, 0.849, and 0.862 for predicting 1-year, 3-year, and 5-year OS, respectively. Additionally, the AUC values for predicting 1-year, 3-year, and 5-year DFS were 0.851, 0.873, and 0.883, respectively.

Conclusion: SIRI has shown promise as a valuable long-term prognostic indicator for forecasting the outcomes of lung carcinoma patients following MWA.

KEYWORDS

lung carcinoma, microwave ablation, biomarkers, systemic inflammatory response index, prognosis

1 Introduction

Incidence and mortality rates of pulmonary carcinoma account for 17.1% and 21.7% of all malignant tumors, respectively, making it a major public health concern that significantly threatens human well-being (1). The treatment of lung carcinoma, both primary and metastatic, has undergone substantial diversification over the past two decades. New therapeutic technologies offer a range of choices, from traditional open surgical resection and video-assisted thoracoscopic surgery (VATS) to more recent highly efficient radiation therapy methods, exemplified by stereotactic body radiation therapy (SBRT) and image-guided thermal ablation treatments such as radiofrequency ablation (RFA), microwave ablation (MWA), and cryoablation. Multitude of techniques and their constant improvement raise the question of which approach is most beneficial for optimal patient outcome. Although SBRT and thermal ablation treatments exhibit lower control rates compared to surgical resection in the overall population of lung carcinoma patients, their primary advantages lie in reducing invasiveness and their impact on respiratory function. Additionally, they demonstrate prognosis comparable to surgery in certain early-stage non-small cell lung carcinoma (NSCLC) (2, 3).

Hence, they offer a potential curative option for patients with early-stage NSCLC or limited metastasis for whom surgical intervention is medically infeasible. Over the past five years, thermal ablation, particularly MWA, has made significant strides in the treatment of lung tumors. MWA holds theoretical advantages over RFA, as it can generate a larger ablation zone in a shorter duration. Initially, this local treatment technique was primarily employed for lung carcinoma patients who were ineligible for surgery or unsuitable candidates for surgical intervention. However, with advancements in technology and the accumulation of clinical experience, its indications have expanded to include the treatment of early-stage multiple primary lung carcinomas (MPLC), especially in cases where tumors cannot be surgically resected through conventional means or when patients are averse to surgery (4). Multiple studies have indicated that MWA can achieve high rates of local control, particularly for smaller ground-glass opacities (GGO) or ground-glass nodules (GGN) (5). The use of next-generation radiofrequency ablation devices and advanced medical imaging techniques such as CT and electromagnetic navigation technology has enhanced the precision of MWA procedures, ensuring more accurate tumor localization and treatment (6).

Although some literature reports that for early-stage NSCLC patients, the survival rates after MWA treatment are comparable to traditional surgical approaches, with a lower incidence of complications (7), there is limited documentation regarding post-MWA recurrence and prognosis for the overall lung carcinoma patient population, therefore, effective biomarkers and accurate prediction of the risk of tumor recurrence after MWA for lung carcinoma and timely intervention are particularly important for improving the disease free survival (DFS) and overall survival (OS).

The deviation from homeostasis is resolved through an intensified immune activation process, often referred to as inflammation. The malignant transformation of cells represents such a non-physiological state, demanding a robust host response. The current consensus suggests a close association between chronic inflammation and the development of tumors. Thus far, it has been postulated that many cancers are either caused by chronic inflammation or induced by inflammatory reactions, making inflammation a characteristic feature of cancer (8). MWA releases a microwave thermal field that induces rapid rotational motion and friction of surrounding molecules, raising temperatures to promote coagulative necrosis in both the tumor and adjacent tissues. Simultaneously, the release of numerous cellular debris and inflammatory factors resulting from necrosis further stimulates the occurrence of tumor immune responses (9). In the meantime, inflammatory and immune responses related to tumors are recognized as crucial factors in the initiation, progression, angiogenesis, and metastasis of tumors (10, 11). These include several inflammatory and immune scores, such as neutrophil lymphocyte ratio (NLR), platelet lymphocyte ratio (PLR) and monocyte-to-lymphocyte ratio (MLR) (12). A recently introduced systemic inflammatory response index (SIRI), calculated from absolute lymphocyte, monocyte, and neutrophil counts in peripheral blood, has emerged as a promising prognostic tool for various cancers, including NSCLC (13, 14). Numerous

investigations have provided evidence that increased SIRI values are linked to unfavorable outcomes among NSCLC patients following surgical resection (15, 16). However, the threshold for SIRI values vary widely across the studies (17, 18) and the prognostic value of preoperative systemic immune-inflammation index in patients with early-stage pulmonary cancer following MWA remains unclear.

This real-world study was conducted to investigate the prognostic value of SIRI in patients with lung carcinoma after MWA treatment. Novel preoperative SIRI-based nomograms were developed to predict the probability of MWA in these patients to help screen high-risk patients and formulating adjuvant therapeutic and individualized strategies.

2 Materials and methods

2.1 Study population

This study included patients who received MWA treatment for lung carcinoma at our hospital from January 2012 to December 2020.

Inclusion Criteria: (1) patients >18 years old; (2) the pathological results confirm NSCLC; (3) not suitable for curative surgery or patient refuses curative surgery; (4) no signs of metastasis; (5) ECOG score of 0-1; (6) no preoperative neoadjuvant therapy; (7) no history of radiation therapy; (8) informed consent from patients and their families regarding the surgical procedure and research protocol.

Exclusion criteria: (1) consideration of metastatic lung carcinoma or lymph node metastasis; (2) concurrent active bacterial or fungal infections; (3) deemed unsuitable for inclusion by the research team after evaluation; (4) incomplete follow-up information.

2.2 Data collection and follow-up

We collected the following data using an electronic medical record system: patient baseline information (age, gender, smoking history, and family history, etc.), laboratory test results (complete blood count, etc.), surgical-related data (surgery time, MWA power, intraoperative bleeding volume, chest drainage volume, pneumothorax, etc.), complications (pneumothorax, hemothorax, pulmonary hemorrhage) and pathological results (pathological type, degree of differentiation, etc.). All patients who were enrolled had venous blood samples collected within 24 hours of admission and underwent a complete blood count analysis. NLR was defined as the ratio of neutrophils to lymphocytes, while PLR was defined as the ratio of platelets to lymphocytes. SII was defined as the neutrophil counts multiplied by the platelet counts, then divided by the lymphocyte counts. SIRI was defined as the monocyte counts multiplied by the neutrophil counts divided by the lymphocyte counts.

We maintained continuous patient follow-up through telephone calls, outpatient visits, or inpatient observation until December 2022 or until the occurrence of death. Follow-up of discharged patients at 6 months, 1 year, and 2 years post-surgery to assess the occurrence of short-term and long-term complications, as well as their survival status, etc. The overall survival (OS) was defined as the time from MWA to death or the last observation, while disease-free survival (DFS) was defined as the time from MWA to recurrence.

2.3 MWA procedure

Computerized Tomography (CT) and magnetic navigation system were employed to guide the MWA procedure. This study is real-world research, and there are no specific eligibility criteria for patient inclusion. The choice between CT or magnetic navigation system-guided surgery can be made through communication between patients and physicians. MWA employs the MTC-3C microwave ablation system (developed by Zhongyuan Medical Device Research Center) with a frequency of 2450 ± 50 MHz and an adjustable continuous wave output power range from 0 to 100W. The effective length of the microwave antenna is 100-180mm, with an outer diameter of 14-20G. The effective ablation range of MWA is 3.5 cm * 3 cm, with an output of 60~80W/6~8 min. In this study, the high-power group is defined as having an output power of 80W, while the low-power group is defined as having an output power of 60W. The surgery is performed by the chief physician, who has extensive experience in tumor ablation to ensure consistency in the quality of the surgery. The radiological manifestation of a successful surgery is that the lesions after ablation is 5-10mm larger than the target lesion.

2.4 Statistical analysis

Continuous variables were compared between groups using either the Student's t-test or the Mann-Whitney U test. Categorical variables were assessed for group comparisons using either Pearson's chi-square test or Fisher's exact test. We utilized Kaplan-Meier curves in conjunction with the log-rank test for survival analysis. The receiver operating characteristics (ROC) curve was used to determine the optimal cut-off value. Cox regression models were used to identify independent prognostic factors associated with OS and DFS. Variables were considered for inclusion in the multivariate analysis if their $P<0.2$ in the univariate analysis. Utilizing independent risk factors identified in Cox analysis, we developed nomograms to predict the long-term prognosis of MWA patients. Statistical significance was defined as two-sided $P<0.05$. All statistical analyses were performed using GraphPad Prism (version 8.0), SPSS (version 26.0) and R software (version 4.3.1).

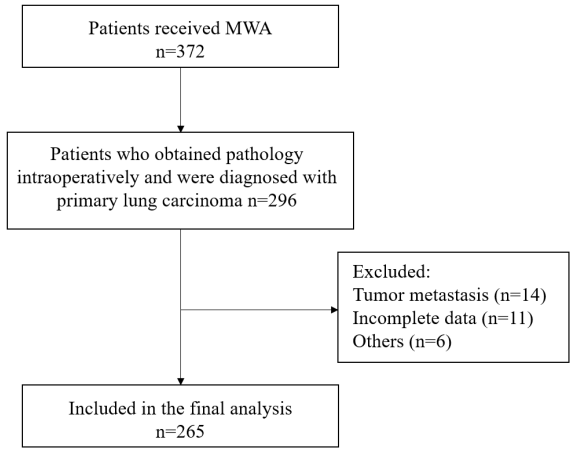


FIGURE 1
Flowchart depicting the composition of the study group. MWA, microwave ablation.

3 Results

3.1 Patient's basic information

This study included a total of 265 lung carcinoma patients who underwent MWA (Figure 1). The average age of all enrolled patients was 69.1 ± 9.7 years, with 177 males and 88 females. The median duration of follow-up was 43.8 months. The patient's comorbidity, tumor-related information, and laboratory test results were as shown in Table 1.

We applied the ROC curve to explore the correlation between preoperative inflammatory biomarkers and long-term prognosis in lung carcinoma patients after MWA. The results showed that SIRI had the best value for predicting long-term prognosis (Supplementary Figure 1). For predicting postoperative OS and DFS, the optimal cutoff values for SIRI were 1.24 and 1.15, respectively (area under the curve [AUC] = 0.796, 0.716, respectively). Based on the OS corresponding cut-off value, divide all patients into two groups for a comparison of basic information. The results indicated that there were statistically significant differences between the two groups in terms of gender, tumor diameter, MWA power, complication, tumor differentiation, gene mutation, neutrophil count, lymphocyte count, monocyte count, etc. (Table 1).

TABLE 1 Characteristics of the entire cohort in terms of demographics and baseline data.

Variables	Total (n=265)	Low SIRI group (n=188)	High SIRI group (n=77)	P
Age	69.1 ± 9.7	68.5 ± 9.4	70.7 ± 10.1	0.103
Gender (male)	177 (66.8)	115 (61.2)	62 (80.5)	<0.001
Comorbidity				
Hypertension	68 (25.7)	44 (23.4)	24 (31.2)	0.189
Diabetes	35 (13.2)	22 (11.7)	13 (16.9)	0.258
Coronary heart disease	15 (5.7)	9 (4.8)	6 (7.8)	0.336
Smoking	50 (18.9)	32 (17.0)	18 (23.4)	0.230
Tumor diameter	2.0 ± 1.0	1.8 ± 1.0	2.4 ± 1.0	<0.001
MWA power (High)	73 (27.5)	53 (28.2)	20 (26.0)	<0.001
Complication	92 (34.7)	61 (32.4)	31 (40.3)	<0.001
Tumor differentiation				**<0.001**
Poor	33 (12.5)	12 (6.4)	21 (27.3)	
Moderate and well	232 (87.5)	176 (93.6)	56 (72.7)	
Gene mutation	66 (24.9)	51 (27.1)	15 (19.5)	<0.001
Neutrophil ($\times10^9$/L)	4.0 ± 1.9	3.2 ± 0.9	5.9 ± 2.2	<0.001
Lymphocyte ($\times10^9$/L)	1.7 ± 0.6	1.8 ± 0.5	1.3 ± 0.6	<0.001
Monocyte ($\times10^9$/L)	0.5 ± 0.2	0.4 ± 0.1	0.5 ± 0.2	<0.001
Platelet ($\times10^9$/L)	207.3 ± 58.8	206.0 ± 54.1	207.9 ± 66.1	0.810

MWA, microwave ablation.

3.2 SIRI and long-term prognosis

We applied Cox regression analysis to explore the factors affecting the long-term prognosis of lung carcinoma patients who underwent MWA, and the results were shown in Tables 2 and 3. Because there was multicollinearity among neutrophil count, lymphocyte count, monocyte count, and SIRI, and SIRI was more systematic and comprehensive compared to the other three, we have therefore only included SIRI in the multivariate analysis.

Multivariable cox regression analysis showed that tumor diameter, MWA power (High), tumor differentiation (Poor), EGFR mutations, and SIRI (High) were independent related factors for OS (hazard ratio [HR] 2.271, P<0.001; HR 0.510, P=0.010; HR 3.983, P<0.001; HR 0.151, P<0.001 and HR 2.583, P<0.001, respectively; Table 2). Tumor diameter, MWA power (High), tumor differentiation (Poor), EGFR mutations, and SIRI (High) were independent related factors for DFS (HR 2.214, P<0.001; HR 0.500, P=0.005; HR 3.604, P<0.001; HR 0.144, P<0.001 and HR 2.391, P<0.001, respectively; Table 3).

Figures 2A, B showed the Kaplan-Meier survival analysis correlating SIRI with the long-term prognosis of lung carcinoma patients receiving MWA. The long-term mortality and tumor recurrence rates of the SIRI (high) group were significantly higher (log-rank test, P<0.001, P<0.001, respectively).

TABLE 2 Cox regression examination investigating the impact of clinicopathological variables on patients' overall survival.

Variables	Univariable Cox regression		Multivariable Cox regression	
	HR (95% CI)	P	HR (95% CI)	P
Age	0.998 (0.977-1.019)	0.844		
Gender (male)	1.213 (0.983-1.523)	0.423		
Tumor diameter	2.122 (1.811-2.487)	<0.001	2.271 (1.865-2.765)	<0.001
MWA power (High)	0.600 (0.369-0.977)	0.040	0.510 (0.306-0.849)	0.010
Complication	1.623 (1.080-2.439)	0.020	1.078 (0.695-1.672)	0.738
Tumor differentiation (Poor)	7.226 (4.496-11.614)	<0.001	3.983 (2.418-6.563)	<0.001
Gene mutation				
No mutations	ref			
EGFR mutations	0.115 (0.042-0.314)	<0.001	0.151 (0.055-0.3418)	<0.001
KRAS mutations	0.453 (0.142-1.446)	0.181	0.269 (0.082-0.885)	0.031
Neutrophil	1.409 (1.277-1.554)	<0.001		
Lymphocyte	0.581 (0.411-0.821)	0.002		
Monocyte	5.485 (1.819-16.537)	0.003		
Platelet	1.002 (0.999-1.005)	0.276		
SIRI (High)	2.753 (1.832-4.137)	<0.001	2.583 (1.675-3.983)	<0.001

MWA, microwave ablation, EGFR, Epidermal Growth Factor Receptor, KRAS, Kirstenratsarcomaviraloncogenehomolog, SIRI, the ratio of monocyte multiply neutrophil to lymphocyte.

TABLE 3 Cox regression examination investigating the impact of clinicopathological variables on patients' disease-free survival.

Variables	Univariable Cox regression		Multivariable Cox regression	
	HR (95% CI)	P	HR (95% CI)	P
Age	0.990 (0.968-1.012)	0.359		
Gender (male)	1.067 (0.589-1.934)	0.831		
Tumor diameter	2.1647 (1.848-2.495)	<0.001	2.214 (1.859-2.637)	<0.001
MWA power (High)	0.670 (0.423-1.062)	0.088	0.500 (0.309-0.810)	0.005
Complication	1.129 (0.730-1.746)	0.585		
Tumor differentiation (Poor)	6.811 (4.295-10.802)	<0.001	3.604 (2.201-5.901)	<0.001
Gene mutation				
No mutations	ref			
EGFR mutations	0.112 (0.041-0.306)	<0.001	0.144 (0.052-0.395)	<0.001
KRAS mutations	0.754 (0.277-2.055)	0.581	0.554 (0.202-1.521)	0.252
Neutrophil	1.325 (1.167-1.505)	<0.001		
Lymphocyte	0.669 (0.486-0.920)	0.014		
Monocyte	1.491 (0.344-6.463)	0.593		
Platelet	1.002 (0.999-1.005)	0.283		
SIRI (High)	2.878 (1.941-4.266)	<0.001	2.391 (1.573-3.635)	<0.001

MWA, microwave ablation, EGFR, Epidermal Growth Factor Receptor, KRAS, Kirstenratsarcomaviraloncogenehomolog, SIRI, the ratio of monocyte multiply neutrophil to lymphocyte.

3.3 Construct nomograms based on independent related factors

We constructed nomograms to predict the long-term prognosis of lung carcinoma patients undergoing MWA based on independent factors identified through multivariate Cox regression analysis (Figures 3A, B). In the nomograms, each factor is assigned a specific score, and then the cumulative overall risk score is calculated based on these scores. Afterward, it is possible to intuitively predict the long-term prognosis of each patient after MWA by drawing a straight line downward.

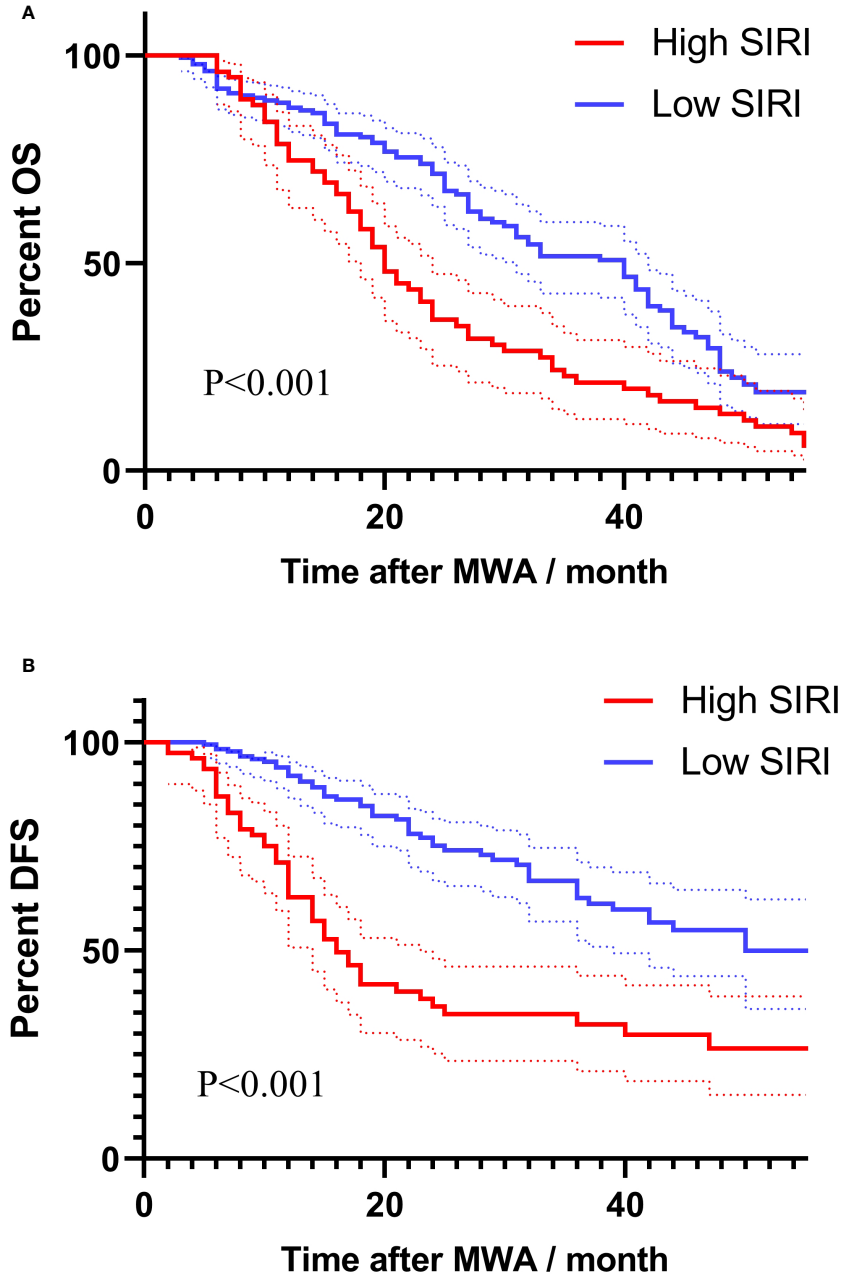


FIGURE 2
Survival analysis using Kaplan-Meier curves based on SIRI for long-term prognosis: **(A)** Kaplan-Meier survival curves based on SIRI for assessing overall survival. **(B)** Kaplan-Meier survival curves based on SIRI for assessing disease-free survival.

The ROC curves were used to evaluate the predictive value of nomograms for the long-term prognosis of lung carcinoma patients undergoing MWA surgery (Figures 4A, B). In OS, the areas under the curves (AUC) corresponding to 1, 2, and 3 years were 0.900, 0.849, and 0.862, respectively. As for DFS, the AUC corresponding to 1, 2, and 3 years were 0.851, 0.873, and 0.883, respectively.

4 Discussion

In this real-world study, the prognostic data of 265 pulmonary carcinoma patients following MWA in our center were retrospectively analyzed. SIRI was validated to be a novel and independent factor for predicting DFS and OS, outperforming SII, NLR, PLR and other inflammatory and immune scores. Furthermore, Our study is the first to explore the potential of preoperative SIRI in predicting survival among patients with lung carcinoma treated with MWA worldwide.

The role of chronic inflammation as a carcinogenic factor can be substantiated through various inflammatory conditions that are associated with an increased risk of cancer. The common underlying mechanism involves carcinogenesis through direct mutations or the activation of cellular cytokine responses. Prior research has traditionally concentrated on the impact of the

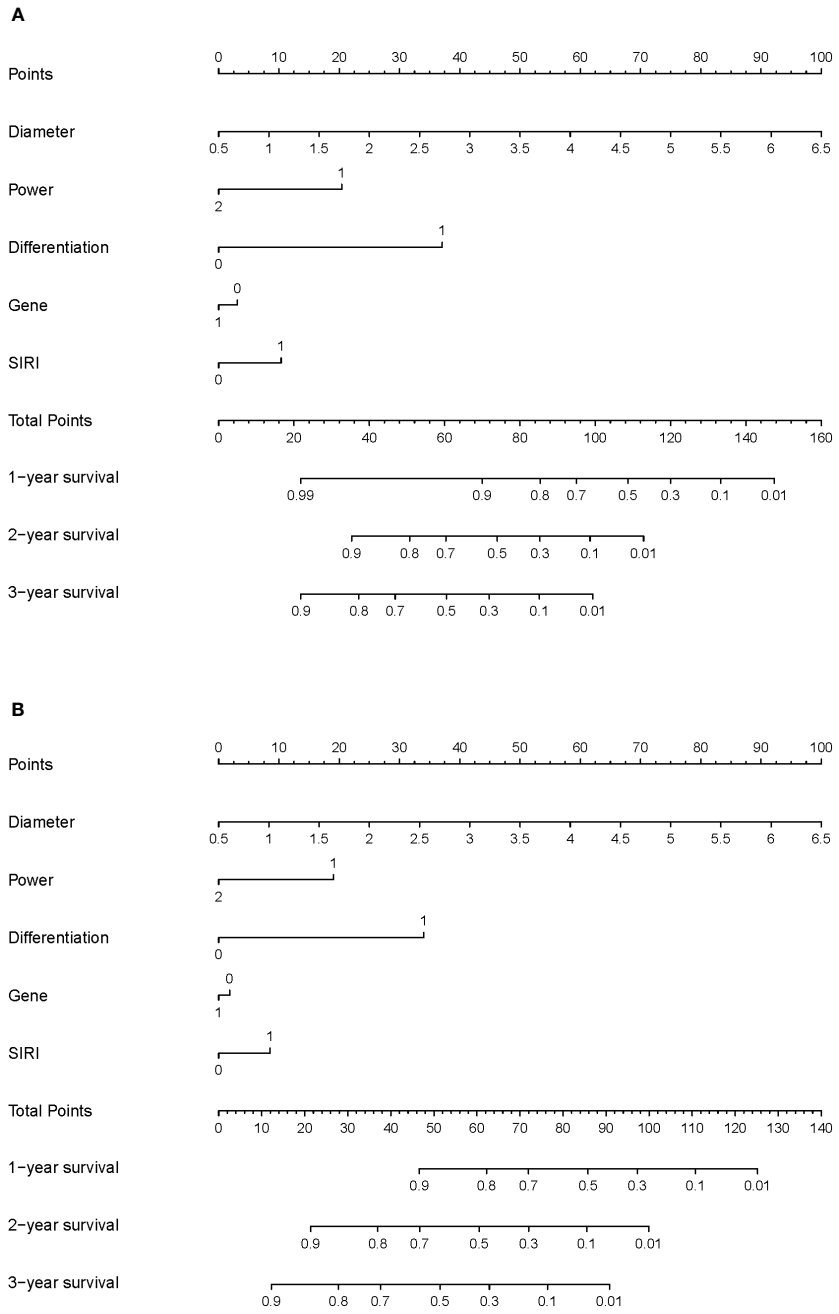


FIGURE 3
Nomogram for the long-term prognosis of lung cancer patients undergoing MWA: **(A)** Nomogram based on independent related factors for assessing overall survival. **(B)** Nomogram based on independent related factors for assessing disease-free survival. In "Power", "1" represents the low power group, and "2" represents the high power group. In "Differentiation", "0" represents the well differentiation group, and "1" represents the poor differentiation group. In "Gene", "0" represents no gene mutation group and ALK mutation, and "1" represents EGFR mutation group. In "SIRI", "0" represents the low SIRI group, and "1" represents the high SIRI group.

inflammatory process on malignant cells, with a focus on the role of genomic instability and mutations as fundamental concepts in the mechanisms underlying cancer development and pathogenesis (19). MWA, as an emerging interventional therapeutic approach, typically operates at two power frequencies, 915 MHz or 2450 MHz. Under the influence of the microwave electromagnetic field, polar molecules within tumor tissue, such as water and protein molecules, undergo extremely rapid vibrations, leading to mutual collisions and friction between molecules, which resulting in the rapid generation of temperatures ranging from 60 to 150 degrees Celsius within a short timeframe, causing coagulative necrosis of cells. Alongside inducing localized inflammatory changes, the coagulated scar tissue and necrotic tissue may also potentially trigger direct mutagenic effects. Inflammatory cytokines released by immune cells, stromal cells, and tumor cells, activated through nuclear factor-κB and signal transduction and transcription

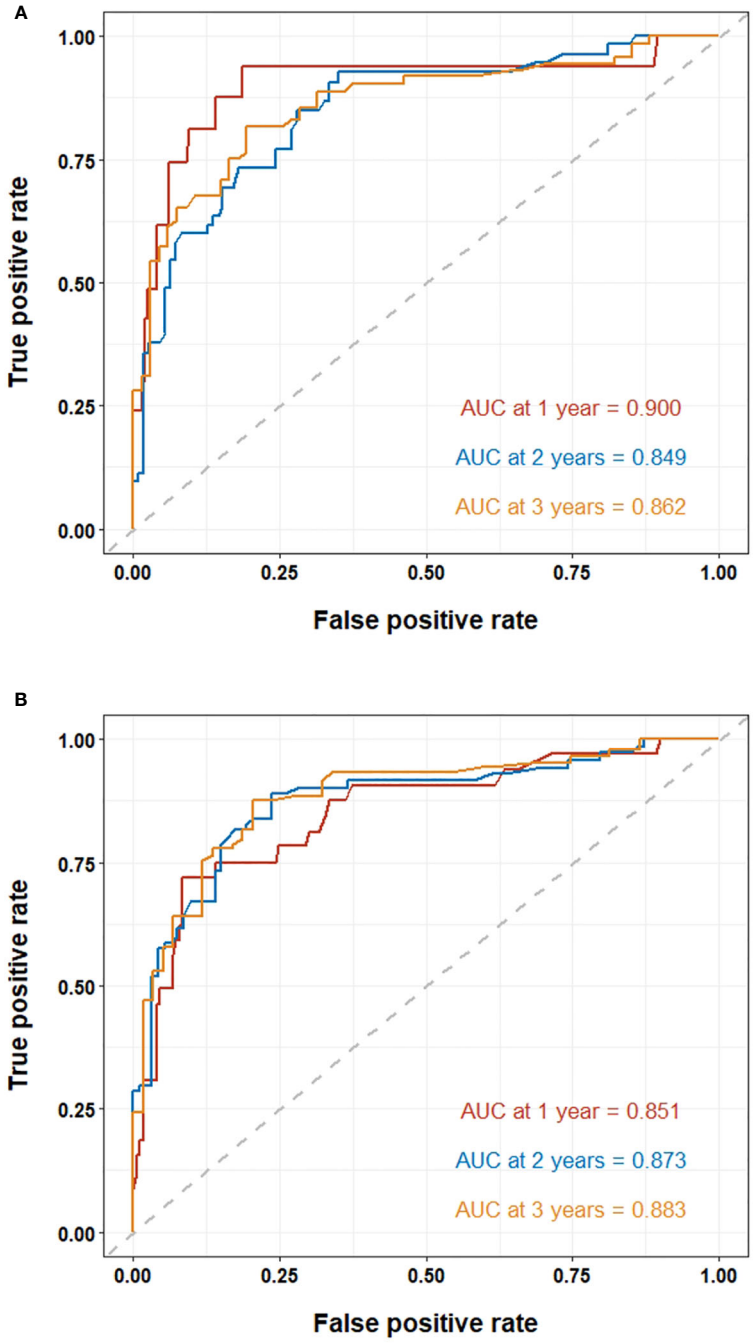


FIGURE 4
ROC curve of the nomogram for lung cancer patients undergoing MWA: **(A)** ROC curves of the nomogram for 1, 2, and 3-year overall survival. **(B)** ROC curves of the nomogram for 1, 2, and 3-year disease-free survival.

activation factors (STAT), stimulate malignant cell survival and proliferation pathways. Reactive oxygen species (ROS) originating from bone marrow cells can initiate carcinogenic and invasive behaviors (20).

In recent years, the concept of non-mutational epigenetic rearrangements that influence tumor growth and immune regulation has gained increasing attention. It's noteworthy that a single inflammatory event can drive sustained genetic and epigenetic changes, referred to as epithelial memory, subsequently triggering malignant progression of epithelial cells. In such instances, the continuous adaptive mutations of oncogenes are considered a physiological mechanism to limit tissue damage caused by inflammation. Despite the primary tumor being cleared, cancer patients may still experience recurrence due to dormant cells at distant sites that are clinically undetectable (21).

The role of immune-inflammatory responses and associated cells in the development of tumors is gradually being uncovered. The potential mechanisms underlying the effects of different

inflammatory biomarkers on prognosis are unclear; however, they are possibly related to changes in the tumor immune microenvironment. Moreover, tumor immune microenvironment changes are closely associated with inflammatory and immune cell distribution in the peripheral blood (22). As previously reported, biomarkers such as NLR and PLR have prognostic implications in various malignant tumors. Lymphocyte count is closely linked to acquired immunity, and its decrease is advantageous for tumor evasion of immune surveillance, playing a crucial role in the tumor defense system. lymphocytes play a crucial role in cancer therapy, and infiltration by lymphocytes stimulates the production of more pro-inflammatory factors, which promote cytotoxic cell death and inhibit tumor cell proliferation and migration (23, 24). On the contrary, neutrophils not only secrete inflammatory mediators to promote tumor proliferation and metastasis, but also enhance the adhesion ability of distant tumor cells. Elevated neutrophil levels in the bloodstream, along with the heightened expression of cancer-promoting and blood vessel growth-stimulating molecules like vascular endothelial growth factor (VEGF), nuclear factor kappa-B (NF-κB), C-X-C motif chemokine ligand 8 (CXCL8), granulocyte colony-stimulating factor (G-CSF), and transforming growth factor-β1 (TGF-β1), contribute to the creation of a conducive environment for tumor growth (25–27). Platelets contribute to the distant metastasis of tumors by promoting the epithelial-mesenchymal transition of tumor cells, safeguarding their migration to distant organs (28). Meanwhile, increased levels of circulating monocytes have demonstrated an ability to forecast poorer patient survival across various tumor models. Monocytes were observed transitioning into tumor-associated macrophages (29), secreting factors such as TNF-α and VEGF to facilitate both tumor expansion and angiogenesis. Additionally, they hinder the immune response against tumors *in vivo*. These monocytes have also been proven to encourage the movement of tumor cells by releasing enzymes that break down the extracellular matrix (30). In combination, these diverse types of white blood cells collectively constitute the immune milieu, exerting a substantial impact on the therapeutic efficacy of treatments. The Systemic Immune-Inflammation Index (SIRI), calculated based on lymphocyte, neutrophil, and monocyte levels, has demonstrated its ability to predict prognosis for malignancies (31–33). However, SIRI remains relatively unexplored in lung carcinoma patients undergoing MWA treatment. The findings of this study distinctly established a direct correlation between elevated SIRI values and unfavorable long-term outcomes within this particular population. In order to obtain low SIRI, additional neoadjuvant or adjuvant therapy should be performed in patients with MWA.

In recent years, as research into genetic mutations in lung carcinoma has advanced, genetic testing for patients has gained significant value in guiding treatment and predicting prognosis. Common genetic mutations in lung carcinoma include EGFR, KRAS, EML4-ALK, Ros1, and c-MET, among others. Personalized targeted therapies directed at specific mutation sites have become a first-line treatment option for patients with these genetic mutations. Among these, EGFR, being one of the more prevalent mutation genes in NSCLC patients, has seen a continuous evolution of EGFR tyrosine kinase inhibitors (TKIs), significantly improving the survival time of EGFR-positive patients. Since the ADJUVANT study first confirmed the significant prognostic benefit of EGFR-TKI application in postoperative EGFR-positive NSCLC patients (34), multiple subsequent clinical studies have similarly affirmed the positive impact of EGFR-TKIs on patient prognosis, significantly extending both DFS and OS compared to traditional chemotherapy. Guidelines have consequently incorporated targeted therapy as the preferred treatment option for postoperative adjuvant gene mutation-positive patients. Based on our findings, the prognosis of patients with identified gene mutations after MWA treatment was notably better than that of patients without gene mutations, with EGFR mutation-positive patients demonstrating superior outcomes compared to KRAS mutation-positive patients. This outcome is likely attributed to gene mutation-positive patients subsequently receiving targeted drug therapy. Third-generation EGFR-TKI drugs have shown significant efficacy in resistant patients with the T790M mutation, expanding the range of available drugs for EGFR mutation-positive patients and leading to overall improved prognosis. However, adjuvant targeted therapy may not be the sole reason for the extended prognosis in gene mutation-positive patients. A key distinction between our study and others lies in the fact that we examined the post-treatment prognosis of patients with identified gene mutations following MWA treatment, whereas most other studies primarily included postoperative patients. Cell experiments have revealed that the thermal effect increases the levels of the T790M mutation within cells, potentially offering resistant EGFR-TKI patients a longer duration of drug efficacy (35). Xu et al. found that patients with limited progression of TKI resistance significantly benefited from the addition of local ablation therapy (36). Clinical research by Ni et al. indicated that when patients who had experienced EGFR-TKI progression underwent MWA, reusing the same TKI drugs significantly extended both OS and PFS (37). Hence, microwave ablation therapy may prolong the prognosis of gene mutation-positive patients by enhancing the effectiveness of targeted treatments.

Furthermore, our research revealed another significant factor affecting the prognosis of MWA patients is the size of the tumor. Clinically, the maximum diameter of the primary tumor plays a pivotal role in tumor TNM staging and has been affirmed as an independent prognostic factor in several studies concerning lung carcinoma (38, 39). Simultaneously, larger-diameter tumors exhibit a significantly higher rate of recurrence after thermal ablation. Lee et al. found that among 30 lung carcinoma patients undergoing RFA, tumors with a diameter smaller than 3 cm achieved complete necrosis in 100% of cases, whereas this proportion dropped to only 23% in larger one (40). This finding aligns with our study results. In our predictive model, we believe that larger tumor size, which is inherently more invasive, and incomplete ablation due to excessive volume, are both associated with poorer postoperative survival in patients. This observation also provides a solid explanation for the comparison of the prognosis of lung carcinoma patients treated with different power levels of MWA in our other real-world study. Moreover, the degree of tumor differentiation is a crucial factor in assessing tumor malignancy. Poorly-differentiated tumors exhibit greater cellular heterogeneity, with significant differences from

normal tissue, and are more prone to local necrosis within the tumor. Sun's team found that among postoperative NSCLC patients, those with poorly differentiated cancer had a higher Hazard Ratio (HR), indicating a greater risk of recurrence compared to other patients (41). In our study, we conducted the first statistical analysis of the relationship between tumor differentiation and post-MWA prognosis in lung carcinoma patients. The conclusions suggest that the OS of poorly-differentiated group was significantly lower than that of highly-differentiated patients, emphasizing that tumor differentiation should also be considered a risk factor for post-MWA prognosis in lung carcinoma.

It is also noteworthy that in pathological classification, NSCLC accounts for approximately 85% of all lung carcinoma patients, among which the prognostic differences and reasons between lung adenocarcinoma (LUAD) and squamous cell carcinoma (LUSC) have not reached a consistent conclusion in multiple clinical studies. Our real-world study included post-MWA lung carcinoma patients of different pathological types, which is the first reported instance globally. Based on our follow-up data, the short-term OS of LUAD patients was significantly better than that of LUSC patients. This aligns with the findings of the Kawase's team, although their study attributes this phenomenon not to the histological differences between LUAD and LUSC but to the more significant comorbidity effects in LUSC patients during surgical resection (42). Fukui et al., in a retrospective study with a large number of patients, also confirmed that non-cancer-related mortality events in LUSC patients are significantly higher than in LUAD patients (43). Hence, LUSC patients often have comorbidities that could be one of the significant factors influencing the total OS. However, we believe that comorbidity may not be the sole reason for this OS difference. Unlike most studies conducted on surgical treatments, the patients selected for MWA treatment in this study mostly had indications that rendered them unsuitable for surgery. The patients' baseline conditions were relatively similar, and the bias brought about by comorbidities can be disregarded. Our study indicated a worse prognosis for LUSC patients, suggesting that there may be more potential explanations for this phenomenon. Wang et al. conducted a retrospective study on Chinese patients and found that platelet levels and systemic inflammation indices were significantly higher in LUSC patients than in LUAD patients, which may contribute to the tumor's development (44). Additionally, differences in gene expression between LUSC and LUAD involve several ontological subgroups, including cell proliferation, DNA replication, and cytoskeleton. Whole-genome analysis suggests significant structural differences in genes between the two. Research has shown that even the same expressed genes may have opposite effects in the signaling networks of LUAD and LUSC (45). Therefore, we believe that while the results show LUSC to be one of the risk factors for a worse prognosis, the impact of different histological types on prognosis awaits further explanation from pathological research.

Our study had several limitations. Firstly, despite being rooted in real-world data, the inherent nature of a single-center retrospective analysis introduces potential bias into the results, necessitating further validation through multi-center prospective studies. Secondly, our study did not incorporate repeated measurements at various time points to observe the dynamic changes of these biomarkers. This aspect will be addressed in our upcoming prospective registry study.

5 Conclusions

Elevated preoperative SIRI is an independent risk factor for poor prognosis in lung carcinoma patients undergoing MWA. Using nomograms based on SIRI to predict the long-term prognosis of these patients can assist clinical physicians in screening high-risk patients and formulating supportive treatments and personalized strategies.

Data availability statement

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

Ethics statement

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and was approved by the Ethics Committee of Beijing Chaoyang Hospital (2023-D.-471).

Author contributions

JW: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Supervision, Visualization, Writing – original draft, Writing – review & editing. S-PC: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Writing – original draft, Writing – review & editing. QZ: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft. YG: Investigation, Methodology, Validation, Writing – original draft. YJ: Investigation, Methodology, Validation, Writing – original draft. YL: Investigation, Methodology, Writing – review & editing. J-BM: Data curation, Investigation, Project administration, Writing – original draft. Y-LF: Conceptualization, Data curation, Formal analysis, Resources, Visualization, Writing – original draft, Writing – review & editing. BH: Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Supervision, Validation, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by Reform and development project of Beijing Institute of Respiratory Diseases, Beijing, PR China (Ggyfz202316) and Beijing Chaoyang Hospital Golden Seed Research Project, PR China (CYJZ202206).

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.2024.1305262/full#supplementary-material

References

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

2. Chan MV, Huo YR, Cao C, Ridley L. Survival outcomes for surgical resection versus CT-guided percutaneous ablation for stage I non-small cell lung cancer (NSCLC): a systematic review and meta-analysis. *Eur Radiol*. (2021) 31:5421–33. doi: 10.1007/s00330-020-07634-7

3. Vogl TJ, Naguib NN, Gruber-Rouh T, Koitka K, Lehnert T, Nour-Eldin NE. Microwave ablation therapy: clinical utility in treatment of pulmonary metastases. *Radiology*. (2011) 261:643–51. doi: 10.1148/radiol.11101643

4. Huang G, Yang X, Li W, Wang J, Han X, Wei Z, et al. A feasibility and safety study of computed tomography-guided percutaneous microwave ablation: a novel therapy for multiple synchronous ground-glass opacities of the lung. *Int J Hyperthermia*. (2020) 37 (1):414–22.

5. Akhan O, Güler E, Akıncı D, Çiftçi T, Köse I. Radiofrequency ablation for lung tumors: outcomes, effects on survival, and prognostic factors. *Diagn Interv Radiol (Ankara, Turkey)*. (2016) 22(1):65–71.

6. Yang X, Ye X, Lin Z, Jin Y, Zhang K, Dong Y, et al. Computed tomography-guided percutaneous microwave ablation for treatment of peripheral ground-glass opacity-Lung adenocarcinoma: A pilot study. *J Cancer Res Ther*. (2018) 14(4):764–71.

7. Li YC, Chen PH, Yeh JH, Hsiao P, Lo GH, Tan T, et al. Clinical outcomes of surgical resection versus radiofrequency ablation in very-early-stage hepatocellular carcinoma: a propensity score matching analysis. *BMC gastroenterology*. (2021) 21(1):418.

8. Denk D, Greten FR. Inflammation: the incubator of the tumor microenvironment. *Trends Cancer*. (2022) 8:901–14. doi: 10.1016/j.trecan.2022.07.002

9. Cao B, Liu M, Wang L, Zhu K, Cai M, Chen X, et al. Remodelling of tumour microenvironment by microwave ablation potentiates immunotherapy of AXL-specific CAR T cells against non-small cell lung cancer. *Nat Commun*. (2022) 13:6203. doi: 10.1038/s41467-022-33968-5

10. Singh R, Mishra MK, Aggarwal H. Inflammation, immunity, and cancer. *Mediators Inflamm*. (2017) 2017:6027305. doi: 10.1155/2017/6027305

11. Abolhassani H, Wang Y, Hammarström L, Pan-Hammarström Q. Hallmarks of cancers: primary antibody deficiency versus other inborn errors of immunity. *Front Immunol*. (2021) 12:720025. doi: 10.3389/fimmu.2021.720025

12. Wang J, Li H, Xu R, The MLR. NLR, PLR and D-dimer are associated with clinical outcome in lung cancer patients treated with surgery. *BMC Pulm Med*. (2022) 22:104. doi: 10.1186/s12890-022-01901-7

13. Liu J, Li S, Zhang S, Liu Y, Ma L, Zhu J, et al. Systemic immune-inflammation index, neutrophil-to-lymphocyte ratio, platelet-to-lymphocyte ratio can predict clinical outcomes in patients with metastatic non-small-cell lung cancer treated with nivolumab. *J Clin Lab Anal*. (2019) 33:e22964. doi: 10.1002/jcla.22964

14. He K, Si L, Pan X, Sun L, Wang Y, Lu J, et al. Preoperative systemic immune-inflammation index (SII) as a superior predictor of long-term survival outcome in patients with stage I-II gastric cancer after radical surgery. *Front Oncol*. (2022) 12:829689. doi: 10.3389/fonc.2022.829689

15. Li S, Yang Z, Du H, Zhang W, Che G, Liu L. Novel systemic inflammation response index to predict prognosis after thoracoscopic lung cancer surgery: a propensity score-matching study. *ANZ J Surg*. (2019) 89:E507–13. doi: 10.1111/ans.15480

16. Rice SJ, Belani CP. Diversity and heterogeneity of immune states in non-small cell lung cancer and small cell lung cancer. *PloS One*. (2021) 16:e0260988. doi: 10.1371/journal.pone.0260988

17. Wang L, Zhou Y, Xia S, Lu L, Dai T, Li A, et al. Prognostic value of the systemic inflammation response index (SIRI) before and after surgery in operable breast cancer patients. *Cancer biomark*. (2020) 28:537–47. doi: 10.3233/CBM-201682

18. Li S, Lan X, Gao H, Li Z, Chen L, Wang W, et al. Systemic Inflammation Response Index (SIRI), cancer stem cells and survival of localised gastric adenocarcinoma after curative resection. *J Cancer Res Clin Oncol*. (2017) 143:2455–68. doi: 10.1007/s00432-017-2506-3

19. Greten FR, Grivennikov SI. Inflammation and cancer: triggers, mechanisms, and consequences. *Immunity*. (2019) 51:27–41. doi: 10.1016/j.immuni.2019.06.025

20. Ni Y, Ye X, Wan C, Ni Q, Yang X, Huang G, et al. Percutaneous microwave ablation (MWA) increased the serum levels of VEGF and MMP-9 in Stage I non-small cell lung cancer (NSCLC). *Int J Hyperthermia*. (2017) 33:435–9. doi: 10.1080/02656736.2017.1284350

21. Yang CL, Wang SB, He WP, Liu JJ. Anti-oxidant and Anti-inflammatory Effects of Ethanol Extract from Polygala sibirica L. var megalopha Fr. on Lipopolysaccharide-Stimulated RAW264.7 Cells. *Chin J Integr Med*. (2023) 29(10):905–13. doi: 10.1007/s11655-023-3602-7

22. Tanaka T. Introduction for inflammation and cancer. *Semin Immunopathol*. (2013) 35:121–2. doi: 10.1007/s00281-012-0360-6

23. Ohsawa M, Hamai Y, Emi M, Ibuki Y, Kurokawa T, Yoshikawa T, et al. Neutrophil-to-lymphocyte ratio as a predictor of postoperative recurrence and prognosis in oesophageal squamous cell carcinoma. *Anticancer Res*. (2022) 42:1499–507. doi: 10.21873/anticanres.15622

24. Zhao Q, Chen G, Ye L, Shi S, Du S, Zeng Z, et al. Treatment-duration is related to changes in peripheral lymphocyte counts during definitive radiotherapy for unresectable stage III NSCLC. *Radiat Oncol*. (2019) 14:86. doi: 10.1186/s13014-019-1287-z

25. Zou JM, Qin J, Li YC, Wang Y, Li D, Shu Y, et al. IL-35 induces N2 phenotype of neutrophils to promote tumor growth. *Oncotarget*. (2017) 8:33501–14. doi: 10.18632/oncotarget.16819

26. Mandaliya H, Jones M, Oldmeadow C, Nordman II. Prognostic biomarkers in stage IV non-small cell lung cancer (NSCLC): neutrophil to lymphocyte ratio (NLR), lymphocyte to monocyte ratio (LMR), platelet to lymphocyte ratio (PLR) and advanced lung cancer inflammation index (ALI). *Transl Lung Cancer Res*. (2019) 8:886–94. doi: 10.21037/tlcr.2019.11.16

27. Korbecki J, Simińska D, Gąssowska-Dobrowolska M, Listos J, Gutowska I, Chlubek D, et al. Chronic and cycling hypoxia: drivers of cancer chronic inflammation through HIF-1 and NF-κB activation: A review of the molecular mechanisms. *Int J Mol Sci*. (2021) 22:10701. doi: 10.3390/ijms221910701

28. Liu Y, Zhang Y, Ding Y, Zhuang R. Platelet-mediated tumor metastasis mechanism and the role of cell adhesion molecules. *Crit Rev Oncol Hematol*. (2021) 167:103502. doi: 10.1016/j.critrevonc.2021.103502

29. Chen X, Li Y, Xia H, Chen YH. Monocytes in tumorigenesis and tumor immunotherapy. *Cells*. (2023) 12:1673. doi: 10.3390/cells12131673

30. Singhal S, Stadanlick J, Annunziata MJ, Rao AS, Bhojnagarwala PS, O'Brien S, et al. Human tumor-associated monocytes / macrophages and their regulation of T cell responses in early-stage lung cancer. *Sci Transl Med*. (2019) 11(479):eaat1500. doi: 10.1126/scitranslmed.aat1500

31. Pacheco-Barcia V, Mondéjar Solís R, France T, Asselah J, Donnay O, Zogopoulos G, et al. A systemic inflammation response index (SIRI) correlates with survival and predicts oncological outcome for mFOLFIRINOX therapy in metastatic pancreatic cancer. *Pancreatology*. (2020) 20:254–64. doi: 10.1016/j.pan.2019.12.010

32. Chao B, Ju X, Zhang L, Xu X, Zhao Y. A novel prognostic marker systemic inflammation response index (SIRI) for operable cervical cancer patients. *Front Oncol*. (2020) 10:766. doi: 10.3389/fonc.2020.00766

33. Hua X, Long ZQ, Huang X, Deng JP, Wen W, He ZY, et al. The preoperative systemic inflammation response index (SIRI) independently predicts survival in postmenopausal women with breast cancer. *Curr Probl Cancer*. (2020) 44:100560. doi: 10.1016/j.currproblcancer.2020.100560

34. Zhong WZ, Wang Q, Mao WM, Xu ST, Wu L, Shen Y, et al. ADJUVANT investigators. Gefitinib versus vinorelbine plus cisplatin as adjuvant treatment for stage II-IIIA (N1-N2) EGFR-mutant NSCLC (ADJUVANT/CTONG1104): a randomised, open-label, phase 3 study. *Lancet Oncol*. (2018) 19:139–48. doi: 10.1016/S1470-2045(17)30729-5

35. Wang J, Ling X, Zhou M, Ding G, Peng B, Wan J. Thermal treatment decreases resistance to osimertinib in non-small cell lung cancer through the EGFR/PI3K/AKT pathway. *Neoplasma*. (2021) 68:535–45. doi: 10.4149/neo_2021_200506N489

36. Xu Q, Liu H, Meng S, Jiang T, Li X, Liang S, et al. First-line continual EGFR-TKI plus local ablative therapy demonstrated survival benefit in EGFR-mutant NSCLC patients with oligoprogressive disease. *J Cancer*. (2019) 10:522–9. doi: 10.7150/jca.26494

37. Ni Y, Ye X, Yang X, Huang G, Li W, Wang J, et al. Microwave ablation as local consolidative therapy for patients with extracranial oligometastatic EGFR-mutant non-small cell lung cancer without progression after first-line EGFR-TKIs treatment. *J Cancer Res Clin Oncol*. (2020) 146:197–203. doi: 10.1007/s00432-019-03043-6

38. Li Z, Xu W, Gu T, Cao X, Wu W, Chen L. Tumor size, but not consolidation-to-tumor ratio, is an independent prognostic factor for part-solid clinical T1 non-small cell lung cancer. *Thorac Cancer*. (2023) 14:602–11. doi: 10.1111/1759-7714.14788

39. Jia B, Zheng Q, Wang J, Sun H, Zhao J, Wu M, et al. A nomogram model to predict death rate among non-small cell lung cancer (NSCLC) patients with surgery in surveillance, epidemiology, and end results (SEER) database. *BMC Cancer*. (2020) 20:666. doi: 10.1186/s12885-020-07147-y

40. Lee JM, Jin GY, Goldberg SN, Lee YC, Chung GH, Han YM, et al. Percutaneous radiofrequency ablation for inoperable non-small cell lung cancer and metastases: preliminary report. *Radiology*. (2004) 230:125–34. doi: 10.1148/radiol.2301020934

41. Sun Z, Aubry MC, Deschamps C, Marks RS, Okuno SH, Williams BA, et al. Histologic grade is an independent prognostic factor for survival in non-small cell lung cancer: an analysis of 5018 hospital- and 712 population-based cases. *J Thorac Cardiovasc Surg*. (2006) 131:1014–20. doi: 10.1016/j.jtcvs.2005.12.057

42. Kawase A, Yoshida J, Ishii G, Nakao M, Aokage K, Hishida T, et al. Differences between squamous cell carcinoma and adenocarcinoma of the lung: are adenocarcinoma and squamous cell carcinoma prognostically equal? *Jpn J Clin Oncol*. (2012) 42:189–95. doi: 10.1093/jjco/hyr188

43. Fukui T, Taniguchi T, Kawaguchi K, Fukumoto K, Nakamura S, Sakao Y, et al. Comparisons of the clinicopathological features and survival outcomes between lung cancer patients with adenocarcinoma and squamous cell carcinoma. *Gen Thorac Cardiovasc Surg*. (2015) 63:507–13. doi: 10.1007/s11748-015-0564-5

44. Wang W, Liu H, Li G. What's the difference between lung adenocarcinoma and lung squamous cell carcinoma? Evidence from a retrospective analysis in a cohort of Chinese patients. *Front Endocrinol (Lausanne)*. (2022) 13:947443. doi: 10.3389/fendo.2022.947443

45. McKay JD, Hung RJ, Han Y, Zong X, Carreras-Torres R, Christiani DC, et al. Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nat Genet*. (2017) 49:1126–32. doi: 10.1038/ng.3892

Frontiers in Oncology

Advances knowledge of carcinogenesis and tumor progression for better treatment and management

The third most-cited oncology journal, which highlights research in carcinogenesis and tumor progression, bridging the gap between basic research and applications to imrpove diagnosis, therapeutics and management strategies.

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