frontiers | Research Topics

The application of multi-omics analysis in translational medicine

Edited by
HaiHui Huang and Madhu Chetty

Published in
Frontiers in Medicine
Frontiers in Systems Biology

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 978-2-8325-7292-4
DOI 10.3389/978-2-8325-7292-4

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

About Frontiers

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

Frontiers journal series

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

Dedication to quality

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

What are Frontiers Research Topics?

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

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

The application of multi-omics analysis in translational medicine

Topic editors
HaiHui Huang — Shaoguan University, China
Madhu Chetty — Federation University Australia, Australia

Citation
Huang, H., Chetty, M., eds. (2026). *The application of multi-omics analysis in translational medicine*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-7292-4

Table of contents

frontiers | Frontiers in Medicine

TYPE Editorial
PUBLISHED 04 December 2025
DOI 10.3389/fmed.2025.1744814

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Victoria Bunik,
Lomonosov Moscow State University, Russia

*CORRESPONDENCE
Haihui Huang
✉ tomyhwang@163.com
Madhu Chetty
✉ madhu.chetty@federation.edu.au

RECEIVED 12 November 2025
REVISED 17 November 2025
ACCEPTED 18 November 2025
PUBLISHED 04 December 2025

CITATION
Huang H and Chetty M (2025) Editorial: The application of multi-omics analysis in translational medicine.
Front. Med. 12:1744814.
doi: 10.3389/fmed.2025.1744814

COPYRIGHT
© 2025 Huang and Chetty. 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: The application of multi-omics analysis in translational medicine

Haihui Huang[1,2]* and Madhu Chetty[3]*

[1]Shaoguan Research Center of the National Engineering Laboratory for Big Data System Computing Technology, Shaoguan University, Shaoguan, China, [2]Guangdong Provincial Laboratory of Traditional Chinese Medicine, Hengqin, China, [3]Institute of Innovation, Science and Sustainability, Federation University Australia, Ballarat, VIC, Australia

KEYWORDS

multi-omics analysis, translational medicine, data integration, personalized medicine, computational biology

Editorial on the Research Topic
The application of multi-omics analysis in translational medicine

The primary goal of translational medicine is to convert fundamental biological discoveries into tangible improvements in human health. Achieving this requires a holistic understanding of the complex molecular networks governing disease. Multi-omics analysis has become an essential paradigm in this endeavor, integrating data from diverse layers such as genomics, transcriptomics, proteomics, and metabolomics. This integration is crucial for bridging the gap between basic research and clinical application, facilitating precise diagnostics and personalized therapies.

However, the application of multi-omics approaches faces challenges, including the complexity of data integration, the interpretation of high dimensional datasets, and the standardization required for clinical implementation. This Research Topic, "*The Application of Multi-omics Analysis in Translational Medicine*", presents 14 articles that navigate these challenges and showcase advancements in this rapidly evolving field. The Research Topic underscores the power of integrative strategies across a spectrum of methodological innovations and complex diseases.

Advancing computational methodologies and data integration

The volume and complexity of multi-omics data demand sophisticated computational tools. A significant focus within this Research Topic is the development of artificial intelligence (AI) and machine learning frameworks, alongside the integration of diverse data types, to enhance predictive accuracy.

Predicting drug response is central to precision oncology. Miao et al. introduced an innovative drug response prediction model (NMDP) to address challenges in feature extraction and data fusion. Their model utilizes an interpretable semi supervised weighted SPCA module and integrates convolution methods with Kolmogorov Arnold Networks, demonstrating superior performance in predicting drug sensitivity.

Prioritizing actionable drug targets from vast genomic landscapes remains a significant hurdle. Gu and Chen developed GETgene AI, a framework that combines network-based prioritization, machine learning, and automated literature analysis powered by advanced language models. Applied to pancreatic cancer, GETgene AI successfully prioritized high priority targets, illustrating how AI driven approaches can accelerate drug discovery.

The integration of molecular data with imaging modalities represents another critical frontier. Huang Y. et al. explored the predictive potential of quantitative histopathological image features (HIF) in glioblastoma. By integrating HIF with genomics, transcriptomics, and proteomics, they found that the integrated multi-omics model significantly enhanced prognostic accuracy compared to single omics approaches.

Similarly, Li et al. developed a radiomics model for predicting chemoradiotherapy response in advanced non-small cell lung cancer. They integrated radiomic features from both the primary lesion and nodal disease with clinical data. This multimodal composite model demonstrated superior predictive performance, emphasizing the value of comprehensive data integration in clinical decision making.

Multi-omics insights into molecular mechanisms of oncological and chronic diseases

Multi-omics research continues to deepen our understanding of tumorigenesis, classification, and the interplay with systemic conditions and chronic diseases.

Loganathan and Doss investigated the interconnected molecular mechanisms between breast cancer and diabetes. Utilizing transcriptomic and exomic analyses across different cohorts, they identified shared pathways related to extracellular matrix organization and immune regulation. Their analysis highlighted the TNF pathway as a central link connecting chronic inflammation, insulin resistance, and tumor growth.

Pugazenthi et al. provided a review of the application of multi-omics analysis for pituitary neuroendocrine tumors (PitNETs). They summarized how integrated approaches have contributed to a deeper understanding of PitNET pathogenesis, revealing molecular subtypes and regulatory networks that inform classification and advance personalized medicine.

The power of multi-omics analysis extends to degenerative and inflammatory diseases. Zhang et al. aimed to identify novel risk genes for intervertebral disc disorder by integrating large scale multi-omics analyses, including transcriptome wide association studies and proteome wide association studies. Their integrative analysis and experimental validation confirmed the pathogenic roles of TMEM190, CILP2, and FOXO3, highlighting CILP2 as a potential druggable target.

Jin et al. focused on periodontitis by integrating transcriptomic and DNA methylation profiles. Their analysis explored the immune microenvironment and utilized machine learning to identify nine key diagnostic biomarkers. Subsequent network pharmacology analysis identified potential targeted drugs, offering new therapeutic avenues.

The nexus of microbiome, metabolism, and host response

The integration of microbiome and metabolomic data with host multi-omics profiles is rapidly emerging as a critical area of translational research, revealing intricate interactions between host metabolism, immune function, and microbial communities.

The connection between gut microbiota and systemic disease is an area of intense investigation. Liu L. et al. reviewed the emerging evidence surrounding the gut microbiota lung axis in lung cancer. They synthesized data indicating that gut dysbiosis is associated with worse prognosis and impacts the efficacy of immune checkpoint blockade, suggesting potential adjunctive therapeutic strategies through microbiome modulation.

The microbiome's role extends to neuroscience. Wang et al. employed a multi-omics approach to unravel the mechanisms of propofol induced psychological dependence. By integrating transcriptomics, metabolomics, and gut microbiome sequencing in a mouse model, they identified significant changes in neuroactive ligand receptor interaction pathways and gut microbial composition, suggesting a complex bidirectional signaling mechanism.

In the context of musculoskeletal health, Liu Y. et al. conducted an integrative analysis of serum microorganisms and serum metabolomics in osteoporosis patients. Their findings revealed distinct microbial compositions and significant differences in lipid metabolism pathways associated with osteoporosis, providing candidate biomarkers for early diagnosis.

Metabolomics also proved valuable in elucidating the mechanisms of traditional therapies. Sun et al. investigated the protective mechanisms of Angelica sinensis polysaccharide (ASP) against recurrent spontaneous abortion. Through metabolomic analysis and assessment of autophagy levels, they found that ASP restores diminished autophagy activity and regulates key metabolic pathways, including glycolysis/gluconeogenesis.

Emerging modalities and perspectives

The Research Topic also highlights the potential of novel biological entities and advanced analytical modalities. The diagnostic and therapeutic potential of exosomes was reviewed by Odehnalová et al.. As carriers of disease specific biomarkers, these extracellular vesicles offer opportunities for non-invasive detection, targeted drug delivery, and regenerative medicine in cancer and neurodegenerative diseases.

Furthermore, Huang F. F. et al. provided neuroimaging evidence for the central mechanisms of acupuncture in non-specific

low back pain through a systematic review and meta-analysis. Utilizing functional neuroimaging data, this study demonstrated that acupuncture modulates pain processing through the insula and limbic system, validating its clinical efficacy and exploring its underlying mechanisms.

The studies compiled in this Research Topic collectively demonstrate the profound impact of multi-omics analysis on translational medicine. By embracing integrative approaches, novel computational methods, and the inclusion of diverse data types such as microbiome profiles and imaging features, these investigations are significantly advancing the field. The insights generated here not only enhance our understanding of complex diseases but also pave the way for more precise diagnostics and personalized therapeutic strategies, bringing us closer to the realization of personalized healthcare.

Author contributions

HH: Writing – original draft, Writing – review & editing. MC: Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the Guangdong Key Construction Discipline Research Capacity Enhancement Project (2022ZDJS049) and the National Natural Science Foundation of China (62102261).

Acknowledgments

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

Conflict of interest

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

Generative AI statement

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

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

Publisher's note

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

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 15 January 2025
DOI 10.3389/fmed.2025.1522503

Check for updates

OPEN ACCESS

EDITED BY
HaiHui Huang,
Shaoguan University, China

REVIEWED BY
Kaiyuan Ji,
Guangzhou Medical University, China
Wenfei Zheng,
Yichang Central People's Hospital, China

*CORRESPONDENCE
Yuan Tan
✉ yuantan_shyfy@163.com
Shuyun Wang
✉ wsyz9999@hotmail.com

†These authors have contributed equally to this work

RECEIVED 04 November 2024
ACCEPTED 02 January 2025
PUBLISHED 15 January 2025

CITATION
Sun Y, Li G, Kong M, Li J, Wang S and Tan Y (2025) Angelica sinensis polysaccharide as potential protectants against recurrent spontaneous abortion: focus on autophagy regulation.
Front. Med. 12:1522503.
doi: 10.3389/fmed.2025.1522503

COPYRIGHT
© 2025 Sun, Li, Kong, Li, Wang and Tan. 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.

Angelica sinensis polysaccharide as potential protectants against recurrent spontaneous abortion: focus on autophagy regulation

Yeli Sun[1†], Guohua Li[2†], Mengwen Kong[1], Junyuan Li[1], Shuyun Wang[3*] and Yuan Tan[3*]

[1]Shanghai Key Laboratory of Maternal Fetal Medicine, Shanghai Institute of Maternal-Fetal Medicine and Gynecologic Oncology, Shanghai First Maternity and Infant Hospital, School of Medicine, Tongji University, Shanghai, China, [2]Shanghai Key Laboratory of Maternal Fetal Medicine, Department of Reproductive Immunology, Shanghai Institute of Maternal-Fetal Medicine and Gynecologic Oncology, Shanghai First Maternity and Infant Hospital, School of Medicine, Tongji University, Shanghai, China, [3]Department of Integrated Traditional Chinese Medicine (TCM) and Western Medicine, Shanghai First Maternity and Infant Hospital, School of Medicine, Tongji University, Shanghai, China

Introduction: Recurrent spontaneous abortion (RSA) represents a significant clinical challenge, with its underlying mechanisms yet to be fully elucidated. Despite advances in understanding, the precise pathophysiology driving RSA remains unclear. Angelica sinensis, a traditional herbal remedy, is frequently used as an adjunctive treatment for miscarriage. However, it remains uncertain whether its primary active component, Angelica sinensis polysaccharide (ASP), plays a definitive role in its therapeutic effects. The specific function and mechanism of ASP in the context of RSA require further investigation.

Methods: In this study, we sought to evaluate autophagy levels at the maternal-fetal interface in RSA patients and in an RSA mouse model treated with ASP, complemented by a comprehensive metabolomic analysis. Autophagy flux in the decidua was compared between eight RSA patients and eight healthy pregnant women. Additionally, changes in autophagy flux were assessed in an RSA mouse model following ASP treatment, with embryos and placental tissues collected for subsequent metabolomic profiling.

Results: Our results revealed a significant reduction in Beclin 1 protein levels in the decidua of RSA patients compared to the normal pregnancy group. Conversely, ASP treatment in the RSA mouse model restored autophagy-related protein expression, including ATG7, ATG16L, and Beclin 1, to levels higher than those observed in the untreated RSA group. Metabolomic analyses further identified significant changes in phosphatidylethanolamine levels between ASP-treated and control groups, with differential metabolites enriched in pathways related to glycolysis/gluconeogenesis, glycerolipid metabolism, and glycine, serine, and threonine metabolism. Functional assays revealed that ASP enhances trophoblast cell proliferation, migration, and invasion.

Conclusion: In summary, our findings demonstrate diminished autophagy activity in RSA patients, while ASP appears to restore autophagy and regulate key metabolic pathways, including glycolysis/gluconeogenesis. These results provide new insights into the protective mechanisms of ASP in RSA, suggesting its potential as a therapeutic intervention for this condition.

KEYWORDS

recurrent spontaneous abortion, Angelica sinensis polysaccharide, autophagy, metabolomics, Beclin 1

1 Introduction

Recurrent spontaneous abortion (RSA) is a prevalent and clinically significant obstetric complication, defined by the occurrence of two or more consecutive pregnancy losses prior to the 28th week of gestation (1). The known causes of RSA are diverse, including infectious agents, chromosomal abnormalities, hormonal and metabolic disorders, antiphospholipid syndrome, and structural abnormalities of the uterus (2–6). However, approximately 50% of RSA cases are classified as idiopathic, with no clearly identifiable underlying cause (7). Emerging evidence suggests that autophagy levels in trophoblast cells of RSA patients are dysregulated (8–11). However, whether RSA is driven by autophagy deficiency or excessive autophagy activation remains poorly understood, necessitating further investigation.

Autophagy is a fundamental cellular process that mediates the transport of intracellular components to lysosomes for degradation and recycling, ensuring cellular homeostasis and adaptation to stress (12). This highly orchestrated pathway is regulated by a network of autophagy-related genes (ATGs) and their associated proteins. Beyond its role in cellular maintenance, autophagy plays a pivotal role in early embryonic development and implantation (11). Dysregulated autophagy has been implicated in various pregnancy complications, including preeclampsia and fetal growth restriction (13, 14), However, investigations into the role of autophagy in RSA remain scarce, and existing findings are often inconsistent, highlighting the need for further focused research.

Angelica sinensis, a cornerstone of traditional Chinese medicine, has been historically employed for the treatment of gynecological disorders (15). Its primary bioactive components include Angelica sinensis polysaccharide (ASP), along with sugars such as xylose, galactose, glucose, arabinose, rhamnose, fucose, and galacturonic acid, with ASP recognized as the most significant therapeutic constituent (16). A growing body of research has highlighted the diverse pharmacological properties of ASP, including hepatoprotective effects (17, 18), anti-cancer activity (19, 20), anti-aging benefits (21–23), antioxidant capacity (24), and immune modulation (25).

Emerging evidence further supports the role of Angelica sinensis extracts in modulating autophagy (26, 27) and restoring immune balance in abortion-prone models (28, 29). Specifically, ASP have been shown to significantly suppress the expression of autophagy-related proteins, including microtubule-associated protein 1 light chain 3 (LC3)II/LC3I, thereby mitigating excessive mitochondrial autophagy (30, 31). Despite these promising findings, the precise regulatory effects of ASP on autophagy, particularly within the context of RSA, and its underlying mechanisms remain inadequately understood.

Thus, this study aims to elucidate whether ASP contributes to the adjunctive treatment of RSA through modulation of autophagy pathways and to identify the associated signaling pathways. Our findings demonstrate a significant reduction in Beclin 1 expression at the maternal-fetal interface in RSA patients, indicating impaired autophagy activity. In contrast, ASP treatment in RSA mice led to a marked upregulation of autophagy-related proteins, including ATG7, ATG16L, and Beclin 1, accompanied by alterations in the glycolysis/gluconeogenesis metabolic pathway. Furthermore, ASP was shown to enhance the proliferation, migration, and invasion of HTR-8/SVneo trophoblast cells, highlighting its potential to support trophoblast function.

2 Materials and methods

2.1 RSA animal model

The CBA/J and DBA/2 mouse strains are well-established models for investigating immune-mediated RSA. In this study, female CBA/J mice were mated with male DBA/2 mice (CBA/J × DBA/2) to induce RSA-like phenotypes, as previously described (1). A total of 30 female CBA/J mice (20 ± 2 g), 10 male DBA/2 mice (22 ± 2 g), and 5 male BALB/c mice (20 ± 2 g), all aged 9 weeks, were obtained from Beijing Huafukang Biotechnology Co., Ltd. During the experimental procedures, female mice were paired with male mice at a 2:1 mating ratio. During the experimental procedures, female mice were paired with male mice in a 2:1 mating ratio. The detection of a vaginal plug was designated as embryonic day 1 of pregnancy. The pregnant females were randomly divided into three experimental groups: a normal control group, a RSA control group and an ASP intervention group. The ASP group received daily oral gavage of ASP at a dose of 400 mg/kg (32), while the control groups were administered an equivalent volume of saline. Treatments commenced on the first day of pregnancy and continued for 2 weeks. At the end of the treatment period, all mice were anesthetized and euthanized in accordance with humane protocols. All experimental procedures were conducted under the approval of the Ethics Committee of Shanghai First Maternity and Infant Hospital.

2.2 Sample collection

Embryos and placentas were carefully harvested from the experimental mice and immediately rinsed with ice-cold phosphate-buffered saline (PBS) to remove residual blood and debris. The cleaned samples were then subjected to rapid flash-freezing in liquid nitrogen to preserve molecular integrity and stored at −80°C until further analysis.

2.3 Sample preparation

Frozen embryo and placenta samples, each weighing 15 milligrams, were carefully transferred into 1.5 mL Eppendorf tubes. To optimize the extraction process, two small steel beads were added to each tube, along with 0.3 mL of a methanol-to-water solution (4:1, vol/vol). Additionally, a reference solution containing 0.3 mg of L-2-chlorophenylalanine dissolved in methanol was included in each tube. The samples were subsequently incubated at −20°C for 30 min to enhance the extraction efficiency.

Following the initial storage period, the samples were subjected to ultrasonic extraction in an ice-water bath for 10 min to ensure thorough processing. They were then briefly stored at −20°C for 2 min before being ground at 60 Hz for 2 min to achieve a uniform mixture. Subsequently, the samples were centrifuged at 4°C and 13,000 rpm for 10 min, facilitating the separation of the supernatant. The collected supernatant was concentrated and dried using a freeze-drying centrifuge, yielding a final volume of 250 μL.

Each dried sample was processed with 300 μL of a methanol–water mixture (1:4, vol/vol). The mixture was vortexed for 30 s and subjected to ultrasonic extraction in an ice-water bath for 3 min. The

samples were then incubated at −20°C for 2 h to ensure thorough extraction. Following incubation, the samples were centrifuged at 13,000 rpm for 10 min at 4°C. A 150 μL aliquot of the supernatant was carefully collected with a crystal syringe, passed through a 0.22 μm microfilter, and transferred to LC vials. These vials were subsequently stored at −80°C to maintain sample integrity prior to liquid chromatography-mass spectrometry (LC–MS) analysis.

To ensure the reliability and consistency of the LC–MS analysis, a pooled sample derived from equal portions of each individual sample was prepared and utilized as a quality control (QC) sample.

2.4 LC–MS analysis

Metabolic profiling of the collected samples was conducted using a cutting-edge UPLC i-Series system (Waters Corporation, Milford, United States) coupled with the advanced VION IMS QTOF mass spectrometer (Waters Corporation, Milford, United States). Chromatographic separation was achieved on a high-resolution UPLC BEH C18 column (1.7 μm, 2.1 × 100 mm) in both positive and negative ionization modes. The mobile phase comprised water with 0.1% formic acid as solvent A and a 2:3 (vol/vol) acetonitrile/methanol mixture containing 0.1% formic acid as solvent B. A linear gradient elution program was employed as follows: 0 min, 1% B; 1 min, 30% B; 2.5 min, 60% B; 6.5 min, 90% B; 8.5 min, 100% B; 10.7 min, 100% B; 10.8 min, 1% B; and 13 min, 1% B. The column was maintained at a constant temperature of 45°C, with a flow rate of 0.4 mL/min. To ensure sample integrity, all samples were stored at 4°C throughout the analysis, with an injection volume precisely set to 1 μLs.

Mass spectrometry data acquisition was conducted using both full-scan mode (m/z range: 50–1,000) and MSE mode to ensure comprehensive coverage and enhanced fragmentation information. In MSE mode, alternating low- and high-energy scans were performed, enabling simultaneous acquisition of precursor and fragment ion data. The low-energy scans were conducted with a fixed collision energy of 4 eV, while high-energy scans employed a collision energy ramp ranging from 20 to 45 eV. Collision-induced dissociation was facilitated using high-purity argon gas (99.999%), with optimized instrument settings as follows: source temperature set to 115°C, desolvation gas temperature maintained at 450°C, cone voltage at 40 V, desolvation gas flow rate at 900 L/h, a scan interdelay of 0.02 s, and a scan time of 0.2 s.

To ensure data reproducibility and evaluate analytical repeatability, QC samples were systematically injected at regular intervals throughout the analysis, typically after every three sample injections. The QC samples, prepared as pooled extracts from all experimental samples, were used to monitor the relative standard deviations of both retention times and peak areas.

The specifications and details of the primary instruments utilized in this study are available in Supplementary material S1.

2.5 Data preprocessing

The raw LC–MS data were processed using Progenesis QI V2.3 software (Nonlinear Dynamics, Newcastle, United Kingdom), incorporating a comprehensive workflow that included baseline filtering, peak detection, integration, retention time correction, peak alignment, and normalization. The data processing pipeline utilized stringent parameters, including a 5% production threshold, 10 ppm product tolerance, and 5 ppm precursor tolerance, to ensure high fidelity and reproducibility. Compound identification was performed through a qualitative analysis using multiple reference databases, including the Human Metabolome Database (HMDB), LipidMaps (V2.3), Metlin, EMDB, PMDB, and a custom in-house database. Accurate mass-to-charge ratios (m/z), secondary fragment patterns, and isotopic distributions were employed as definitive criteria for compound annotation, ensuring precise and reliable metabolite identification.

2.6 Statistical analysis

The acquired data underwent rigorous preprocessing to ensure reliability and accuracy. Peaks with more than 50% missing values across groups (ion intensity = 0) were excluded. Zero values were imputed with half of the minimum detected value, and compounds were filtered based on qualitative criteria. Specifically, compounds scoring fewer than 36 points on a 60-point scale were deemed invalid and subsequently removed. Data from both positive and negative ion modes were integrated into a unified data matrix. To evaluate the overall distribution and confirm the stability of the analytical workflow, the consolidated matrix was subjected to principal component analysis (PCA) using the R programming environment.

To identify differential metabolites between experimental groups, we applied orthogonal partial least squares discriminant analysis (OPLS-DA) and partial least squares discriminant analysis (PLS-DA). Model quality was rigorously evaluated through 7-fold cross-validation and 200 response permutation tests to mitigate the risk of overfitting. The variable importance in projection (VIP) scores derived from the OPLS-DA model were utilized to quantify each variable's contribution to group separation. Metabolites were considered differentially expressed if they met the criteria of a VIP score greater than 1.0 and a p-value less than 0.05, determined using a two-tailed Student's t-test.

2.7 Kyoto encyclopedia of genes and genomes enrichment analysis

Pathway enrichment analysis of differential metabolites was conducted using their KEGG IDs, leveraging the KEGG database[1] and the analytical platform developed by Shanghai Oebiotech Co., Ltd.[2] Enrichment of metabolic pathways was determined using a hypergeometric test, with a significance threshold set at $p \leq 0.05$. A lower p-value indicated a higher degree of significance in the differences observed across metabolic pathways. Detailed calculation formulas and methodologies are provided in Supplementary material S2.

1 https://www.kegg.jp/

2 https://cloud.oebiotech.cn/task/

2.8 Ethical approval and clinical sample collection

This study was conducted with the approval of the Ethics Committee of the Obstetrics and Gynecology Hospital Affiliated to Tongji University (Ethical Approval Number: 22Y11922400). Clinical samples, including villi and decidual tissues, were collected from January 2024 to May 2024 at Shanghai First Maternity and Infant Hospital (also known as the Obstetrics and Gynecology Hospital Affiliated to Tongji University). The study population consisted of 8 patients diagnosed with RSA (RSA group) and 8 women with normal pregnancies (NC group).

The inclusion criteria for the RSA group comprised patients with a history of two or more consecutive unexplained spontaneous miscarriages occurring prior to 28 weeks of gestation. For the control group, participants were individuals undergoing elective termination of normal pregnancies, carefully matched to the RSA group based on baseline characteristics and with no prior history of spontaneous miscarriage.

Exclusion criteria encompassed any history of infections, reproductive tract abnormalities, endocrine disorders, or other identified causes of miscarriage. Baseline clinical characteristics for both groups are presented in Table 1, with additional details available in Supplementary material S3.

2.9 Western blotting

Total protein was extracted using RIPA lysis buffer (WB6001, Shanghai Wayo Biotechnology, Shanghai, China), and protein concentrations were quantified using the bicinchoninic acid (BCA) method (23,235, Thermo Scientific, Waltham, United States). Equal amounts of protein samples were resolved on SDS-PAGE gels and subsequently transferred onto PVDF membranes (IPVH00010, Millipore, Massachusetts, United States). Membranes were blocked with 5% non-fat milk at room temperature for 1 h, followed by overnight incubation at 4°C with primary antibodies (42,867, Cell Signaling Technology, Boston, United States). The following day, membranes were incubated with secondary antibodies for 1 h at room temperature. Immunoreactive proteins were visualized using the Tanon 5,200 imaging system (Tanon, Shanghai, China).

Grayscale intensities of protein bands were quantified using ImageJ software (NIH, Manassas, MD, United States). The relative expression of target proteins was normalized to internal controls, and mean values along with standard deviations were calculated for each group. Statistical comparisons were performed using two-tailed *t*-tests, with statistical significance defined as $p < 0.05$. All antibodies used in this experiment were obtained from the autophagy antibody kit supplied by Cell Signaling Technology.

2.10 Cell culture

HTR8-Svneo cells, a human chorionic trophoblast-derived cell line, were procured from the cell bank of Shanghai First Maternity and Infant Hospital. The cells were maintained in DMEM/F12 medium (C3130-0500, Biological Industries, Kibbutz Beit Haemek, Israel) supplemented with 10% fetal bovine serum and 1% penicillin–streptomycin (15140-122, Grand Island Biological Company, Montana, United States). Cultures were incubated in a humidified atmosphere of 95% air and 5% carbon dioxide at 37°C to ensure optimal growth conditions.

2.11 Drug preparation

ASP (Yuanye, Shanghai, China) was dissolved in complete culture medium to prepare a series of concentrations: 0 μg/mL, 0.001 μg/mL, 0.01 μg/mL, 0.1 μg/mL, 1 μg/mL, and 10 μg/mL. The solutions were then sterilized by filtration through a 0.22 μm pore-sized membrane filter to ensure sterility prior to subsequent experiments.

2.12 Assessment of cell proliferation capacity

HTR8 cells were seeded into 96-well plates at a density of 3,000 cells per well, with 100 μL of culture medium supplemented with specified concentrations of ASP. Each group included six replicates. After cell adhesion, the Cell Counting Kit-8 (CCK8, MedChemExpress, New Jersey, United States) reagent was added to the wells following the manufacturer's protocol. A blank control, containing culture medium and CCK8 reagent without cells, was included to account for background absorbance. The optical density (OD) at 450 nm was measured using a microplate reader, with the first measurement recorded as Day 1. Subsequent measurements were performed at 24 h intervals to monitor cell proliferation dynamics.

The net OD was determined by subtracting the OD value of the blank control from that of the experimental wells. Comparative analysis of OD values across groups was conducted, and proliferation curves were generated using GraphPad Prism (version 8.0.2). Statistical significance was assessed via repeated measures analysis, followed by the least significant difference (LSD) method for post-hoc comparisons.

TABLE 1 Clinical characteristics of the RSA group and NC group.

	NC group (Mean ± SEM)	RSA group (Mean ± SEM)	*p*-value	95% CI	
				Down	Up
Count	8	8			
Age (year)	32.25 ± 1.61	32.75 ± 1.16	0.805	−4.761	3.761
BMI (kg/m^2)	21.53 ± 1.00	22.79 ± 1.18	0.443	−4.738	2.222
Gestational age (week)	7.75 ± 0.53	9.13 ± 0.61	0.110	−3.104	0.354

2.13 Wound healing assay

To evaluate cell migration, a wound healing assay was performed. Cells were seeded into 6-well plates at a density of 600,000 cells per well and cultured in serum-free medium containing varying concentrations of the tested drug. Once the cell monolayer reached approximately 90% confluence, a sterile 200-μL pipette tip was used to create a uniform, vertical scratch across the well. Detached cells and debris were carefully removed by washing the wells 2–3 times with PBS. Images of the wound area were captured using an inverted microscope (Leica, Wetzlar, Germany) at predefined time intervals. At each time point, the culture medium was replenished to maintain optimal conditions. The wound area was quantified using ImageJ software to assess the rate of wound closure over time.

2.14 Transwell migration assay

To assess cell migration, 800 μL of medium containing 20% serum and the respective drug treatment was added to the lower chamber of the Transwell system, which was then placed in a 24-well plate. A total of 200 μL of cell suspension (containing 100,000 cells per well) was seeded into the upper chamber. The system was incubated at 37°C in a humidified incubator with 5% CO_2 for 16 h. Following incubation, cells that had migrated to the lower surface of the membrane were fixed with 4% paraformaldehyde at room temperature for 20 min. The fixed cells were stained with 0.1% crystal violet solution for 30 min. Non-migrated cells on the upper surface of the membrane were carefully removed using a cotton swab. The membranes were then rinsed with PBS to eliminate excess stain. Migrated cells were visualized and imaged under an inverted microscope for quantitative analysis.

2.15 Transwell invasion assay

In contrast to the migration assay, the invasion assay incorporates an additional step to assess cell invasive capabilities. The upper chamber is pre-coated with 100 μL of medium containing 10% Corning matrigel matrix (Corning, New York, United States) to mimic the extracellular matrix. The chamber is incubated at 37°C for 1 h to allow the Matrigel to solidify, after which the supernatant is carefully removed. The subsequent procedures, including cell seeding and incubation, follow the same protocol as described for the migration assay.

2.16 Data acquisition and statistical analysis

Image analysis was conducted using ImageJ software, while data visualization and statistical evaluations were performed with GraphPad Prism (version 8.0.2). Results are presented as the mean ± standard error of the mean. Statistical comparisons between groups were performed using one-way analysis of variance. *Post hoc* analyses were carried out using either the LSD or Bonferroni multiple comparison tests to assess the significance of intergroup differences. Statistical significance was defined as follows: $p < 0.05$ (*), $p < 0.01$ (**), or $p < 0.001$ (***).

2.17 Transmission electron microscopy

Cells were harvested via centrifugation and promptly fixed in a 2.5% glutaraldehyde solution at 4°C for a minimum of 3 h to preserve cellular structures. Following fixation, samples were washed three times with 0.1 M phosphate buffer and subsequently post-fixed in 1% osmium tetroxide at 4°C for 3 h. The specimens were then subjected to three additional washes with phosphate buffer, sequentially dehydrated in graded ethanol, and embedded in Epon 812 resin to ensure optimal preservation and sectioning quality.

Ultrathin sections, approximately 70 nm in thickness, were prepared using an ultramicrotome (Leica UC6) and carefully mounted onto copper grids coated with formvar support films. The sections were stained with uranyl acetate for 30 min to enhance contrast, followed by counterstaining with lead citrate for 15 min. Finally, the stained sections were visualized and imaged using a transmission electron microscope (Thermo Fisher Talos 120) operated at 120 kV.

3 Results

3.1 Clinical characteristics of the RSA and NC groups

The diagnostic criteria for RSA were defined according to the ESHRE guidelines (33, 34). The clinical characteristics of participants in the RSA group and the NC group are summarized in Table 1 and Supplementary material S3. No significant differences were observed between the two groups in terms of age (30–34 years), body mass index (BMI, 20–24), or gestational age (7–10 weeks) ($p > 0.05$).

3.2 Reduced autophagy levels at the maternal-fetal interface of RSA patients

To investigate autophagy activity at the maternal-fetal interface, we examined decidual tissues from RSA patients and compared them to those of the NC group. As illustrated in Figure 1, autophagy flux was assessed across 8 samples from the NC group and 8 samples from the RSA group. Western blot analysis revealed a significant reduction in the expression levels of autophagy-related proteins, including ATG5, ATG7, and ATG16L, in the RSA group. Notably, Beclin 1 levels were significantly decreased (p-value <0.05). Although the expression of several other proteins implicated in the autophagy pathway did not show significant differences, the overall autophagy levels in the RSA group displayed a clear downward trend.

3.3 ASP elevate autophagy levels at the maternal-fetal Interface in RSA model mice

Previous studies have demonstrated that ASP promotes autophagy activation (35) and improves outcomes in RSA animal models (36). Furthermore, ASP has been reported to confer protective effects during pregnancy (32). Based on these findings, we hypothesize that ASP may mitigate RSA by activating protective autophagy pathways. Data presented in Supplementary material S4 illustrate the miscarriage status of mice in the RSA model. WB analysis revealed that the

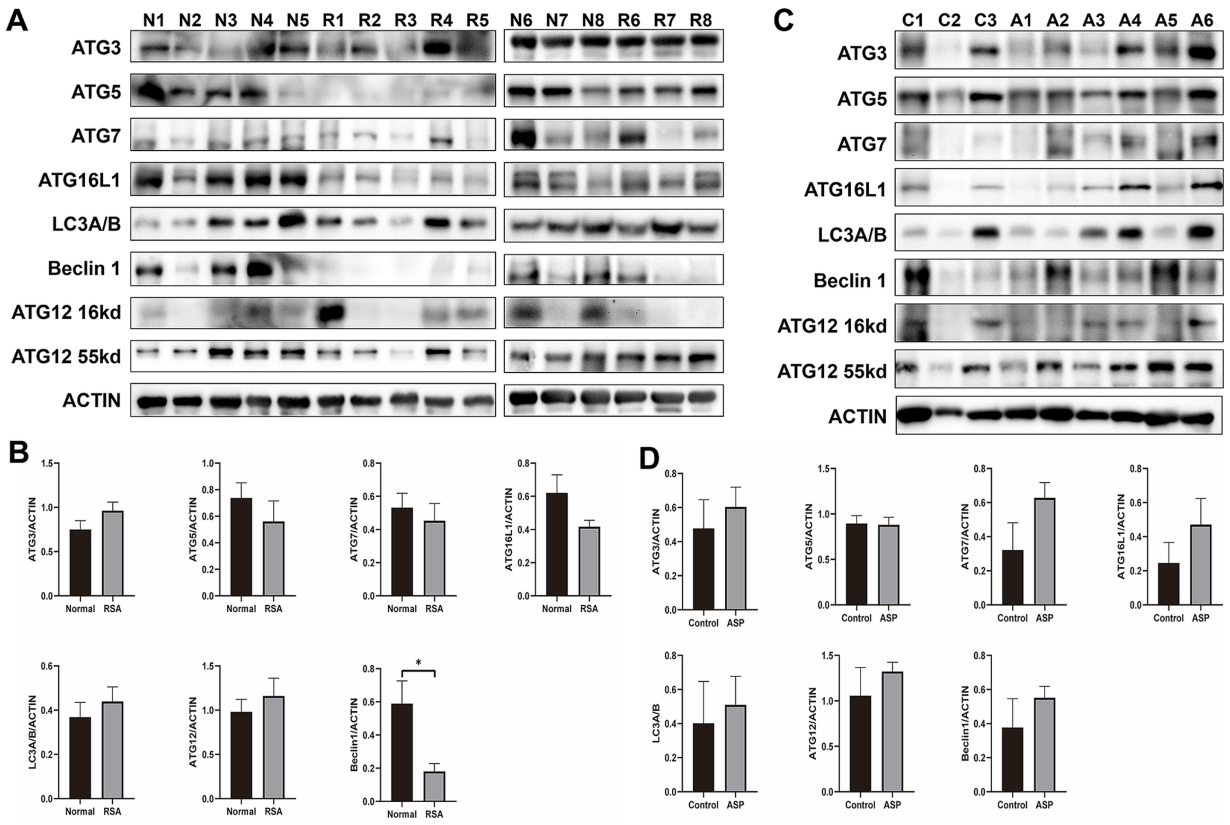


FIGURE 1
Western blot analysis of autophagy level changes in clinical samples from the RSA group and RSA model mice following ASP intervention. **(A)** Western blot analysis of autophagy-related proteins in samples from the RSA group compared to the Normal group. **(B)** Histogram showing the quantification of Western blot band intensities from **(A)**. **(C)** Western blot analysis of autophagy-related proteins in RSA model mice, comparing the ASP intervention group to the control group. **(D)** Histogram illustrating the quantification of Western blot band intensities from **(C)**.

expression levels of autophagy-related proteins, including ATG7, ATG16L, and Beclin 1, were significantly elevated in the ASP-treated group compared to the untreated RSA group (Figures 1C,D). Additionally, our previous metabolomic analysis highlighted enrichment of differential metabolites in the autophagy pathway, with pathway activity upregulated in the ASP-treated group relative to controls (Supplementary material S5).

3.4 Metabolite profiling of ASP and control groups

To investigate the metabolic alterations induced by ASP in the context of RSA, a metabolomic analysis was conducted comparing samples from the ASP-treated and control groups (Figure 2A). An OPLS-DA model revealed a clear and optimized class separation, demonstrating robust model fitting and effectively capturing the metabolic changes induced by ASP exposure (Figure 2B). Among the identified metabolites, 55 were significantly downregulated, and 42 were upregulated in the ASP group compared to the control group (Figure 2C; Supplementary materials S6, S7). Supporting our hypothesis, phosphatidylethanolamine (PE) was prominently altered between the two groups. The differential metabolites identified belong to several chemical classes, including benzene and substituted derivatives, carboxylic acids and derivatives, and fatty acyls, etc. (Supplementary materials S8, S9). Pathway enrichment analysis using the KEGG database highlighted significant enrichment of these metabolites in pathways such as glycolysis/gluconeogenesis, glycerolipid metabolism, glycine, serine, and threonine metabolism, nicotinate and nicotinamide metabolism, glyoxylate and dicarboxylate metabolism, Fc gamma R-mediated phagocytosis, and the Apelin signaling pathway (Figure 2D; Supplementary material S10). Subsequently, autophagosomes were observed by TEM. TEM further provided direct evidence of autophagic activity. Autophagosomes were visualized in the decidual tissues of the normal mouse model (Figure 2E), the RSA mouse model (Figure 2F), and the ASP-treated RSA mouse model (Figure 2G).

3.5 ASP enhances proliferation, migration and invasion of human chorionic trophoblast cells

The impact of ASP on the proliferation of HTR8 cells was assessed using the CCK-8 assay, revealing that ASP significantly promoted cell proliferation in a dose-dependent manner (Figure 3A). Consistently, transwell migration and scratch wound healing assays demonstrated a marked enhancement in the migratory capacity of HTR8 cells upon ASP treatment (Figures 3B,C,F,G). Additionally, the transwell invasion assay

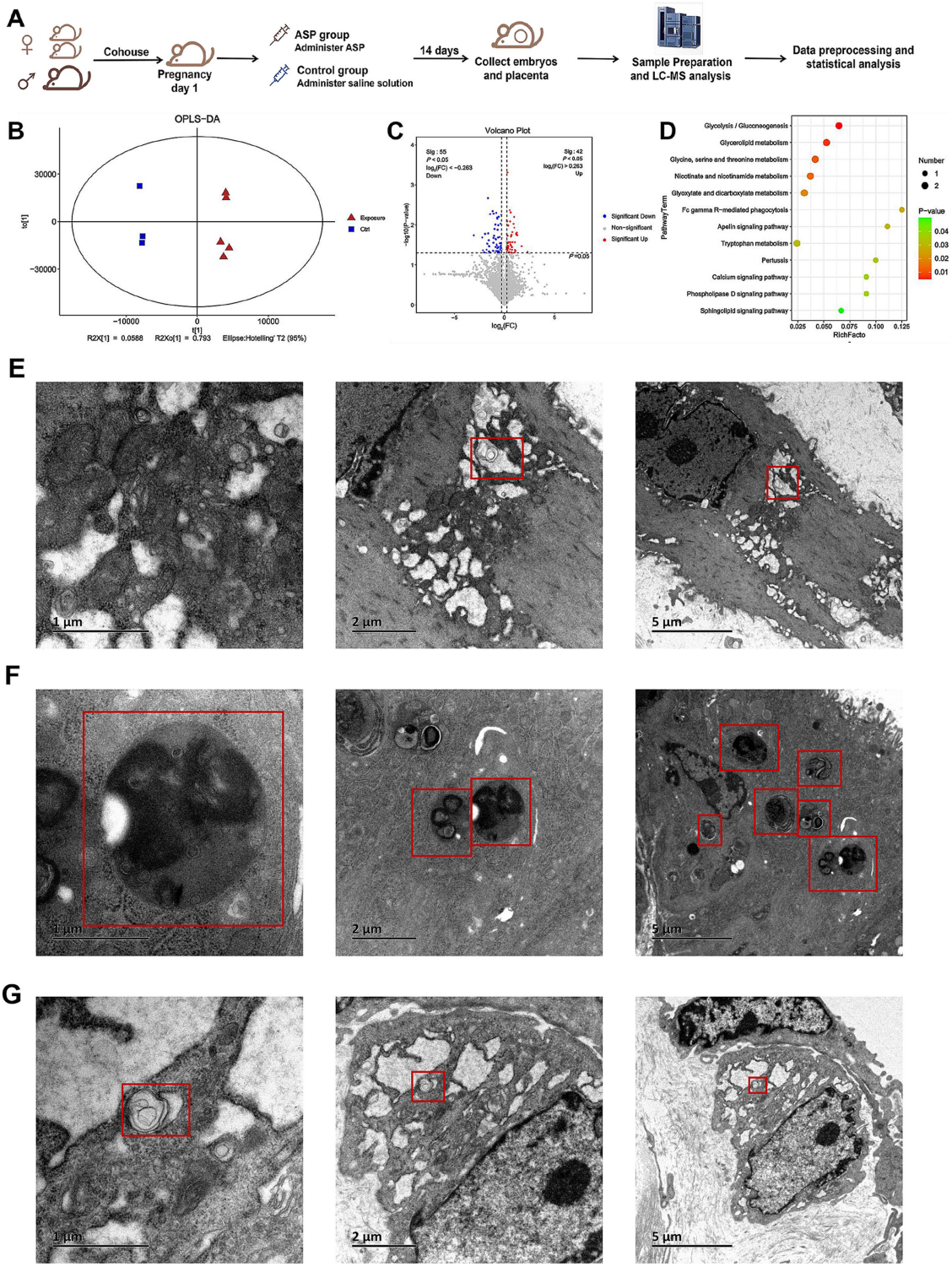


FIGURE 2
Establishment of the mouse model, metabolomics analysis, and transmission electron microscopy (TEM) images of autophagosomes. **(A)** workflow for constructing the mouse model. Female CBA/J mice and male DBA/2 mice were paired in a 2:1 ratio to establish the RSA model. On the first day of pregnancy, female mice were randomly assigned to either the ASP group or the control group, receiving ASP or an equivalent volume of saline, respectively. Samples were collected after 14 days for subsequent LC–MS analysis and data processing. **(B)** OPLS-DA analysis demonstrates a clear separation between the ASP and control groups. **(C)** Volcano plot of differential metabolites between the ASP and control groups. Each point in the figure represents a metabolite. The x-axis represents the $\log_2$(FC) value of the comparison between the two groups, while the y-axis represents the $-\log_{10}$(*p*-value). Red points indicate metabolites with $p < 0.05$ and fold change (FC) > 1, and blue points indicate metabolites with $p < 0.05$ and FC < 1. Gray points indicate non-significant differences ($p > 0.05$). **(D)** Bubble chart showing KEGG enrichment analysis of selected differential metabolites. **(E)** TEM revealed autophagosomes in the decidual tissues of mice. Autophagosomes in the decidua of mice from the normal control group. **(F)** Autophagosomes in the decidua of mice from the RSA control group. **(G)** Autophagosomes in the decidua of mice from the ASP intervention RSA group.

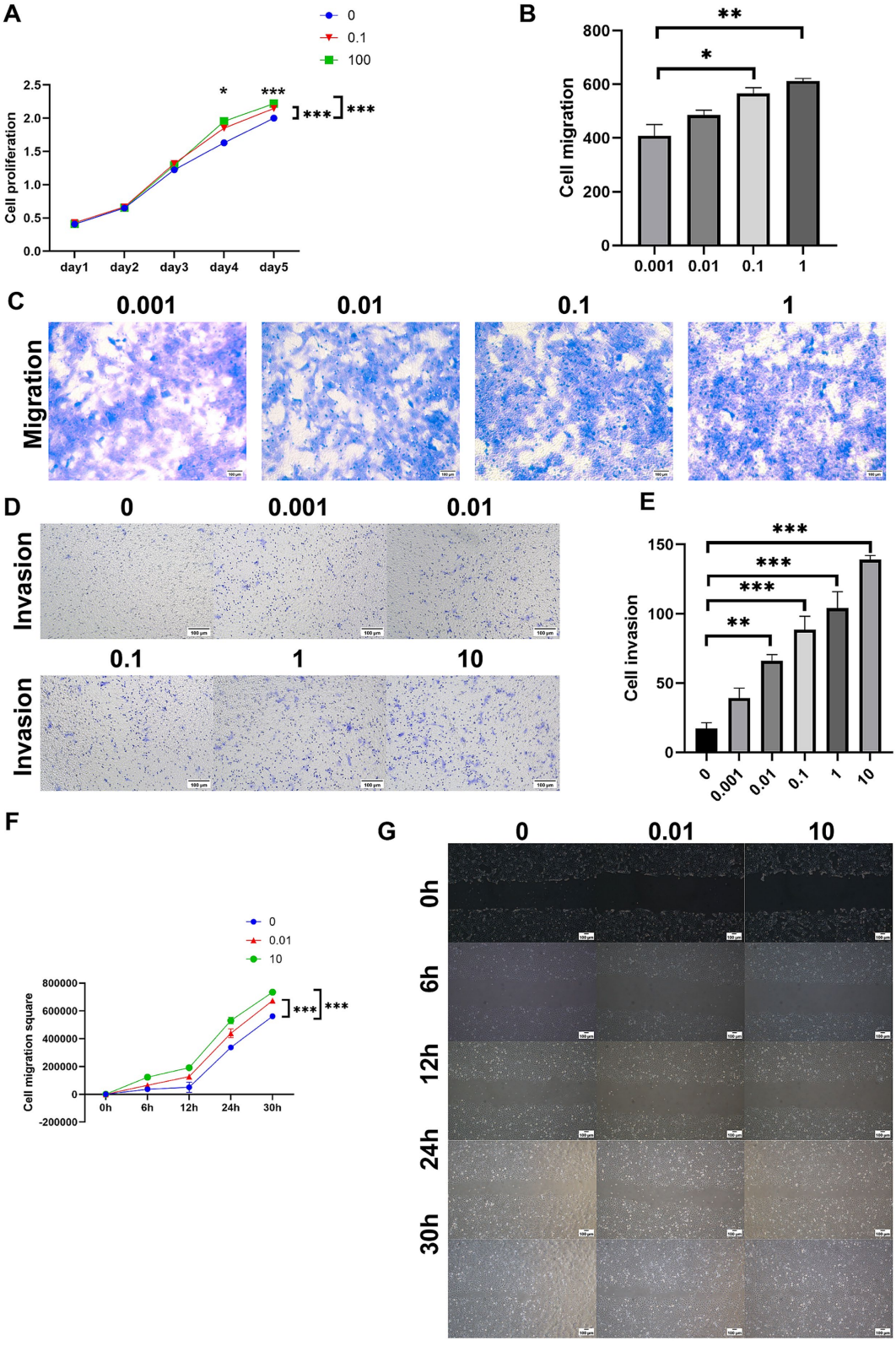


FIGURE 3
The effects of different concentrations of ASP (0, 0.001, 0.01, 0.1, 1, 10, 100 μg/mL) on the phenotype of HTR8 cells. **(A)** CCK8 assay showing the effect of ASP treatment at concentrations of 0, 0.1, and 100 μg/mL on the proliferation of HTR8 cells. **(B,C)** Transwell migration assay evaluating the effect of ASP treatment at concentrations of 0.001, 0.01, 0.1, and 1 μg/mL on the migration ability of HTR8 cells. **(D,E)** Transwell invasion assay showing the effect of ASP treatment at concentrations of 0, 0.001, 0.01, 0.1, 1, and 10 μg/mL on the invasion ability of HTR8 cells. **(F,G)** Scratch wound assay evaluating the effect of ASP treatment at concentrations of 0, 0.01, and 10 μg/mL on the migration ability of HTR8 cells.

further confirmed that ASP significantly facilitated the invasive ability of HTR8 cells (Figures 3D,E). Collectively, these findings indicate that ASP serves as a potent enhancer of trophoblast cell proliferation, migration, and invasion, underscoring its potential role in promoting trophoblast function in a dose-dependent manner.

4 Discussion

The relationship between RSA and autophagy remains an area of limited investigation, with findings to date presenting inconsistencies. Most research has predominantly focused on autophagy and its upstream and downstream signaling pathways, while metabolic aspects remain underexplored. Some studies have reported elevated autophagy levels in the chorionic tissues of RSA patients (11). Conversely, other investigations have observed a downregulation of autophagy-related genes in the chorion of RSA patients (9, 37, 38), suggesting that suppressed autophagy may lead to aberrant alterations in decidual natural killer (dNK) cell phenotypes, potentially contributing to pregnancy loss (10). Further studies investigating the immune microenvironment at the maternal-fetal interface have highlighted significant upregulation of autophagy-related proteins, such as Beclin 1, LC3B II/I, and BNIP3, in decidual macrophages of RSA patients (8). Impaired decidualization, a key factor influencing RSA, has been associated with reduced autophagy levels and disrupted uterine decidualization in RSA patients (39). Moreover, preclinical research indicates that hypericin, a bioactive compound, exerts protective effects against abortion in a rat model by enhancing autophagy (40). These findings collectively underscore the complex and multifaceted role of autophagy in RSA. However, the precise interplay between autophagy, immune regulation, and metabolism at the maternal-fetal interface remains to be fully elucidated. Further studies are warranted to clarify the mechanistic links between autophagy and RSA pathogenesis, which may pave the way for novel therapeutic strategies targeting this pathway.

Our experimental findings reveal a significant reduction in Beclin-1 levels in RSA patients compared to those with normal pregnancies, accompanied by a decreasing trend in ATG5, ATG7, and ATG16L expression. The autophagy pathway is initiated by the unc-51-like autophagy-activating kinase (ULK) complex, which orchestrates upstream signals to activate downstream processes. Beclin-1, a pivotal component of the autophagy-specific vacuolar protein sorting 34 (VPS34) complex I, plays a critical role in catalyzing the production of phosphatidylinositol-3-phosphate (PI3P) (PI3P). The generation of PI3P facilitates the recruitment of autophagy-related machinery, including the ATG16L1-ATG5-ATG12 complex, ATG3, and ATG7. These components work synergistically to conjugate ATG8 family members—encompassing the LC3 and GABARAP subfamilies—with PE, a key step in promoting autophagosome maturation and subsequent autophagic flux (12, 41).

Therefore, our findings indicate that autophagy levels at the maternal-fetal interface are diminished in RSA patients compared to those with normal pregnancies, aligning with previously reported observations (9, 37, 38). Importantly, in the RSA mouse model, treatment with ASP partially restored autophagy activity, suggesting a potential mechanism by which ASP confers protective effects in RSA. Beyond its influence on autophagy, ASP has demonstrated broader benefits in pregnancy-related contexts. For example, ASP has been shown to mitigate iron-deficiency anemia in pregnant rats by modulating the hepcidin-FPN1 axis (32). Furthermore, Angelica sinensis extracts, such as Ligustilide, have been reported to enhance pregnancy outcomes by improving endometrial receptivity and promoting angiogenesis within the endometrium (42). Additionally, Angelica sinensis has been implicated in alleviating metabolic disturbances in abortion-prone mice through the regulation of glycerolipid metabolism and has been shown to exert immunomodulatory effects (29, 43).

Simultaneously, ASP has demonstrated the ability to regulate autophagy through diverse signaling pathways. In the context of osteoarthritis, ASP has been reported to induce autophagy via activation of the ERK1/2 pathway (35). Similarly, ASP can mitigate chemotherapy-induced hepatotoxicity by enhancing autophagy through the MEK/ERK signaling cascade (44). Additionally, studies in a rat model of idiopathic pulmonary fibrosis revealed that Angelica sinensis exerts its autophagy-inducing effects via modulation of the mammalian target of rapamycin (mTOR) pathway (45).

In summary, ASP provide a degree of protection during pregnancy and can influence autophagy levels through different signaling pathways. Despite extensive evidence supporting the protective role of Angelica sinensis in pregnancy, research specifically investigating the contribution of ASP in miscarriage remains limited. The phytochemical composition of Angelica sinensis is highly complex, encompassing various bioactive compounds such as ASP, ligustrazine, laurene, ferulic acid, and vanillic acid (46). This complexity underscores the need for targeted studies to identify whether ASP represents the primary active component responsible for its therapeutic effects and to elucidate the molecular mechanisms involved. Our findings provide preliminary evidence suggesting that ASP may mitigate RSA by activating autophagy. However, further comprehensive investigations are required to validate these observations and explore the precise mechanisms underlying this protective effect.

Metabolomics has emerged as a robust and unbiased analytical approach, offering a comprehensive overview of an individual's metabolic profile (47). In this study, a metabolomic analysis was performed to compare the metabolic profiles of the ASP-treated group and the control group, revealing significant differences in PE levels between the two. PE plays a critical role in the autophagy pathway, serving as an essential lipid for the conjugation of LC3-I, facilitating its conversion into the autophagosome-associated form, LC3-II (48). These findings highlight the potential mechanistic link between ASP treatment and autophagy regulation.

KEGG pathway analysis highlighted significant alterations in metabolic pathways, including Glycolysis/Gluconeogenesis, Glycerolipid metabolism, and Glycine, serine, and threonine metabolism. Glycolysis/Gluconeogenesis has been implicated in impaired decidualization in pregnant rats (49) and is associated with defective trophoblast invasion in preeclampsia patients, as previously reported (50). Similarly, disruptions in Glycerolipid metabolism have been identified as potential biomarkers for

idiopathic infertility in *in vitro* fertilization (IVF) patients (51) and are linked to an increased risk of gestational diabetes in pregnant women (52). Consistent with our findings, metabolomic analyses of plasma from RSA patients have also revealed changes in Glycolysis/Gluconeogenesis and Glycerolipid metabolism (53). Both pathways have been further associated with preterm birth (54) and gestational diabetes risk (55). These studies collectively underscore the critical impact of metabolic dysregulation on pregnancy, highlighting the intricate connection between altered metabolic states and pregnancy complications.

However, there is a lack of relevant research when it comes to trends in the levels of specific metabolites within these pathways or whether intervention leads to reversals of these trends, particularly in RSA. Our findings reflect the possible metabolic mechanisms through which ASP exerts protective effects against RSA by regulating autophagy, providing a comprehensive overview of the metabolic profile changes induced by ASP at the maternal-fetal interface. Although these discoveries do not delve into deeper metabolic explorations, they lay a foundation and perspective for further investigation into the pathogenesis of RSA and its treatment.

However, this study has some limitations. Firstly, although our study initially found that ASP exerts protective effects on RSA by regulating autophagy, we did not further conduct dynamic validation to comprehensively observe autophagic flux. Instead, we focused only on autophagy levels at specific time points. This verification is crucial for establishing the reliability and applicability of the research findings. In future projects, we plan to perform additional experiments to observe autophagic flux in RSA and ASP intervention. These experiments may include observing LC3-labeled cells under a fluorescence microscope, using lysosome-specific fluorescent dyes to assess lysosomal function, measuring sequestosome 1 levels at different time points after intervention, and combining lysosomal inhibitors to validate autophagic flux activity. Secondly, the sample size in our study is relatively small, with only 8 participants in each clinical group, and the mouse experiments included only 6 and 3 samples from the ASP and control groups, respectively. This limitation restricts the generalizability of the findings and may affect the applicability of the results. Before ASP can be considered a protective factor for RSA, more in-depth validation in larger independent cohorts is necessary. Although our current metabolomics sequencing results indicate that PE is an important differential metabolite and that ASP affects key metabolic pathways, we have not yet conducted further experiments to explore how the metabolic changes induced by ASP specifically contribute to the pathological mechanisms of RSA. In future studies, we plan to exogenously add PE and other key metabolites to evaluate their effects on cell functions. Furthermore, in an RSA animal model with ASP intervention, we aim to measure the activity of key metabolic enzymes and lipid metabolic enzymes in critical pathways to investigate whether ASP improves RSA by regulating these enzyme activities and influencing metabolic pathways. At the same time, we will utilize specific inhibitors or gene knockout or knockdown methods to study the expression of key metabolic enzymes in cell models and assess their effects on cellular functions and the autophagy pathway.

5 Conclusion

In summary, we studied the levels of autophagy in the maternal-fetal interface of RSA patients and healthy pregnant individuals, and examined the changes in autophagy levels in ASP-treated RSA model mice, followed by a metabolomic analysis and cell phenotype assays. Our findings suggest that ASP may exert protective effects against RSA by activating autophagy while influencing pathways such as Glycolysis/Gluconeogenesis, Glycerolipid metabolism, and Glycine, serine, and threonine metabolism. However, further research and validation are necessary. Our results may provide insights for exploring the pathogenesis of RSA and offer evidence for the therapeutic effects of ASP in treating RSA.

Data availability statement

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

Ethics statement

The studies involving humans were approved by Ethics Committee of Shanghai First Maternity and Infant Hospital. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. The animal study was approved by Ethics Committee of Shanghai First Maternity and Infant Hospital. The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

YS: Data curation, Validation, Writing – original draft, Writing – review & editing. GL: Investigation, Resources, Writing – review & editing. MK: Conceptualization, Investigation, Writing – review & editing. JL: Conceptualization, Data curation, Investigation, Writing – review & editing. SW: Project administration, Resources, Writing – review & editing. YT: Project administration, Resources, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Science and Technology Commission of Shanghai Municipality (22Y11922400).

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.

Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2025.1522503/full#supplementary-material

SUPPLEMENTARY MATERIAL S1
The main instruments used during the LC-MS process, along with their models/specifications and manufacturers.

SUPPLEMENTARY MATERIAL S2
Calculation method for *P*-value in KEGG analysis.

SUPPLEMENTARY MATERIAL S3
Detailed clinical information of RSA group and NC group.

SUPPLEMENTARY MATERIAL S4
Miscarriage status in normal control group and RSA model group.

SUPPLEMENTARY MATERIAL S5
Bubble chart and bar chart of metabolic pathway KEGG enrichment analysis. **(A)** and **(C)** Bubble chart and bar chart of pathways with *P*-values less than 0.05. **(B)** and **(D)** Bubble chart and bar chart of top 20 pathways ranked by -lg(*P*-value). **(E)** and **(G)** Bubble chart and bar chart of pathways with *P*-values less than 0.05 and upregulated. **(F)** and **(H)** Bubble chart and bar chart of top 20 pathways ranked by -lg(*P*-value) and upregulated. **(I)** and **(J)** Bubble chart and bar chart of top 20 pathways ranked by -lg(*P*-value), downregulated and with *P*-values less than 0.05.

SUPPLEMENTARY MATERIAL S6
Summary of differential metabolites.

SUPPLEMENTARY MATERIAL S7
Heatmap of hierarchical clustering of the top 50 significant differential metabolites ranked by variable importance in projection.

SUPPLEMENTARY MATERIAL S8
Pie charts showing the classification of metabolites. **(A)** Pie chart of metabolite classes. **(B)** Pie chart of metabolite subclasses. **(C)** Pie chart of metabolite superclasses.

SUPPLEMENTARY MATERIAL S9
Data matrix of classification pie charts.

SUPPLEMENTARY MATERIAL S10
Summary table of metabolite pathways.

References

1. Chen X, Song QL, Ji R, Wang JY, Cao ML, Guo DY, et al. JPT2 affects trophoblast functions and macrophage polarization and metabolism, and acts as a potential therapeutic target for recurrent spontaneous abortion. *Adv Sci (Weinh)*. (2024) 11:e2306359. doi: 10.1002/advs.202306359

2. Deng T, Liao X, Zhu S. Recent advances in treatment of recurrent spontaneous abortion. *Obstet Gynecol Surv*. (2022) 77:355–66. doi: 10.1097/OGX.0000000000001033

3. de Assis V, Giugni CS, Ros ST. Evaluation of recurrent pregnancy loss. *Obstet Gynecol*. (2024) 143:645–59. doi: 10.1097/AOG.0000000000005498

4. Chen X, Song QL, Wang JY, Ji R, Cao ML, Guo DY, et al. FKBP5 regulates trophoblast-macrophage crosstalk in recurrent spontaneous abortion through PI3K/AKT and NF-κB signaling pathways. *Free Radic Biol Med*. (2023) 209:55–69. doi: 10.1016/j.freeradbiomed.2023.10.380

5. Yang M, Ong J, Meng F, Zhang F, Shen H, Kitt K, et al. Spatiotemporal insight into early pregnancy governed by immune-featured stromal cells. *Cell*. (2023) 186:4271–4288.e24. doi: 10.1016/j.cell.2023.08.020

6. Tang L, Xu X-H, Xu S, Liu Z, He Q, Li W, et al. Dysregulated Gln-Glu-α-ketoglutarate axis impairs maternal decidualization and increases the risk of recurrent spontaneous miscarriage. *Cell Rep Med*. (2023) 4:101026. doi: 10.1016/j.xcrm.2023.101026

7. Tang C, Hu W. The role of Th17 and Treg cells in normal pregnancy and unexplained recurrent spontaneous abortion (URSA): new insights into immune mechanisms. *Placenta*. (2023) 142:18–26. doi: 10.1016/j.placenta.2023.08.065

8. Yang Y, Liu H, Zhao Y, Geng C, Chao L, Hao A. Grim-19 deficiency promotes decidual macrophage autophagy in recurrent spontaneous abortion. *Front Endocrinol (Lausanne)*. (2022) 13:1023194. doi: 10.3389/fendo.2022.1023194

9. Chen M, Shi J-L, Zheng Z-M, Lin Z, Li M-Q, Shao J. An abnormal LPA/LPAR1-NHE1 axis leads to the autophagy deficiency of trophoblast cells in recurrent spontaneous abortion. *Reproduction*. (2023) 166:357–68. doi: 10.1530/REP-23-0224

10. Liu N, Shen H, Wang Z, Qin X, Li M, Zhang X. Autophagy inhibition in trophoblasts induces aberrant shift in CXCR4+ Decidual NK cell phenotype leading to pregnancy loss. *J Clin Med*. (2023) 12:7491. doi: 10.3390/jcm12237491

11. Wang P, Zhao C, Zhou H, Huang X, Ying H, Zhang S, et al. Dysregulation of histone deacetylases inhibits trophoblast growth during early placental development partially through TFEB-dependent autophagy-lysosomal pathway. *Int J Mol Sci*. (2023) 24:11899. doi: 10.3390/ijms241511899

12. Debnath J, Gammoh N, Ryan KM. Autophagy and autophagy-related pathways in cancer. *Nat Rev Mol Cell Biol*. (2023) 24:560–75. doi: 10.1038/s41580-023-00585-z

13. Zhang H, Zheng Y, Liu X, Zha X, Elsabagh M, Ma Y, et al. Autophagy attenuates placental apoptosis, oxidative stress and fetal growth restriction in pregnant ewes. *Environ Int*. (2023) 173:107806. doi: 10.1016/j.envint.2023.107806

14. Sun J, Yu M, Du W, Zhu S, Chen Z, Tao J, et al. The cGAS-STING pathway promotes the development of preeclampsia by upregulating autophagy: mechanisms and implications. *Int Immunopharmacol*. (2024) 128:111531. doi: 10.1016/j.intimp.2024.111531

15. Batiha GE-S, Shaheen HM, Elhawary EA, Mostafa NM, Eldahshan OA, Sabatier J-M. Phytochemical constituents, folk medicinal uses, and biological activities of genus Angelica: a review. *Molecules*. (2022) 28:267. doi: 10.3390/molecules28010267

16. Nai J, Zhang C, Shao H, Li B, Li H, Gao L, et al. Extraction, structure, pharmacological activities and drug carrier applications of Angelica sinensis polysaccharide. *Int J Biol Macromol*. (2021) 183:2337–53. doi: 10.1016/j.ijbiomac.2021.05.213

17. Xu Y, Wang X-C, Jiang W, Hu J-N. Angelica sinensis polysaccharides modified selenium nanoparticles for effective prevention of acute liver injury. *Int J Biol Macromol*. (2024) 263:130321. doi: 10.1016/j.ijbiomac.2024.130321

18. Luo L, Zhang H, Chen W, Zheng Z, He Z, Wang H, et al. Angelica sinensis polysaccharide ameliorates nonalcoholic fatty liver disease via restoring estrogen-related receptor α expression in liver. *Phytother Res*. (2023) 37:5407–17. doi: 10.1002/ptr.7982

19. Guo W, Wang W, Lei F, Zheng R, Zhao X, Gu Y, et al. Angelica sinensis polysaccharide combined with cisplatin reverses cisplatin resistance of ovarian cancer by inducing ferroptosis via regulating GPX4. *Biomed Pharmacother*. (2024) 175:116680. doi: 10.1016/j.biopha.2024.116680

20. Cai Y, Wang Y, Su W, Zhou X, Lu C. Angelica sinensis polysaccharide suppresses the Wnt/β-catenin-mediated malignant biological behaviors of breast cancer cells via the miR-3187-3p/PCDH10 axis. *Biochem Pharmacol*. (2024) 225:116295. doi: 10.1016/j.bcp.2024.116295

21. Ni S, Yi N, Yuan H, Li D, Chen X, Zhuang C. Angelica sinensis polysaccharide improves mitochondrial metabolism of osteoarthritis chondrocytes through PPARγ/SOD2/ROS pathways. *Phytother Res*. (2023) 37:5394–406. doi: 10.1002/ptr.7979

22. Liu H, Wei G, Wang T, Hou Y, Hou B, Li X, et al. Angelica keiskei water extract mitigates age-associated physiological decline in mice. *Redox Rep*. (2024) 29:2305036. doi: 10.1080/13510002.2024.2305036

23. Tuo W, Wang S, Shi Y, Cao W, Liu Y, Su Y, et al. Angelica sinensis polysaccharide extends lifespan and ameliorates aging-related diseases via insulin and TOR signaling pathways, and antioxidant ability in Drosophila. *Int J Biol Macromol*. (2023) 241:124639. doi: 10.1016/j.ijbiomac.2023.124639

24. Niu Y, Xiao H, Wang B, Wang Z, Du K, Wang Y, et al. Angelica sinensis polysaccharides alleviate the oxidative burden on hematopoietic cells by restoring 5-fluorouracil-induced oxidative damage in perivascular mesenchymal progenitor cells. *Pharm Biol*. (2023) 61:768–78. doi: 10.1080/13880209.2023.2207592

25. Xiao H, Wang Y, Wang Z, Wang B, Hu L, Hou J, et al. Angelica sinensis polysaccharides ameliorated 5-fluorouracil-induced damage of early B cell progenitors by alleviating oxidative stress of IL-7 producing mesenchymal stem and progenitor cells. *Biomed Pharmacother*. (2023) 167:115599. doi: 10.1016/j.biopha.2023.115599

26. Shi Y, Xiao L, Yin Y, Wei L. Ligustilide inhibits tumour necrosis factor-alpha-induced autophagy during C2C12 cells differentiation. *Biomed Pharmacother*. (2015) 69:42–6. doi: 10.1016/j.biopha.2014.11.002

27. Wei Q, He F, Rao J, Xiang X, Li L, Qi H. Targeting non-classical autophagy-dependent ferroptosis and the subsequent HMGB1/TfR1 feedback loop accounts for alleviating solar dermatitis by senkyunolide I. *Free Radic Biol Med*. (2024) 223:263–80. doi: 10.1016/j.freeradbiomed.2024.08.004

28. Bi S-J, Yue S-J, Bai X, Feng L-M, Xu D-Q, Fu R-J, et al. Danggui-Yimucao herb pair can protect mice from the immune imbalance caused by medical abortion and stabilize the level of serum metabolites. *Front Pharmacol*. (2021) 12:754125. doi: 10.3389/fphar.2021.754125

29. Bi S-J, Huang Y-X, Feng L-M, Yue S-J, Chen Y-Y, Fu R-J, et al. Network pharmacology-based study on immunomodulatory mechanism of danggui-yimucao herb pair for the treatment of RU486-induced abortion. *J Ethnopharmacol*. (2022) 282:114609. doi: 10.1016/j.jep.2021.114609

30. Wang J, Yang L. Effect of Angelica sinensis polysaccharides on AMPK signaling pathway and mitochondrial autophagy in kidney of diabetic nephropathy KK-ay mice. *Tradit Chin Med Herbs*. (2023) 54:3189–96. doi: 10.7501/j.issn.0253-2670.2023.10.016

31. Wang J. Mechanism of action of Angelica sinensis polysaccharide on mitochondrial autophagy in the kidneys of diabetic KK-ay mice through the AMPK pathway. *Gansu University of Chinese Medicine*. (2023). doi: 10.27026/d.cnki.ggszc.2023.000411

32. Zhang Y, Guo T, Huang L, He Z, Wang J, Mei H, et al. Protective effect of Angelica sinensis polysaccharide on pregnant rats suffering from iron deficiency anemia via regulation of the hepcidin-FPN1 axis. *Int J Biol Macromol*. (2024) 256:128016. doi: 10.1016/j.ijbiomac.2023.128016

33. Youssef A, Lashley L, Dieben S, Verburg H, Van Der Hoorn M-L. Defining recurrent pregnancy loss: associated factors and prognosis in couples with two versus three or more pregnancy losses. *Reprod Biomed Online*. (2020) 41:679–85. doi: 10.1016/j.rbmo.2020.05.016

34. The ESHRE Guideline Group on RPLBender Atik R, Christiansen OB, Elson J, Kolte AM, Lewis S, et al. ESHRE guideline: recurrent pregnancy loss. Hum Reprod Open. (2018) 2018:hoy004. doi: 10.1093/hropen/hoy004

35. Xu C, Ni S, Zhuang C, Li C, Zhao G, Jiang S, et al. Polysaccharide from Angelica sinensis attenuates SNP-induced apoptosis in osteoarthritis chondrocytes by inducing autophagy via the ERK1/2 pathway. *Arthritis Res Ther*. (2021) 23:47. doi: 10.1186/s13075-020-02409-3

36. Yan Hu ED. Intervention effect of Angelica sinensis polysaccharide on rats with recurrent sponta neous abortion and study on the mechanism. *Matern Child Health Care China*. (2022) 37:1293–7. doi: 10.19829/j.zgfybj.issn.1001-4411.2022.07.038

37. Tan H-X, Yang S-L, Li M-Q, Wang H-Y. Autophagy suppression of trophoblast cells induces pregnancy loss by activating decidual NK cytotoxicity and inhibiting trophoblast invasion. *Cell Commun Signal*. (2020) 18:73. doi: 10.1186/s12964-020-00579-w

38. Yang D, Ding J, Wang Y, Yuan M, Xian S, Zhang L, et al. YY1-PVT1 affects trophoblast invasion and adhesion by regulating mTOR pathway-mediated autophagy. *J Cell Physiol*. (2020) 235:6637–46. doi: 10.1002/jcp.29560

39. Li Z, Dai F, Zhu R, Zhang Y, Chen J, Chen L, et al. Dysregulation of CREB5 impairs Decidualization and maternal-fetal interactions by inhibiting autophagy in recurrent spontaneous abortion. *Reprod Sci*. (2024) 31:1983–2000. doi: 10.1007/s43032-024-01474-2

40. Wei A, Song Y, Ni T, Xiao H, Wan Y, Ren X, et al. Hyperoside attenuates pregnancy loss through activating autophagy and suppressing inflammation in a rat model. *Life Sci*. (2020) 254:117735. doi: 10.1016/j.lfs.2020.117735

41. Durgan J, Lystad AH, Sloan K, Carlsson SR, Wilson MI, Marcassa E, et al. Non-canonical autophagy drives alternative ATG8 conjugation to phosphatidylserine. *Mol Cell*. (2021) 81:2031–2040.e8. doi: 10.1016/j.molcel.2021.03.020

42. He D, Song Y, Xiao H, Shi S, Song H, Cui T, et al. Ligustilide enhances pregnancy outcomes via improvement of endometrial receptivity and promotion of endometrial angiogenesis in rats. *J Nat Med*. (2024) 78:42–52. doi: 10.1007/s11418-023-01739-1

43. Bai F-Y, Bi S-J, Yue S-J, Xu D-Q, Fu R-J, Sun Y, et al. The serum lipidomics reveal the action mechanism of Danggui-Yimucao herbal pair in abortion mice. *Biomed Chromatogr*. (2023) 37:e5717. doi: 10.1002/bmc.5717

44. Li C, Liu S, Zheng J, Xue Y. Angelica sinensis polysaccharide (ASP) attenuates diosbulbin-B (DB)-induced hepatotoxicity through activating the MEK/ERK pathway. *Bioengineered*. (2021) 12:3516–24. doi: 10.1080/21655979.2021.1950280

45. Sun C, Liu H, Chi B, Han J, Koga Y, Afshar K, et al. Improvement of idiopathic pulmonary fibrosis through a combination of Astragalus radix and Angelica sinensis radix via mammalian target of rapamycin signaling pathway-induced autophagy in rat. *J Thorac Dis*. (2024) 16:1397–411. doi: 10.21037/jtd-24-28

46. Long Y, Li D, Yu S, Shi A, Deng J, Wen J, et al. Medicine-food herb: Angelica sinensis, a potential therapeutic hope for Alzheimer's disease and related complications. *Food Funct*. (2022) 13:8783–803. doi: 10.1039/D2FO01287A

47. Bauermeister A, Mannochio-Russo H, Costa-Lotufo LV, Jarmusch AK, Dorrestein PC. Mass spectrometry-based metabolomics in microbiome investigations. *Nat Rev Microbiol*. (2022) 20:143–60. doi: 10.1038/s41579-021-00621-9

48. Klionsky DJ, Abeliovich H, Agostinis P, Agrawal DK, Aliev G, Askew DS, et al. Guidelines for the use and interpretation of assays for monitoring autophagy in higher eukaryotes. *Autophagy*. (2008) 4:151–75. doi: 10.4161/auto.5338

49. Yu G, Wang J, Liu Y, Luo T, Meng X, Zhang R, et al. Metabolic perturbations in pregnant rats exposed to low-dose perfluorooctanesulfonic acid: an integrated multi-omics analysis. *Environ Int*. (2023) 173:107851. doi: 10.1016/j.envint.2023.107851

50. Xu P, Zheng Y, Liao J, Hu M, Yang Y, Zhang B, et al. AMPK regulates homeostasis of invasion and viability in trophoblasts by redirecting glucose metabolism: implications for pre-eclampsia. *Cell Prolif*. (2023) 56:e13358. doi: 10.1111/cpr.13358

51. Batushansky A, Zacharia A, Shehadeh A, Bruck-Haimson R, Saidemberg D, Kogan NM, et al. A shift in Glycerolipid metabolism defines the follicular fluid of IVF patients with unexplained infertility. *Biomol Ther*. (2020) 10:1135. doi: 10.3390/biom10081135

52. Wang Y, Wu P, Huang Y, Ye Y, Yang X, Sun F, et al. BMI and lipidomic biomarkers with risk of gestational diabetes in pregnant women. *Obesity (Silver Spring)*. (2022) 30:2044–54. doi: 10.1002/oby.23517

53. Li X, Yin M, Gu J, Hou Y, Tian F, Sun F. Metabolomic profiling of plasma samples from women with recurrent spontaneous abortion. *Med Sci Monit*. (2018) 24:4038–45. doi: 10.12659/MSM.907653

54. Mercer GV, Stapleton D, Barrett C, Ringer LCM, Lambe S, Critch A, et al. Identifying placental metabolic biomarkers of preterm birth using nuclear magnetic resonance of intact tissue samples. *Placenta*. (2023) 143:80–6. doi: 10.1016/j.placenta.2023.10.006

55. Scott HD, Buchan M, Chadwick C, Field CJ, Letourneau N, Montina T, et al. Metabolic dysfunction in pregnancy: fingerprinting the maternal metabolome using proton nuclear magnetic resonance spectroscopy. *Endocrinol Diabetes Metab*. (2021) 4:e00201. doi: 10.1002/edm2.201

frontiers | Frontiers in Medicine

TYPE Review
PUBLISHED 13 March 2025
DOI 10.3389/fmed.2025.1539714

Check for updates

OPEN ACCESS

EDITED BY
HaiHui Huang,
Shaoguan University, China

REVIEWED BY
Medicharla Venkata Jagannadham,
University of Hyderabad, India
Jayanta Mondal,
University of Texas MD Anderson Cancer Center, United States

*CORRESPONDENCE
Milan Jakubek
✉ Milan.Jakubek@lf1.cuni.cz
Pavel Martásek
✉ Pavel.Martasek@lf1.cuni.cz

RECEIVED 04 December 2024
ACCEPTED 06 February 2025
PUBLISHED 13 March 2025

CITATION
Odehnalová N, Šandriková V, Hromadka R, Skaličková M, Dytrych P, Hoskovec D, Kejík Z, Hajduch J, Vellieux F, Vašáková MK, Martásek P and Jakubek M (2025) The potential of exosomes in regenerative medicine and in the diagnosis and therapies of neurodegenerative diseases and cancer. *Front. Med.* 12:1539714.
doi: 10.3389/fmed.2025.1539714

COPYRIGHT
© 2025 Odehnalová, Šandriková, Hromadka, Skaličková, Dytrych, Hoskovec, Kejík, Hajduch, Vellieux, Vašáková, Martásek and Jakubek. 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 potential of exosomes in regenerative medicine and in the diagnosis and therapies of neurodegenerative diseases and cancer

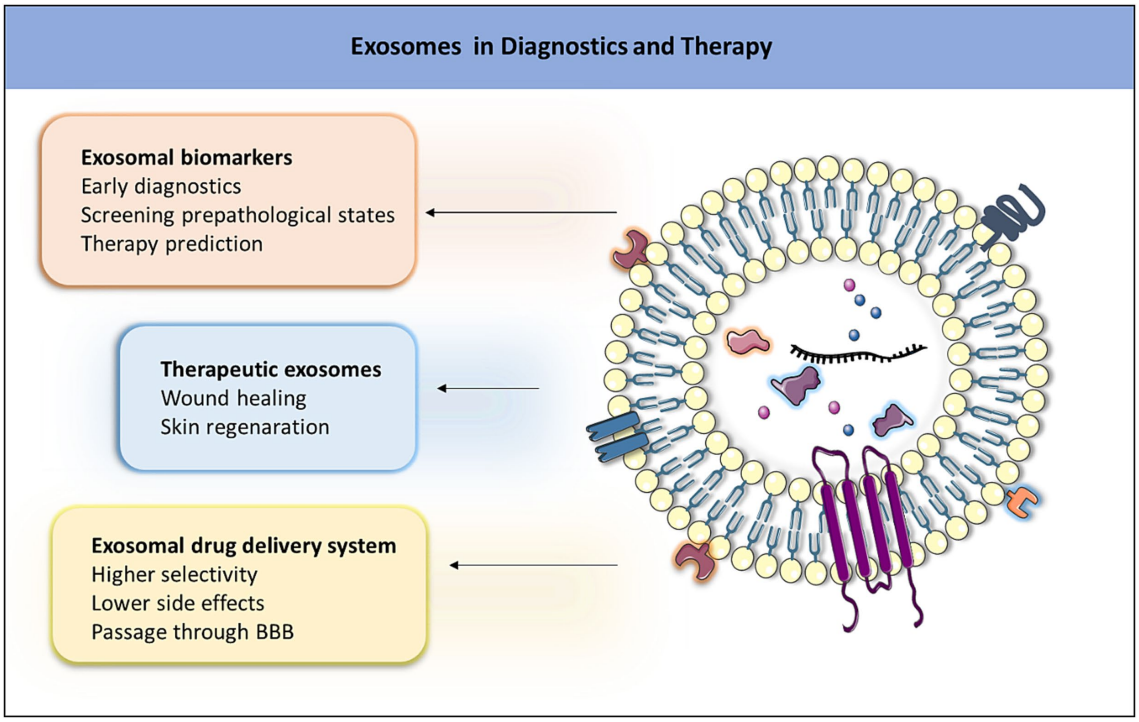
Nikola Odehnalová[1,2], Viera Šandriková[1], Róbert Hromadka[1,2], Markéta Skaličková[2,3], Petr Dytrych[4], David Hoskovec[4], Zdeněk Kejík[2,3,5], Jan Hajduch[2,3,6], Frédéric Vellieux[2,3], Martina Koziar Vašáková[7], Pavel Martásek[3]* and Milan Jakubek[2,3,5]*

[1]NEXARS Research and Development Center C2P s.r.o, Chlumec nad Cidlinou, Czechia, [2]BIOCEV, First Faculty of Medicine, Charles University, Vestec, Czechia, [3]Department of Paediatrics and Inherited Metabolic Disorders, First Faculty of Medicine, Charles University and General University Hospital, Prague, Czechia, [4]Department of Surgery-Department of Abdominal, Thoracic Surgery and Traumatology, First Faculty of Medicine, Charles University and General University Hospital, Prague, Czechia, [5]Department of Analytical Chemistry, University of Chemistry and Technology, Prague, Czechia, [6]The Department of Chemistry of Natural Compounds, University of Chemistry and Technology, Prague, Czechia, [7]Department of Respiratory Medicine, First Faculty of Medicine, Charles University and Thomayer Hospital, Prague, Czechia

Exosomes, nanosized extracellular vesicles released by various cell types, are intensively studied for the diagnosis and treatment of cancer and neurodegenerative diseases, and they also display high usability in regenerative medicine. Emphasizing their diagnostic potential, exosomes serve as carriers of disease-specific biomarkers, enabling non-invasive early detection and personalized medicine. The cargo loading of exosomes with therapeutic agents presents an innovative strategy for targeted drug delivery, minimizing off-target effects and optimizing therapeutic interventions. In regenerative medicine, exosomes play a crucial role in intercellular communication, facilitating tissue regeneration through the transmission of bioactive molecules. While acknowledging existing challenges in standardization and scalability, ongoing research efforts aim to refine methodologies and address regulatory considerations. In summary, this review underscores the transformative potential of exosomes in reshaping the landscape of medical interventions, with a particular emphasis on cancer, neurodegenerative diseases, and regenerative medicine.

KEYWORDS

exosomes, regenerative medicine, biomarker, cancer, neurodegenerative diseases



GRAPHICAL ABSTRACT
The figure was partly generated using Servier Medical Art, provided by Servier, licensed under a Creative Commons Attribution 3.0 unported license.

1 Introduction

Extracellular vehicles (EVs) are cellular structures released by cells into the extracellular space and have recently become a focal point of research due to their multifunctional role in many biological processes (1, 2). According to the new classification, EVs are divided into several types based on their biogenesis (e.g., exosome, microvesicle, apoptosome, and autophagic EVs), concept (e.g., oncosome, matrix vesicle, stress EVs, and migrasome), and size (e.g., small EVs and large EVs) (3). In general, EVs were divided into three types including apoptotic bodies (apoptosomes), which are the largest EVs with a size range between 1 and 5 μm and are released during programmed cell death (apoptosis). They contain cellular organelles and fragmented DNA and are cleared by phagocytic cells. Microvesicles, typically 100–1,000 nm in diameter, are shed directly from the plasma membrane through outward budding, a process in which a portion of the cellular membrane protrudes outward from the cell surface. They contain proteins, lipids, polysaccharides, and nucleic acids and are involved in intercellular communication and signaling (2, 3). Exosomes (Figure 1), the smallest EVs ranging in size from 30 to 150 nm, are lipid bilayer vesicles and were discovered three decades ago by Pan and Johnstone during investigations of reticulocyte maturation (2–5). Recent studies discovered small exosomes (Exo-S) and large exosomes (Exo-L). Exo-S are in the size range 40–80 nm and contain exosomal tetraspanin marker CD63, while Exo-L (80–150 nm) contain CD9 (3). Initially perceived as cellular waste products responsible for eliminating unnecessary cellular components, our understanding of exosomes has undergone a paradigm shift over the years, revealing their multifaceted functions in cellular communication and signaling (1, 2).

Exosomes not only represent a promising material for the diagnosis of serious pathological states but can also be effectively utilized for medicinal applications and drug transport. Given their significance, this review presents the medicinal potential of exosomes, encompassing areas such as regenerative medicine, early diagnosis, and drug treatment. In addition, to provide a more comprehensive understanding, the review rigorously assesses exosome biogenesis, isolation, and characterization.

2 Biogenesis of exosomes

Understanding the biogenesis of exosomes is crucial for advancing the knowledge of their biological functions, their roles in diseases, and potential applications in therapeutics (1, 3). Their life cycle is a complex process involving three main steps: biogenesis, transport, and release (2, 3). The whole process is illustrated in Figure 2 and is initiated by an inward cell membrane budding (6). During this invagination of the plasma membrane, a portion of a cellular membrane undergoes inward folding, and a cup-shaped structure containing extracellular proteins, lipids, metabolites, and cell membrane proteins is formed. This leads to the formation of early endosomes (EEs), which subsequently mature and transform into late endosomes (LEs). Maturation involves the inward budding of the EEs membrane, leading to the sequestration of EE cytoplasmic contents and intraluminal vesicle (ILVs) formation within the endosomal lumen. During ILV formation, specific cargo molecules such as proteins, lipids, and nucleic acids are selectively sorted into the ILVs. LEs containing ILVs are called multi-vesicular bodies (MVBs). The fate of MVBs is determined by the specific proteins present on their surface, which in turn influence various intracellular pathways involved in cargo sorting and trafficking. The MVBs can either fuse with lysosomes or autophagosomes to be degraded or fuse with a plasma membrane to release the contained ILVs as exosomes (2, 6).

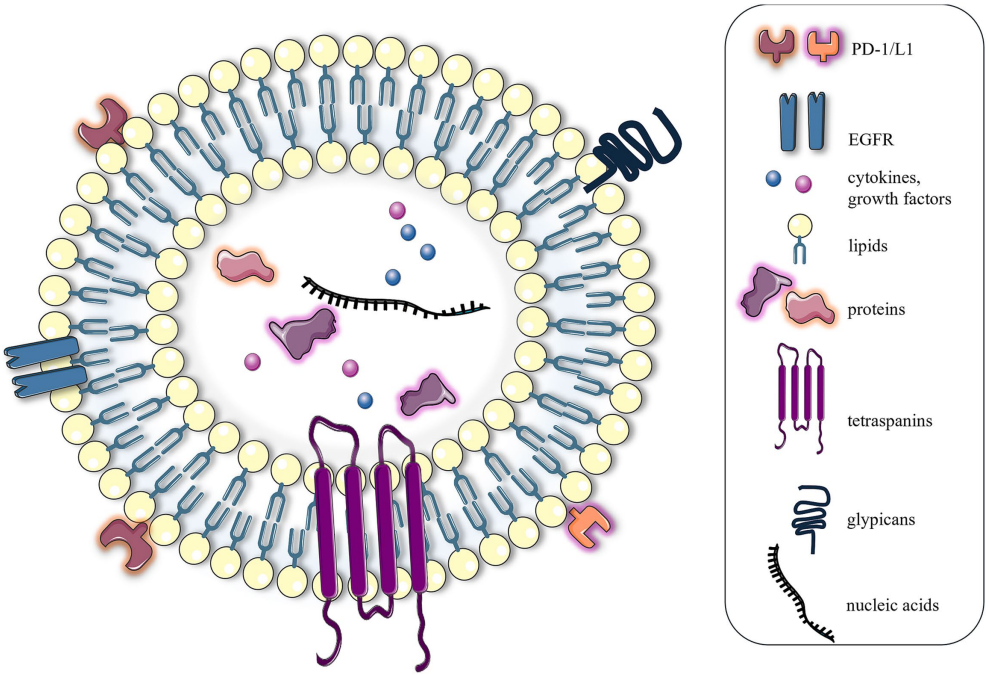


FIGURE 1
Structure of exosomes. From a structural perspective, exosomes can be defined as lipid nanoparticles characterized by a phospholipid bilayer membrane. The exosomal membrane is enriched with a diverse array of proteins and saccharide markers, including immunomodulatory molecules, such as PD-1 and PD-L1, and hormone receptors, such as EGFR, tetraspanins, and glypicans. The internal composition of exosomes comprises a variety of biomolecules, including intracellular and cytoskeletal proteins, nucleic acids, growth factors, and cytokines. The figure was partly generated using Servier Medical Art, provided by Servier, licensed under a Creative Commons Attribution 3.0 unported license.

3 Isolation and characterization methods of exosomes

Isolation of exosomes is challenging due to the complexity of biological fluids (7, 8). The most common isolation methods include ultracentrifugation, ultrafiltration, size exclusion chromatography, polymer precipitation, tangential flow filtration, and immunoaffinity approaches. A comparison of these methods is provided in Table 1. The optimal isolation strategy should be selected based on the application field, as well as the volume and number of biological samples (7–9). Furthermore, the application of exosomes directly influences the use of subsequent characterization methods (9, 10). This chapter does not aim to provide a detailed explanation of the principles behind individual isolation and characterization methods of exosomes but rather to offer a comprehensive overview that assists researchers in selecting the most appropriate approaches based on considering the specific requirements of different experimental and application contexts.

For diagnostic purposes, exosomes from various biological fluids, including blood (serum and plasma), saliva, and urine, are used in non-invasive diagnostics due to their ability to carry specific molecular information (9, 11). Blood-derived exosomes provide systemic insights, reflecting the physiological and pathological state of the entire body, making them ideal for detecting various diseases such as cancer, neurodegenerative, cardiovascular, and autoimmune diseases (12–16). Salivary exosomes, primarily originating from the salivary glands and the oral cavity, are valuable for diagnosing oral diseases and gastrointestinal tract disorders (17–19). Urinary exosomes, secreted by epithelial cells of the urinary tract, are effective in diagnosing renal diseases, bladder cancer, and prostate conditions (20–22).

When utilizing exosomes for diagnostic purposes, the primary goal is to identify disease-specific exosomal markers. To determine disease-specific markers, it is essential to identify markers that differ in presence or expression level between samples from diseased and healthy patients. This requires the comparison and processing of a large number (tens to hundreds) of biological samples for statistical relevance (23, 24). This procedure thus involves processing large numbers of samples with small volume, typically in the maximum of a few milliliters. For limited numbers of samples, typically in the range of lower tens, differential ultracentrifugation (DUC) is considered the gold standard and is the most widely used method. It usually involves several consecutive rounds of centrifugation with increasing centrifugal force and centrifugation time to remove cells, cell debris, and larger microvesicles. The final step at 100,000 g or higher serves to precipitate the exosomes (12, 25–27).

Because DUC is time-consuming, for handling tens to hundreds of samples, the final ultracentrifugation step is often replaced with commercial kits based on precipitation. Several commercial kits use a polyethylene glycol precipitation technology, including the Total Exosome Isolation Kit (Invitrogen), ExoQuick-TC Exosome Precipitation Solution (System Biosciences), miRCURY Exosome Kits (QIAGEN), Exo-Prep (HansaBioMed), PureExo Exosome Isolation Kit (101Bio), ExoGAG (NasasBiotech), Exosome Precipitation Solutions (Immunostep), and the miRCURY Exosome Isolation Kit (Exiqon) (28–32). These methods result in the isolation of exosomes referred to as total exosomes. In neurodegenerative diseases (NDs), specific subpopulations of exosomes are isolated from the total

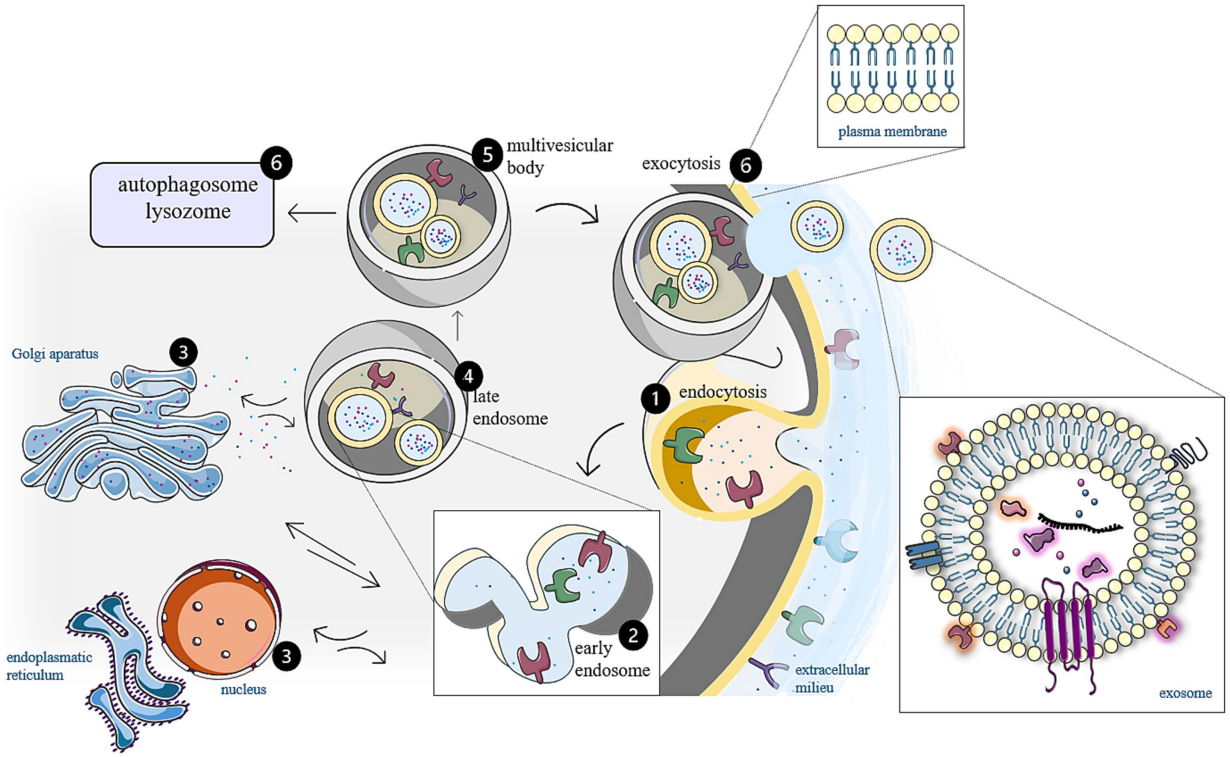


FIGURE 2
Biogenesis of exosomes. Extracellular membranes are characterized by the presence of numerous transmembrane proteins, including various receptors. (1) Upon ligand binding, receptor-mediated endocytosis is initiated, facilitated by actin filaments, which are integral components of the cytoskeleton. This process results in the invagination of the membrane surrounding the receptor, leading to the formation of an early endosome within the cell. (2) In this early endosome, the bilayer phospholipid membrane exhibits an orientation that is opposite to that of the cytoplasmic membrane, causing the extracellular domains of the transmembrane proteins to be directed inward, toward the lumen of the endosome. (3) Within the cellular context, endosomes are integral components of the complex endosomal-lysosomal system that interacts in a parallel manner (in both directions) with various organelles (e.g., Golgi apparatus and endoplasmic reticulum). This interaction allows endosomes and exosomes to encapsulate biomolecules derived from diverse cellular compartment. (4) During this process, early endosomes undergo maturation into late endosomes, which possess the normal orientation of transmembrane proteins and (5) generate intraluminal vesicles that will ultimately become exosomes. (6) Following the fusion of multi-vesicular bodies containing these intraluminal vesicles with the cytoplasmic membrane, exosomes are released into the extracellular environment. The figure was partly generated using Servier Medical Art, provided by Servier, licensed under a Creative Commons Attribution 3.0 unported license.

exosome pool using immunoaffinity methods with specific antibodies, such as anti-L1CAM, anti-NCAM, anti-MOG, or anti-GLAST. These antibodies bind specifically neural, oligodendrocyte, or astrocyte exosomes, which are the most relevant for identifying markers of NDs as Alzheimer disease (AD), multiple sclerosis (MS), dementia, or schizophrenia (33–36).

After the isolation of total exosomes or specific exosomal subpopulations, the characterization of these exosomes and the identification of specific disease markers are performed (10). The most common markers include specific proteins or RNAs. For protein marker identification, techniques such as tandem of liquid chromatography and mass spectrometry (LC–MS) are used, while next-generation sequencing (NGS) is used for RNA. Once specific disease markers are identified, it is crucial to determine whether their exosomal expression differs significantly between diseased and healthy patients in a statistically relevant manner. To quantify markers, techniques such as enzyme-linked immunosorbent assay (ELISA) or quantitative reverse transcription polymerase chain reaction (qRT-PCR) are commonly used (37–42).

Currently, exosomes are being extensively investigated as drug delivery systems. Due to their nanoscale dimensions, they can deliver therapeutic agents specifically to tumor sites (43). Tumor tissues are characterized by a high absorption capacity and poor drainage (44), which facilitates the selective accumulation of exosomes with prolonged circulation times within these tissues. In addition, it is noteworthy that exosomes exhibit favorable permeability across BBB (45). Furthermore, the surface of exosomes can be effectively modified with various ligands that are specific to target cells, thereby significantly enhancing the selectivity of exosomes for these cells (46). As cell-derived products, exosomes are generally considered to be safer than conventional nanoparticles, particularly metal-based nanoparticles (47). Their unique structure, comprising a hydrophilic core and a lipid bilayer, allows them to transport a wide range of drug types (46). However, a significant limitation in the clinical application of exosomes is that cells produce only small quantities of various types of exosomes, which restricts their usability in therapeutic contexts (48).

In regenerative medicine and therapeutic applications, exosomes are isolated from cell culture media. The most commonly used cell sources are mesenchymal stem cells. Similar to diagnostic applications, where DUC is a preferred method for smaller sample quantities, DUC is also commonly employed for isolating exosomes from cell culture media for volumes ranging from tens to hundreds of milliliters. However, a drawback of the final ultracentrifugation step is that the

TABLE 1 Advantages and disadvantages of the most common isolation techniques of exosomes.

Isolation technique	Advantages	Disadvantages	References
Ultracentrifugation	Low cost, separation of large volumes, low operating expenses, compatibility with a wide range of samples	Time consumption, low purity, inappropriateness for small volumes, diminishing biological activity of exosomes, high equipment cost	(8, 25, 26, 216, 217)
Ultrafiltration	Simplicity, fast, absence of special equipment	Potential deformation of exosomes, moderate purity, loss of exosomes, particularly problematic for isolating from small volumes, reduced purification efficiency due to clogging of membrane pores	(218–221)
Size exclusion chromatography	High purity, scalable, cost-effective for large-scale processing, availability of commercial kits, preservation of biological activity	Requirement of dedicated equipment, low yield, target product dilution	(25, 218, 220, 222, 223)
Precipitation techniques	High morphological and functional quality of exosomes, fast, simplicity, compatibility with low sample volumes, availability of commercial kits	Low purity, contamination of precipitating agent, not suitable for large sample volumes	(28, 29, 221, 224, 225)
Immunoaffinity techniques	High purity of specific exosomes	Loss of exosomes with lower expression levels, challenge separating exosomes from the bound antibodies, low capacity, low yields	(216, 226)
Tangential flow filtration	Processing of large sample volumes, high recovery rate of exosomes with minimal loss, preservation of exosome integrity, simultaneous concentration and buffer exchange (diafiltration), scalable for industrial production, reduction of processing time	High initial cost of equipment and consumables, require pre-filtration steps to remove large particles or debris, potential risk of membrane fouling leading to reduced efficiency over time	(52, 53, 55, 56)

high speeds lead to exosomal damage and the sedimentation of impurities, which diminishes exosomal therapeutic activity (12, 49, 50). Therefore, the final ultracentrifugation step is often replaced by density gradient ultracentrifugation (DGUC), utilizing sucrose or iodixanol gradients (OptiPrep™), where the exosomal fraction is not only less damaged but also better purified (49–51).

For larger volumes, in the range of hundreds of milliliters to liters, a combination of other techniques is employed due to the limited capacity of centrifuges. These primarily include ultrafiltration (UF), utilizing polymer filters of various pore sizes to remove cells, cell debris, and microvesicles, followed by tangential flow filtration (TFF) to remove contaminating proteins, to concentrate the sample, and to perform diafiltration of exosomes into the desired buffer (52–54). For final exosome purification, size exclusion chromatography (SEC) is employed, followed by a final TFF step to concentrate the sample and transfer it into suitable application buffers, most commonly PBS (55–57). The advantage of the combination of techniques such as UF, TFF, and SEC is that they are suitable for practical use on an industrial scale (55, 56). In addition, the isolated exosomes exhibit high purity and preserved therapeutic activity, which is usually subsequently confirmed by *in vitro* tests such as scratch or transwell assays (54, 58).

According to the recommendations of the International Society for Extracellular Vesicles (ISEV), exosomal samples should generally be characterized for size, concentration, and the presence of exosomes (59, 60). The presence of exosomes is typically confirmed by detecting at least one transmembrane protein (commonly tetraspanins: CD9, CD63, and CD81) or a GPI-anchored protein (e.g., integrins), along with one cytoplasmic lipid (e.g., sphingolipids, ceramides, and cholesterol) or cytoplasmic protein (e.g., ALIX, TSG101, and HSP70) (59, 60). These specific markers can be assayed using methods such as Western blot or ELISA (61, 62). In addition, electron microscopy is often employed to provide images of typical exosomal morphology, further confirming the presence of exosomes in the sample (63, 64).

The size and concentration of exosomes, expressed as particles per milliliter, are commonly determined using techniques such as dynamic light scattering (DLS), nanoparticle tracking analysis (NTA), and resistive pulse sensing (RPS). However, these techniques are not specific to exosomes and may overestimate their concentration (65, 66). Alternatively, exosomes can be labeled with specific dyes, where the measured concentration corresponds only to the positively stained population. This approach allows for more accurate quantification of exosomes and can be performed using nanoflow cytometry (nanoFCM). Nevertheless, underestimation of concentration may occur if the sample is not properly titrated (67, 68). In some cases, the concentration of exosomes is determined based on the total protein concentration per milliliter (μg/mL) using the Bradford assay (10). All these parameters must be thoroughly assessed to ensure sufficient sample purity and demonstrate that the observed therapeutic effect is primarily induced by exosomes rather than by potential contaminants.

4 Therapeutic applications of exosomes in biomedicine

As mentioned earlier, exosomes are natural intercellular communicators in normal biological processes but also in pathologies. They transport proteins, lipids, and nucleic acids specific for their parenteral pathogenic cells. From a clinical perspective, most applications use exosomes as biomarkers of diseases (69). The content of the exosome has been shown to be disease-specific, such as in NDs, prion diseases, viral infections, and cancer (69). Furthermore, exosomes play a transformative role in regenerative medicine, offering

innovative therapeutic interventions. Their bioactive cargo, including growth factors and signaling molecules, has demonstrated significant potential in modulating immune responses and promoting tissue repair (70). In addition, exosomes serve as carriers for therapeutic cargo loading, holding promise for targeted drug delivery in disease treatment (69, 71, 72). Currently, 150 clinical trials registered on ClinicalTrials.gov are investigating exosome-based therapies for various diseases (73). The majority of applications utilizing exosomes for both therapy and diagnosis focus on their utilization in the fields of cancer and NDs, as depicted in Figure 3 (74). For this reason, this review will focus on the use of exosomes in these diseases.

4.1 Exosomes in regenerative medicine

Nowadays, one of the main applications of exosomes is in regenerative medicine, a promising field dedicated to the regeneration and reconstruction of diseased or injured organs and tissues (72). Since the 1970s, mesenchymal stem cells (MSCs) have been under investigation in this field for their multipotent characteristics and their ability to migrate to injury sites. They are used in regenerative medicine due to their robust self-renewal capacity and ability to differentiate into adipogenic, chondrogenic, osteogenic, endothelial, neural, and epithelial cells, as proven in both *in vivo* and *in vitro* experiments (70, 73–75). MSCs, like every cell in the human body, release exosomes, which have started to be extensively researched due to their regenerative properties. Currently, at least 31 clinical trials are exploring the use of exosomes derived from MSCs (MSCs-EXOs) as an alternative to basic MSCs therapy (73). MSCs-EXOs have shown comparable or superior therapeutic efficacy compared to MSCs alone (72, 73). They exhibit lower immunogenicity, present an enhanced safety profile by avoiding concerns related to uncontrolled differentiation and tumorigenicity associated with live cells, and offer improved storage conditions, simplifying logistical challenges compared to live cell storage. In addition, there are no ethical issues, and their small size allows for sterilization by filtration (73–77). The comparison between exosomal and stem cell therapy is provided in Table 2.

The greatest attention in regenerative medicine is focused on skin healing, a tissue which plays a crucial multifunctional role as a protective barrier, a temperature regulator, and facilitating tactile and pain sensations (78). However, the use of exosomes for wound treatment could be challenging due to their rapid clearance from the

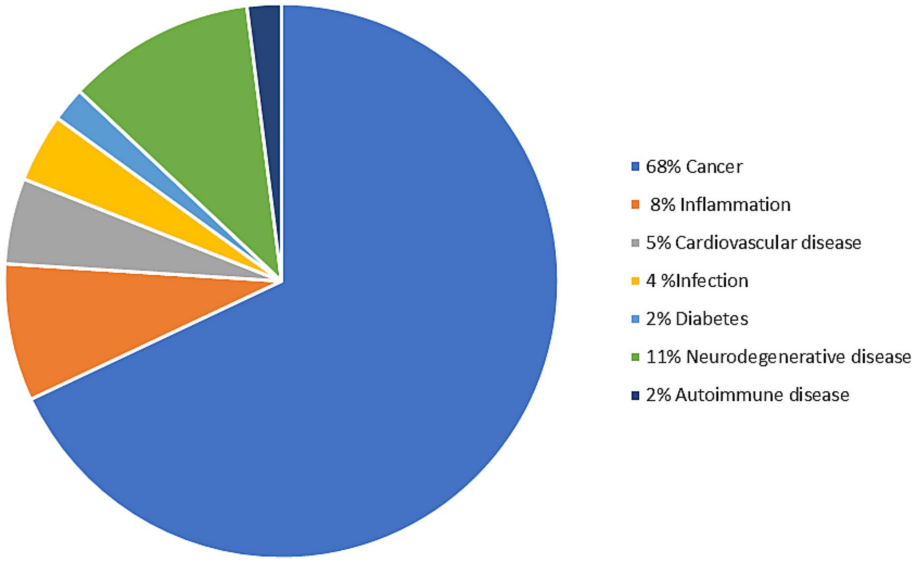


FIGURE 3
Distribution of exosome therapy and diagnostics concerning the target diseases (74). The figure was partly generated using Servier Medical Art, provided by Servier, licensed under a Creative Commons Attribution 3.0 unported license.

TABLE 2 Exosomes versus stem cells in preparation and therapy.

	Advantages	Disadvantages
MSC preparation	Ease to isolate, easy to expand at a large scale, highly proliferative, well established FDA guidelines	Harsh storage and transportation conditions
Exosome preparation	Small size, stable upon freezing and thawing	Difficult to isolate and purify, no established regulations and standards
Therapeutic application of MSCs	Multilineage differentiation potential, extensive preclinical and clinical studies	Immunogenicity, oncological complications, fusion toxicity, ethical issues
Therapeutic application of exosomes	Minimal risk of immune responses and tumor formation, no ethical issues, multiple delivery routes, can be engineered to specifically target and deliver drug cargoes	Rapid clearance from blood after administration

application site, which limited therapeutic effectiveness (79). To address this limitation, current research focuses on the combination of exosomes and biomaterials. This innovative approach extends the retention time of exosomes on the wound surface without compromising their biological activity, enabling the development of advanced exosome-based therapies (80, 81). Hydrogels as biomaterials exhibit a synergistic effect in exosome-induced wound healing and can serve as a versatile platform for the incorporation of therapeutic exosomes, enhancing their efficacy in tissue regeneration. To date, hyaluronic acid, gelatin, chitosan, and polypeptide-based hydrogels have been used for encapsulating exosomes from different cell sources (75, 80, 81). Chitosan is the often-used material for hydrogel preparation. Chitosan hydrogel enriched with MSCs-EXOs, specifically human endometrial stem cell-derived and human placenta-derived exosomes, demonstrated notable wound closure ability by promoting the formation of new epithelial cells, significant retention of MSCs-EXOs at injury sites, promotion of angiogenesis, and acceleration of the recovery of ischemic hind limbs (82). For the best contact of the hydrogel with the wounded skin and effective wound filling, thermosensitive hydrogels are used. Thermosensitive pluronic hydrogel combined with human umbilical cord MSCs-EXOs (HUCMSCs-EXOs) significantly accelerated wound closure, promoted angiogenesis, and improved skin healing of chronic diabetic wounds (83). For effective diabetic wound treatment, polyvinyl alcohol/alginate nanohydrogel with HUCMSCs-EXOs and an injectable antibacterial polypeptide-based hydrogel with adipose derived MSCs-EXOs were used. This type of hydrogels demonstrated the ability to promote proliferation, migration, and angiogenesis of human umbilical vein endothelial cells, expediting diabetic wound closure thus presenting a novel approach for complete skin regeneration (84, 85).

Further applications in regenerative medicine are focused on hard tissue regeneration, essential for bone and cartilage repair. Traditionally, MSCs, scaffolds, and growth factors are used in this field. While scaffolds have been proven beneficial for bone regeneration, the avascular nature of cartilage poses unique challenges (75). Osteoarthritis (OA), a prevalent joint disease extending beyond cartilage, demands innovative regenerative procedures (81). Promisingly, exosome-integrated scaffolds and MSCs-EXO therapy show potential in OA treatment. Articular cavity injection with HUCMSCs-EXOs in PBS demonstrated significant efficacy in preventing severe damage to knee articular cartilage in a rat OA model. These therapies not only promoted chondrocyte proliferation and migration but also exhibited anti-apoptotic effects and reversed cellular injuries. Moreover, HUCMSCs-EXOS played a crucial role in regulating the polarization of M2 macrophages, fostering chondrocyte survival by producing anti-inflammatory cytokines to suppress adverse inflammation (86). To enhance therapeutic efficiency and retention time *in vivo*, HUCMSCs-EXOs were engineered to specifically target chondrocytes and encapsulated within hyaluronic acid hydrogel, presenting a "two-phase" releasing system. This approach synergistically facilitated OA cartilage repair in a rat model and proved the rejuvenating effects of HUCMSCs-EXOs on aging chondrocytes in OA, offering a promising cell-free OA treatment strategy (87). Therapeutic potential of bone marrow MSCs-EXOs was explored in the context of mitochondrial dysfunction and oxidative stress in OA. 3D printed scaffolds, composed of extracellular matrix, gelatin methacrylate, and exosomes, effectively restored chondrocyte mitochondrial function, enhanced chondrocyte migration, and polarized the synovial macrophage response *in vitro*. Notably, a 3D printed scaffold significantly facilitated cartilage regeneration in a rabbit model, highlighting its potential as an early treatment strategy for OA (88).

In recent years, MSCs-EXOs have also been used increasingly in ophthalmology. MSCs-EXOs have shown promise in various applications, such as promoting ocular tissue regeneration and addressing vision-related disorders. MSCs-EXOs have explored the therapeutic potential of stem cell exosomes in treating ocular surface diseases, corneal injuries, and retinal degenerative conditions (89). Exosomes derived from bone marrow stem cells have demonstrated the ability to enhance corneal epithelialization and maintain corneal transparency in diabetic mice (90). MSCs-EXOs have been explored for their regenerative effects on corneal injuries (91). In retinal diseases, including age-related macular degeneration and retinitis pigmentosa, MSCs-EXOs have been investigated for their neuroprotective and regenerative properties. These exosomes may influence retinal cell survival, angiogenesis, and anti-inflammatory responses, offering a novel approach for treating degenerative conditions affecting the retina (92). As exosome research in regenerative medicine continues to progress, MSCs-EXOs are beginning to be applied into various fields, such as periodontitis (93, 94). Research on exosomes in regenerative medicine is expected to gain prominence, highlighting the growing need to integrate exosome-based regenerative therapies into clinical practice (95).

4.2 Exosomes in early diagnosis

As already mentioned, exosomes are promising biomarkers for the diagnosis of several diseases. Most attention has been paid to cancer (Figure 4; Supplementary Table S1) (95–105) with less focus on NDs.

Tumor-derived exosomes (TEXs) have emerged as critical players in cancer progression. Cancer cells release TEXs in large quantities, leading to a rapid increase in the concentration of total exosomes in the serum or plasma of cancer patients, which correlates with poor prognosis (106). TEXs provide immunosuppressive effects by inducing dysfunction in various immune cells, thereby suppressing anti-tumor immune responses (107). Initial interactions between TEXs and immune cells occur through ligand–receptor recognition, followed by either direct fusion with the plasma membrane or receptor-mediated uptake, resulting in the release of TEX cargo into the cytoplasm of immune cells. While T lymphocytes do not efficiently internalize TEXs, they interact with surface molecules to trigger sustained Ca^{2+} flux and downstream signaling cascades, ultimately altering the transcriptome of recipient cells. In contrast, phagocytic cells, such as dendritic cells and macrophages, rapidly internalize TEXs (108).

Macrophages are a prominent component of the tumor microenvironment (TME) and can make up more than half of the tumor mass in some cases. Moreover, tumor progression is associated mainly with macrophages. Macrophages can be induced to either tumor-suppressive immunological type M1 or tumor-promoting inflammatory M2 macrophages (109, 110). Tumor-associated macrophages predominantly exhibit an M2 phenotype, characterized by the secretion of pro-angiogenic factors and cytokines that promote angiogenesis, tumor growth, and metastasis (111, 112). Many studies

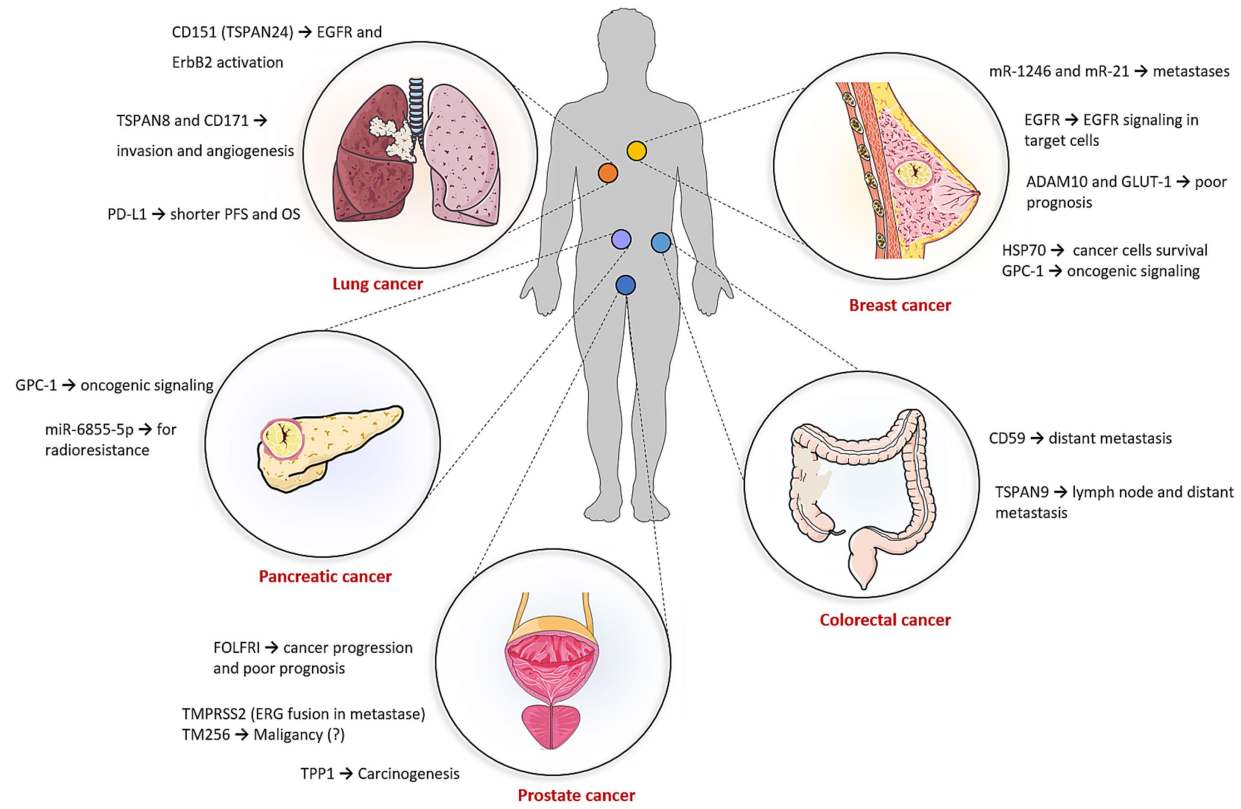


FIGURE 4
Potential of exosomes in cancer diagnostics. The applicability of exosomes in cancer diagnostics is detailed in Supplementary Table S1. The figure was partly generated using Servier Medical Art, provided by Servier, licensed under a Creative Commons Attribution 3.0 unported license.

have shown evidence on the role of TEXs in macrophage M2 polarization to promote tumor progression. For instance, exposure of macrophages to TEXs decreased the expression of M1 markers, such as IFNγ, while upregulating IL-1β, a marker of inflammation, suggesting that TEXs may help maintain a macrophage phenotype supportive of tumor survival and proliferation (113). In triple-negative breast cancer (TNBC), TEXs have been reported to promote M2 macrophage polarization, facilitating lymph node metastasis. Co-culture of TNBC-TEXs with macrophages led to significant morphological changes and an increased expression of M2 markers, including Fizz1, CD206, and Arg-1. An orthotopic TNBC model further confirmed the role of TEXs in driving M2 polarization, with enhanced tumor growth and axillary lymph node metastasis observed *in vivo* (114).

In cancer metastasis, epithelial–mesenchymal transition (EMT) is the key process during which cancer cells lose their epithelial properties and adopt a more mesenchymal and invasive phenotype. Generally, EMT initiation is characterized by loss of cell–cell adhesions and apicobasal polarity, leading to the formation of cells with increased migratory and invasion capabilities that are able to invade the extracellular matrix (115, 116). At a molecular level, EMT involves the downregulation of epithelial-type proteins, such as E-cadherin, and the acquisition of mesenchymal markers, such as vimentin (117). TEXs from bladder cancer cells have been shown to induce EMT-like changes in urothelial cells, enhancing their invasive potential. This effect was mediated by TEX-induced upregulation of vimentin and downregulation of E-cadherin through the TGF-β1 signaling pathway (118). Similar EMT-inducing effects of TEXs have been observed in glioblastoma, lung carcinoma, and gastric cancer models (119–121). TEXs can also influence EMT through their miRNA cargo. For example, miR-23a within TEXs promotes EMT by inhibiting E-cadherin synthesis in lung carcinoma and melanoma cells, while miR-191 and let-7a, present in TEXs from patients with melanoma, gastric, and colorectal cancers, have also been implicated in EMT regulation (122–126). In addition, TEX-derived miR-105 has been shown to promote vascular invasion by downregulating ZO-1 in endothelial cells. Notably, elevated miR-105 levels in the serum of breast cancer patients have been correlated with metastatic progression and poor prognosis (127). Given their critical role in modulating the TME, promoting EMT, and facilitating immune evasion, profiling TEXs in blood and other body fluids holds significant promise as a non-invasive method for cancer diagnosis and prognosis (128).

The WHO Global Cancer Observatory (GLOBOCAN) 2022 registry provides a list of the most common types of cancers. The three most common types are lung (12.4%), breast (11.5%), and colorectal (9.6%) cancers. Lung cancer is the leading cause of cancer death worldwide (129). Surface-enhanced Raman spectroscopy (SERS) of exosomes, combined with AI deep learning software, allows for accurate diagnosis of early-stage lung cancer. The deep learning model was trained with SERS signals of exosomes derived from normal and lung cancer cell lines and subsequently its ability to detect cancer was verified using exosome samples from patients' blood. The model identified the lung cancer patients and even detected stage I patients with an accuracy of 90.7% (130). Regarding TEX protein biomarkers,

CD151, TSPAN8, and CD171 were overexpressed in lung cancer samples. Of these, CD171 has been associated with EMT, metastases, and poor prognosis (96, 131, 132). Determination of PD-L1 can also provide important diagnostic information. Akbar et al. reported that exoPD-L1 from non-small cell lung cancer (NSCLC) patients and healthy controls showed a significantly higher difference than corresponding serum and tissue PD-L1 (97). For example, exoPD-L1 was found in all the patients (100%), while tissue PD-L1 was observed only in 71% patients. Furthermore, exoPD-L1 can also be used for the prediction of immune inhibitors efficiency. ROC curve analysis of change from baseline in exoPD-L1 levels between responders and non-responders showed 87% sensitivity and 100% specificity ($p = 0.0015$), indicating strong discriminatory power. An increasing number of studies have demonstrated that non-coding RNAs are closely correlated with the initiation and development of lung cancer as well (133). Detection of exosomal long non-coding RNA as potential lung cancer diagnosis is also performed in a current clinical study (NCT03830619) (134). Breast cancer is the most common type of cancer and the second leading cause of cancer-related deaths in women. In this type of cancer, media from breast cancer cell lines were analyzed to identify specific exosomal proteins such as glucose transporter-1, glypican-1, and the metalloproteinase domain-containing protein 10 (98). Epidermal growth factor receptor (EGFR) is a transmembrane protein that plays a key role in cell signaling pathways involved in cell growth, proliferation, and survival and is often overexpressed or mutated in various cancers. It was reported that triple-negative breast cancer cells (MDA-MB-468) can produce exosomes with encapsulated EGFR (protected from EGFR inhibitor), which can induce EGFR signaling in target cells, thereby promoting cancer progression or resistance to therapy (99). Typical non-coding RNA cancer biomarkers are miR-1246 and miR-21, which were also overexpressed in breast cancer cell lines (100). The plasma level of miR-1246 was measured in breast cancer patients and healthy controls using an Au nanoflare probe. This biomarker-based probe distinguished breast cancer patients from healthy individuals with 100% sensitivity and 92.9% specificity (95). Currently, a clinical trial (NCT02662621) is underway, focusing on an exosomal detection protocol for diagnosing various cancers, including breast cancer. The study indicates that exosomes displaying the stress protein HSP70 on their membrane may serve as cancer-specific exosomes (134). In colorectal cancer, tetraspanin-1 was found to be upregulated in plasma exosomes from patients compared to healthy controls, showing 75.7% sensitivity (135). Pancreatic cancer, one of the deadliest cancer types, involves TEXs that carry the glypican-1 biomarker. This co-receptor for various signaling molecules regulates key processes such as cell growth, motility, and differentiation (102). Currently, two clinical trials are underway to assess the efficacy of diagnostic exosomes in colon and liver cancers (NCT03432806) and in pancreatic cancer (NCT03334708). These studies aim to identify biomarkers circulating in blood and analyze the corresponding tissues (134). While previous studies have focused on TEXs from blood, urinary exosomes play a role in urological tumors (136). In prostate cancer, biomarkers such as TM256 and LAMTOR proteins found in urinary exosomes exhibited very high sensitivity (105). In addition, potential biomarkers such as TPP1, TMPRSS2, and FOLR1 were highly upregulated in urinary exosomes derived from the bladder. Notably, despite being histologically tumor-free at cystectomy, patients' urinary exosomes displayed a carcinogenic metabolic profile, likely originating from undetected or partially transformed cancer cells (137).

Current diagnosis of NDs relies on clinical assessments, medical history, imaging techniques, and diagnostic tests based on observed symptoms (138, 139). However, predicting these diseases remains challenging due to the typically late-stage diagnosis. Emerging research highlights the potential of exosomes as biomarkers for early detection, offering a promising approach for more effective and timely diagnoses in the future (Figure 5) (140). Early diagnosis of NDs is crucial for enabling timely interventions that can significantly improve patient outcomes (141).

In Alzheimer's disease (AD), neural-derived blood exosomes have emerged as a valuable source of AD-related overexpressed protein markers, including total tau, P-T181-tau, P-S396-tau, and Aβ42 (33). Exosomal synaptic proteins, such as growth-associated GAP43, neurogranin, SNAP25, and synaptotagmin 1, show promise in predicting AD at asymptomatic stages, potentially detecting the disease 5 to 7 years before cognitive impairment occurs (34). In addition, miRNA markers, including miR-137, miR-181c, miR-9, and miR-29a/b, are downregulated in blood serum and offer additional potential for early AD detection (142, 143). Among other types of biomarkers, lipids such as LDL-C, TG, and HDL-C are associated with AD (144). Furthermore, it was also shown that exosomal protein markers offer comparable diagnostic capacity to cerebrospinal markers, further enhancing the potential of blood-based diagnostics for AD (145). Exosomal levels of Aβ42, Aβ40, and P-T181-tau were also measured in the blood plasma of patients diagnosed with schizophrenia (36). Specifically, these markers were determined in neural and astrocytic exosomes. While Aβ42 levels were higher in astrocytic exosomes than in neural exosomes, other markers were similar between these two groups. It was also found that higher astrocytic P-T181-tau levels were associated with worse executive functioning, and astrocytic Aβ42 levels were more sensitive and specific in differentiating diagnostic groups. Other types of neural cells explored in the context of NDs are oligodendrocytes. These cells are associated with MS. Oligodendrocyte-derived extracellular vesicles showed higher concentrations of myelin basic protein, which could serve as potential biomarkers across diverse MS phenotypes (35). For Parkinson disease (PD), it was found that the most reliable biomarker of blood neural exosomes is α-synuclein (146). In amyotrophic lateral sclerosis (ALS), exosomal biomarkers, such as neurofilament light chain, have been identified in both cerebrospinal fluid and blood. This marker can serve for early diagnosis and monitoring disease progression (147).

4.3 Exosomes in targeted delivery and disease treatment

The two main areas in which exosome applications are greatly investigated and hold great potential are cancer and NDs (74, 76). In cancer treatment, significant attention is paid to nanocarriers which can entrap chemotherapeutic drugs and deliver them to the diseased site, reducing the side effects associated with the systemic administration of conventional anticancer drugs (148, 149). In recent years, exosomes started to be explored as promising nanocarriers (48) that can affect tumor growth, metastasis, and even sensitize cancer cells to conventional therapies (Figure 6).

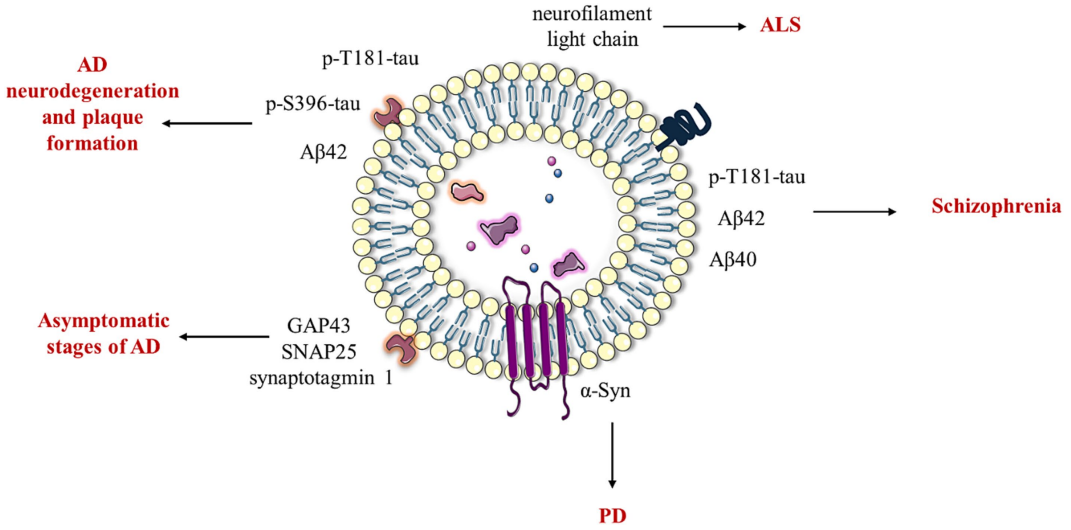


FIGURE 5
Potential of exosomes in the diagnosis of neurological diseases. In AD, protein markers such as total tau, phosphorylated T181-tau (p-T181-tau), phosphorylated S396-tau (p-S396-tau), and amyloid-beta 42 (Aβ42) derived from neural-derived blood exosomes are indicative of neurodegeneration and plaque formation. Notably, GAP43, SNAP25, and synaptotagmin 1 can be detected even in asymptomatic stages of AD. Furthermore, exosomal Aβ42, Aβ40, and P-T181-tau have been implicated in schizophrenia. In the context of PD and ALS, exosomal α-synuclein and neurofilament light chain exhibit promising potential as biomarkers for disease monitoring and progression. The figure was partly generated using Servier Medical Art, provided by Servier, licensed under a Creative Commons Attribution 3.0 unported license.

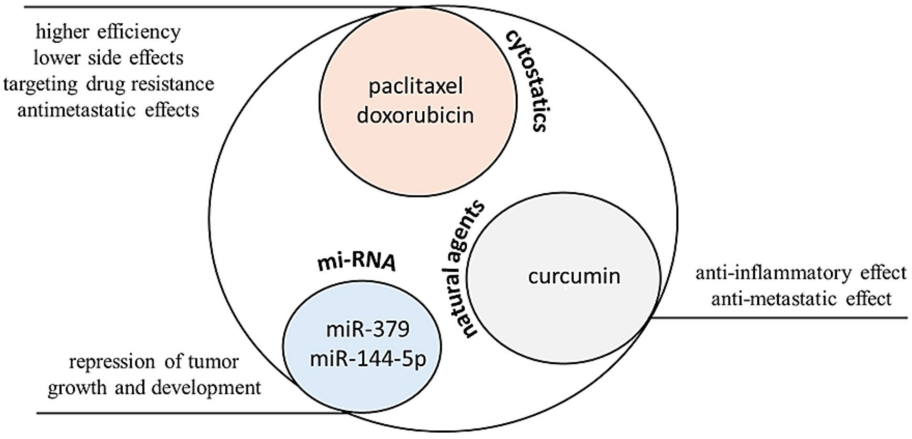


FIGURE 6
Exosomes represent a promising delivery system for the transport of anticancer agents. In murine cancer models, specifically in breast and ovarian carcinoma, exosomal formulations exhibit lower cardiotoxicity and higher efficacy compared to free doxorubicin. In the case of paclitaxel, exosomal formulations provide significantly improved intracellular delivery into 3LL-M27 cells (Lewis carcinoma expressing P-glycoprotein, which is associated with drug resistance) when compared to liposomal formulations and polystyrene nanoparticles. Furthermore, in these murine models, exosomes have been shown to suppress the development of metastases. Regarding natural agents, curcumin-loaded exosomes display promising efficacy against colon and pancreatic carcinomas, exhibiting antimetastatic effects and reduced inflammation. Exosomes are also suitable for the transport of biological agents such as microRNAs. For instance, exosomes loaded with miR-379 and miR-144-5p demonstrate potent effects against breast cancer and pancreatic ductal adenocarcinoma. The figure was partly generated using Servier Medical Art, provided by Servier, licensed under a Creative Commons Attribution 3.0 unported license.

Their ability to target specific cell types and deliver therapeutic cargos makes them a valuable asset in the fight against malignancies (150–153). For example, exosomes loaded with doxorubicin (ExoDOX) were used in a mouse model of breast and ovarian cancers. It was found that ExoDOX are less cardiotoxic than free DOX, enabling the use of higher concentrations of ExoDOX, thus increasing the efficacy of DOX (154). ExoDOX conjugated with gold nanoparticles (ExoDOX-GNPs) were used *in vitro* with lung cancer cell lines and normal lung fibroblasts. The pH sensitive conjugation bond enables the enhanced rate of drug release under acidic conditions and successful uptake of the ExoDOX-GNPs by the recipient cells. Cell viability assays indicated that ExoDOX-GNPs exhibit preferential cytotoxicity toward cancer cells and have minimal activity on non-cancerous cells (155). Exosomes loaded with paclitaxel (ExoPTX) were developed. The incorporation of paclitaxel into exosomes enhanced drug cytotoxicity more than 50-fold in drug-resistant

MDCKMDR1 (Pgp+) cells (156). Furthermore, ExoPTX displayed a potent anticancer effect in a mouse model of murine Lewis Lung Carcinoma pulmonary metastases. In addition to separate therapeutics, other compounds are also used as exosomal cargo. Exosomes loaded with gemcitabine in combination with the survivin protein with mutation T34A induce apoptosis and enhance gemcitabine-killing effects in pancreatic adenocarcinoma cells (157). For cancer treatment, exosomes can also deliver nucleic acids, including miRNAs, siRNAs, and mRNAs (158). For instance, exosomes containing miR-379, a potential tumor suppressor, showed a significant reduction in the rate of tumor formation and growth for the *in vitro* and *in vivo* therapies of breast cancer (159). Exosomes loaded with miR-145-5p, which inhibits pancreatic ductal adenocarcinoma (PDAC) cell proliferation, invasion, and increase apoptosis, significantly reduced the growth of xenograft tumors in a PDAC mouse model (160). Natural products are also used in exosomal cancer therapies, including the use of curcumin. This molecule can mitigate cancer initiation and metastasis (161, 162). Exosomes loaded with curcumin induced the apoptosis of pancreatic cancer cells and significantly delayed brain tumor growth with reduced inflammation when delivered to a GL26 brain tumor model via an intranasal route (163, 164). Plant exosomes with curcumin are used also in Phase I clinical trial investigating the ability of plant exosomes to deliver curcumin to normal and malignant colon tissue (NCT01294072). This study is now recruiting patients (165). Grape exosomes are investigated in preliminary active clinical trial to abrogate oral mucositis induced by combined chemotherapy and radiation in head and neck cancer patients (NCT01668849) (165). Vaccination with tumor antigen-loaded dendritic cell-derived exosomes on patients with unresectable NSCLC lung cancer responding to induction chemotherapy was explored in another clinical study (NCT01159288), where the first phase of this study has now been completed. The primary endpoint was progression-free survival at 4 months after chemotherapy cessation, with a target of at least 50% of patients achieving this endpoint. However, this target was not met as only 32% of patients experienced disease stabilization at 4 months (165, 166). A study of mesenchymal stromal cell-derived exosomes with KrasG12D siRNA for metastatic pancreas cancer patients harboring the KrasG12D mutation is now recruiting patients for the NCT03608631 clinical trial (166). G12D is the most common KRAS mutation detected in carcinomas and confers a unique structural conformation that influences downstream signaling and may lead to its potent oncogenic activity (167).

It is noteworthy that anti-tumor exosomes could be produced by activated T cells themselves (Figure 7). It is well established that tumor cells often express programmed death-ligand 1 (PD-L1) to deactivate T cells through the activation of programmed death-1 (PD-1) signaling pathways (168). Furthermore, research has shown that these tumor cells also release exosomes containing PD-L1, which exhibit similar immunosuppressive functionality. However, Qiu et al. have reported that activated T cells produce exosomes containing PD-1, which serve to neutralize exosomes carrying PD-L1 (169) and induce the degradation of PD-L1 on the surface of TNBC cells. This dynamic interplay highlights the potential of T-cell-derived exosomes in counteracting tumor-mediated immune evasion.

The ability of exosomes to cross the blood–brain barrier (BBB) provides the opportunity for their use in the treatment of NDs (45, 170). Many publications highlight the application of MSCs-EXOs in the treatment of NDs (Figure 8) (171, 172).

For example, the effect of bone marrow-derived MSCs-EXOs was examined in rats with induced Parkinson's disease (PD). Rats treated with these exosomes showed significant improvements in motor function and histopathological outcomes, demonstrating a greater suppression of PD symptoms compared to L-DOPA treatment, which

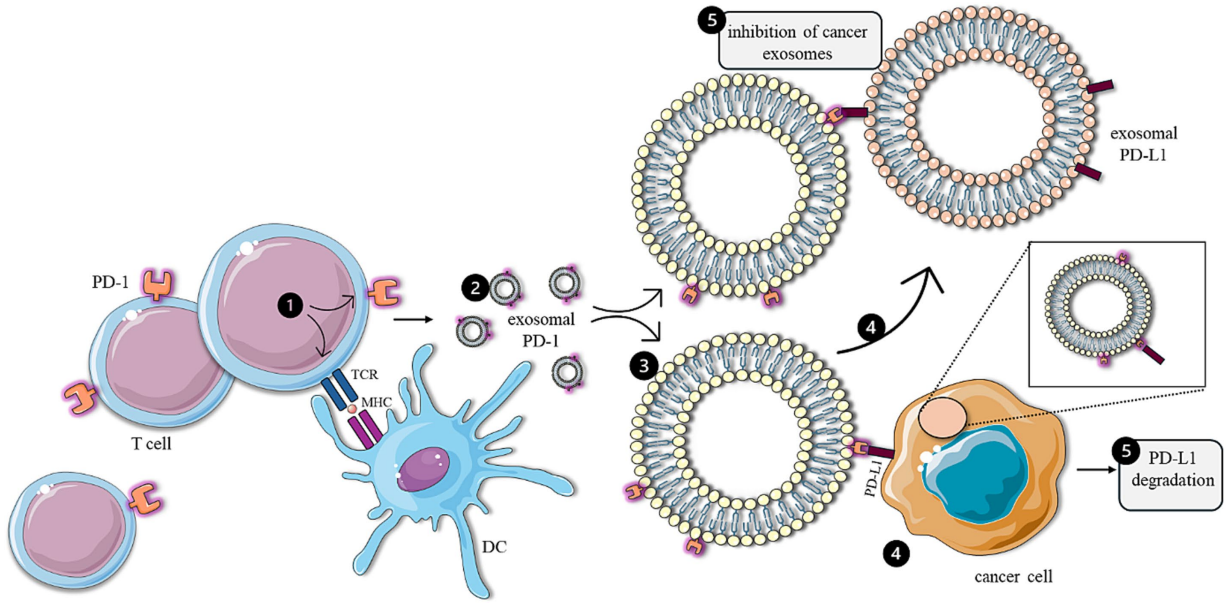


FIGURE 7
Anti-tumor effects of T-cell-derived exosomes. Following activation by antigen-presenting cells (1), T cells generate exosomes containing PD-1 (2), which directly attenuate the immunosuppressive effects of cancer cells (4) by facilitating the degradation of PD-L1 (5) present on the surface of tumor cells. In addition, T-cell-derived exosomes (3) inhibit the functionality of immunosuppressive exosomes expressing PD-L1 (5), produced by cancer cells. The figure was partly generated using Servier Medical Art, provided by Servier, licensed under a Creative Commons Attribution 3.0 unported license.

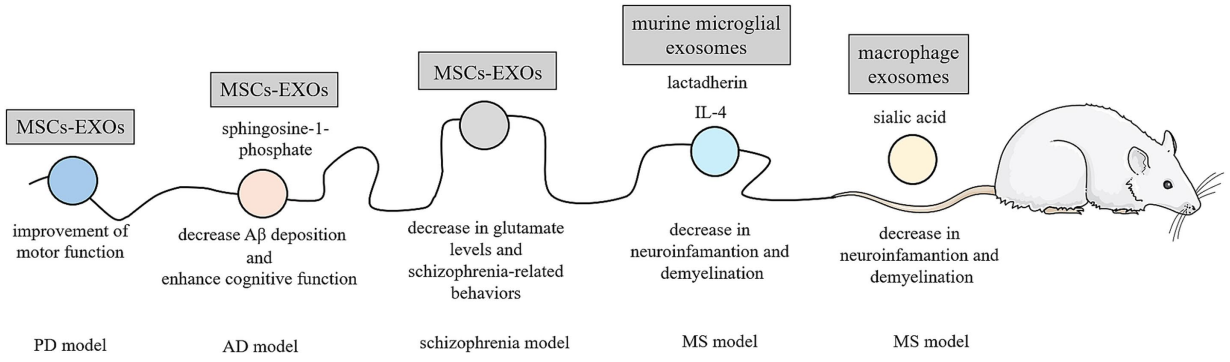


FIGURE 8
Exosomes in the treatment of neurodegenerative diseases. Due to their ability to cross the BBB, exosomes are being investigated for the treatment of neurodegenerative diseases. In a rat model of PD, bone marrow-derived mesenchymal stem cell exosomes (MSCs-EXOs) have been shown to improve motor function. Similarly, MSCs-EXOs with a high content of sphingosine-1-phosphate have been found to decrease Aβ deposition and enhance cognitive function in a mouse model of AD. In a mouse model of schizophrenia, MSCs-EXOs have been reported to reduce cerebrospinal fluid glutamate levels and alleviate schizophrenia-related behaviors. Exosomes derived from murine microglial cells contain IL-4, an anti-inflammatory cytokine, and lactadherin, a phagocytic "eat me" signal, which reduce neuroinflammation in a mouse model of experimental autoimmune encephalomyelitis (a MS model). Similarly, effects were observed with macrophage-derived exosomes that were modified with derivatives of sialic acid, containing resveratrol. The figure was partly generated using Servier Medical Art, provided by Servier, licensed under a Creative Commons Attribution 3.0 unported license.

is a medication commonly used to treat PD (173). In AD treatment, bone marrow MSCS-EXOs with high content of sphingosine-1-phosphate (second messenger downregulated in the AD tissue) were injected into double transgenic AD mice (174). Their application led to reduce Aβ deposition and promote cognitive function recovery. According to the observed results, sphingosine kinase inhibitor (SKI-II) or sphingosine-1-phosphate 1 receptor blocker (VPC23019) repress the therapeutic effects of exosomes. Intracerebral injection of bone marrow MSCS-EXOs in a preclinical mouse model of early stage of AD suggests the possibility of intervening before overt clinical manifestations. The study indicated that bone marrow MSC-EXOs are effective at reducing the Aβ plaque burden and the number of dystrophic neurites in both cortex and hippocampus (175). Adipose-MSCs-EXOs administered intranasally and HUCMSCs-EXOs administered intravenously have also shown promise in AD treatment. These treatments effectively improved neurological damage in entire brain regions, increased newborn neurons, powerfully rescued memory deficits, and reduced Aβ expression in transgenic AD mice (176–179). MSCS-EXOs were also used in a mouse model of schizophrenia, where they improved the core schizophrenia-like behavior and biochemical markers of schizophrenia (e.g., cerebrospinal fluid glutamate level) (180). In addition to MSCs-EXOs, macrophage-derived EXOs loaded with resveratrol, nature antioxidant, or exosomes from murine microglia cell line containing anti-inflammatory cytokine IL-4 were also used in mice with MS (181, 182). The efficacy of this approach can be increased by suitable derivatization of the exosome membrane with compounds such as lactadherin ("eat me" signal for the phagocytes) or sialic acid derivatives (BBB transport). Both exosome agents significantly inhibited inflammatory responses in the CNS and the peripheral system in a mouse model and effectively improved the clinical evolution of MS *in vivo* (182, 183). In clinical studies, there is just a single trial registered in CLinicalTrials.gov. This study evaluates the safety and the efficacy of adipose MSC-EXOs in AD (NCT04388982). It was found that the intranasal administration of MSCs-EXOs was safe and well tolerated for a semi-weekly treatment frequency. A dose of at least 4×10^8 particles was selected for a randomized phase II and phase III clinical trial in further steps (181).

5 Future direction

Exosomes are increasingly recognized for their pivotal role in the diagnostic realm of cancer and NDs. Nonetheless, a multitude of high-impact clinical trials have unveiled their promising potential in diagnosing a spectrum of additional serious conditions. Hyperuricemia (HUA) is acknowledged as a significant risk factor for chronic heart failure (CHF), a disease frequently linked to elevated morbidity and mortality rates (184, 185). Notably, fluctuations in miRNA expression have been correlated with cardiovascular diseases, including CHF and HUA (186, 187). Analysis of miRNA patterns in serum exosomes revealed that miR-27a-5p was upregulated ($p < 0.01$), while miR-139-3p was downregulated ($p < 0.01$) in patient groups (CHF with HUA) (188). When used in combination, these markers exhibited an AUC of 0.899 (95%) with 79.2% sensitivity and 91.7% specificity. Moreover, exosomes are integral to the pathogenesis of osteoarthritis, with those isolated from synovial fluids emerging as promising diagnostic instruments for affected patients (189, 190). However, exosome isolation methods can be invasive. In contrast, urine biomarkers present a non-invasive alternative for osteoarthritis diagnostics (191), positioning urine-derived exosomes as potential diagnostic assets. Cao et al. introduced a nanopolymer modified with an exosome-affinity component (CD63 aptamer and distearoyl phosphoethanolamine) (192). This innovative approach leads to aggregation of exosomes upon binding, facilitating easy centrifugation (4,000 g for 3 min at pH 6). Notably, these precipitated exosomes can be re-dissolved in a basic pH environment. Metabolomic analyses of urine exosomes have identified 30 biomarkers, including catechol (AUC = 0.917, $p < 0.001$), which effectively differentiate between healthy individuals and those with early osteoarthritis.

In the future, exosomes such as natural nanocarriers should emerge as a novel therapeutic alternative in the fields of oncology, immunology, and regenerative medicine. Future research may focus on refining loading techniques, optimizing targeting strategies, and exploring novel applications in diverse disease contexts (193). For example, in ND treatment to enhance targeted drug delivery efficiency, exosomes can be surface-modified with RVG protein/peptides to specifically bind to the acetylcholine receptor expressed on neuronal cells (194). Comparing RVG-tagged MSC-EXOs to MSCS-EXOs, EXOs tagged with RVG exhibit improved targeting to the cortex and hippocampus after being administered intravenously, plaque deposition and Aβ levels are decreased sharply, and the activation of astrocytes is obviously reduced compared to the observations made in the group of AD mice treated only with MSCs-EXOs. In the group of AD mice injected with RVG-EXOs, there is a significant improvement in learning and memory capabilities with reduced plaque deposition and Aβ levels (195).

Exosomes can also be combined with other biomaterials or inorganic materials for biomedical uses. These hybrid nanoparticles can be further loaded with specific cargo or drug and surface engineered to increase the local concentration of the particles at the diseased site, thereby reducing toxicity and side effects and maximizing therapeutic efficacy (196). The surface engineered exosomal hybrid nanoparticles with specific cargo loading, and modifications conferring desired properties such as pH sensitivity or photosensitivity can be called "ExoBots." This term reflects the hybrid nature of these structures, which incorporate elements from both biological (exosomes) and technological (robotics) realms to form advanced nanoparticle system. ExoBots combine the advantageous properties of exosomes and other nanoparticles, holding great promise for advancing therapeutic interventions across various biomedical applications. For instance, an engineered core-shell hybrid system was prepared for the *in vivo* treatment of PD mice. This hybrid system consisted of a curcumin-loaded polymer nanoparticle core and an RVG-modified exosome shell. This hybrid was able to clear α-synuclein aggregates, reduce their cytotoxicity in neurons, and improve the motor behavior of PD mice (197).

Another type of hybrid is composed from exosomes and liposomes. Long circulating and pH sensitive hybrids loaded with DOX were investigated for anti-tumor effect on a mouse model of breast cancer. The results indicated that this hybrid system may be a promising nanocarrier for the treatment of breast cancer, reducing toxicity and inhibiting metastasis mainly in the lungs (198). This hybrid approach was also used to overcome chemotherapy resistance in OC. The hybrid system was developed by fusing cRGD-modified liposomes loaded with miR-497 and triptolide with CD47-expressing tumor-derived exosomes. RGD peptides specifically bind to integrin receptors overexpressed on the surface of many cancer cells. Overexpression of miR-497 may overcome OC chemotherapy resistance, and triptolide was confirmed to possess a superior killing effect on cisplatin-resistant cell lines. The *in vitro* results indicated that these hybrids were efficiently taken up by tumor cells, thus significantly enhancing tumor cell apoptosis and exerting significant anticancer activity without any negative effects observed *in vivo*. These hybrids may provide a translational strategy to overcome cisplatin-resistant OC (199). In addition, paclitaxel was used as the cargo of hybrids for *in vivo* colon cancer treatment. The study revealed that hybrids significantly suppressed tumor growth in a colon tumor-bearing mouse model, reduced the expression of M2 type tumor-associated macrophages, and decreased regulatory T cells (200). Many other types of hybrids were explored, such as thermosensitive hybrids for improved treatment of metastatic peritoneal cancer, pH sensitive macrophage hybrids loaded with DOX for tumor targeted drug delivery, or long circulated pH sensitive hybrids loaded with dasatinib for pancreatic cancer treatment. All these ExoBots show positive therapeutic results and can thus serve as potential therapeutics for cancer treatment (201–203).

Recently, alongside the extensive research on exosomes derived from eukaryotic cells, there has been a growing interest in exploring plant-derived exosomes (P-ELNs; puerarin) as a novel source of exosomes with potential applications in various biotechnological and therapeutic fields. Polyphenolic compounds found in various plant exosomes show great promise in treating serious health disturbances. High-impact studies (204, 205) suggest that their therapeutic effects may be partially attributed to their ability to target ferroptosis, a process linked to numerous pathological states (206, 207).

For instance, exosome-like nanovesicles derived from *P. lobata* roots have been shown to alleviate alcoholic intoxication, enhance alcohol metabolism, and reduce alcohol levels in the liver and serum of mouse models (204). These effects are associated with the induction of acetaldehyde dehydrogenase activity and a decrease in glutathione peroxidase 4 and glutathione levels, as well as with the suppression of acyl-CoA synthetase long-chain family member 4, likely contributing to the repression of ferroptosis. In addition, *Robinia pseudoacacia L.* flower-derived P-ELNs, when administered orally, significantly reduce hypoxia-induced ferroptosis and mucosal injury in the gastrointestinal tract of mouse models (205). Their effects are mediated through the modulation of HIF-1α and HIF-2α expression, subsequently influencing ROS production and lipid peroxidation via NOX4 and ALOX5 pathways.

Another promising application of P-ELNs is their potential to mitigate obesity, a well-known risk factor for cancer, metabolic disorders, cardiovascular diseases, and inflammation (208–210). The global prevalence of obesity has been on the rise, making this research particularly relevant. Wang et al. reported that turmeric-derived ELNs exhibit potent anti-obesity effects, achieving weight reductions of 8.68 and 14.56% through intragastric and subcutaneous delivery, respectively (211). This effect is linked to the stimulation of adipocyte apoptosis, the induction of lipolysis, and the inhibition of lipogenesis, highlighting the therapeutic potential of these natural compounds. However, the biological functions of P-ELNs are not fully understood, standard isolation protocols are lacking, and P-ELNs are a promising new frontier in precision medicine. Their plant origin offers advantages in terms of biocompatibility, scalability, and reduced immunogenicity. Moreover, bioactive molecules from plants are associated with disease-preventive effects making P-ELNs an attractive alternative to mammalian EVs in future biomedical innovations (212–214).

6 Conclusion

Although further fundamental research, especially regarding the biogenesis of exosomes and the optimization of their isolation techniques and characterization methods, is necessary, their significant potential has been demonstrated in several biomedical areas, particularly in regenerative medicine, disease diagnosis, and treatment. The continued

growth of the exosome market is also evidenced by the recent announcement of four collaborations between pharmaceutical companies, including two potentially worth close to $1 billion each. One of the largest is Lilly's partnership with Evox Therapeutics of Oxford, UK. In this deal, which could bring in up to $1.2 billion in milestone payments, CNS-targeting exosomes developed by Evox will be loaded with RNA interference and antisense oligonucleotide therapies from Lilly, targeting up to five undisclosed targets. In another major deal, potentially worth over $900 million, Carmine Therapeutics has partnered with Takeda to develop gene therapies for two undisclosed rare diseases targets (215). Exosomes may thus be the future of medicine, used as ExoBots programmed to deliver specific drugs to specific locations within the organism with minimal side effects and high therapeutic efficacies.

Author contributions

NO: Writing – original draft. VŠ: Writing – original draft. RH: Formal analysis, Writing – review & editing. MS: Writing – original draft. PD: Writing – original draft. DH: Writing – review & editing. ZK: Writing – review & editing. JH: Writing – original draft. FV: Writing – review & editing. MV: Writing – original draft. PM: Writing – review & editing, Formal analysis. MJ: Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was supported by the project of C2P (NEXARS IP 073-23-04). This study was supported by the projects of Charles University in Prague (SVV260637; SVV260521; UNCE 204064; Cooperatio DIAG). This study was also supported by the Ministry of Education, Youth, and Sports grant no. LM2023053 (EATRIS-CZ), the Technology Agency of the Czech Republic within project TN02000122. The study was also supported by the Ministry of Health grant no. RVO-VFN 64165. We are also grateful for the support from the project National Institute for Cancer Research (Program EXCELES, ID Project No. LX22NPO5102) funded by the European Union—Next Generation. The study was also supported by the National Institute for Neurological Research (Programme EXCELES, ID Project No. LX22NPO5107) funded by the European Union—Next Generation EU. The study was supported from the project MULTIOMICS_CZ (Programme Johannes Amos Comenius, Ministry of Education, Youth and Sports of the Czech Republic,__ ID Project CZ.02.01.01/00/23_020/0008540) co-funded by the European Union. We also acknowledge the Operational Programme Research, Development, and Education within the project Center for Tumor Ecology—Research of the Cancer Microenvironment Supporting Cancer Growth and Spread (reg. no. CZ.02.1.01/0.0/0.0/16_019/0000785).

Conflict of interest

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

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

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2025.1539714/full#supplementary-material

References

1. Bayraktar HRME, Abd-Ellah KHGMF, Amero P, Chavez-Reyes A, Rodriguez-Aguayo C. Exosomes: from garbage bins to promising therapeutic targets. *Int J Mol Sci*. (2017) 18:538. doi: 10.3390/ijms18030538

2. Burtenshaw D, Regan B, Owen K, Collins D, McEneaney D, Megson IL, et al. Exosomal composition, biogenesis and profiling using point-of-care diagnostics-implications for cardiovascular disease. *Front Cell Dev Biol*. (2022) 10:853451. doi: 10.3389/fcell.2022.853451

3. Sheta M, Taha EA, Lu Y, Eguchi T. Extracellular vesicles: new classification and tumor immunosuppression. *Biology*. (2023) 12:110. doi: 10.3390/biology12010110

4. Pan BT, Teng K, Wu C, Adam M, Johnstone RM. Electron microscopic evidence for externalization of the transferrin receptor in vesicular form in sheep reticulocytes. *J Cell Biol*. (1985) 101:942–8. doi: 10.1083/jcb.101.3.942

5. Johnstone RM, Adam M, Hammond JR, Orr L, Turbide C. Vesicle formation during reticulocyte maturation. Association of plasma membrane activities with released vesicles (exosomes). *J Biol Chem*. (1987) 262:9412–20. doi: 10.1016/S0021-9258(18)48095-7

6. Kalluri R, LeBleu VS. The biology, function, and biomedical applications of exosomes. *Science*. (2020) 367:eaau6977. doi: 10.1126/science.aau6977

7. Chen J, Li P, Zhang T, Xu Z, Huang X, Wang R, et al. Review on strategies and Technologies for Exosome Isolation and Purification. *Front Bioeng Biotechnol*. (2021) 9:811971. doi: 10.3389/fbioe.2021.811971

8. Yakubovich EI, Polischouk AG, Evtushenko VI. Principles and problems of exosome isolation from biological fluids. *Biochem Suppl Ser A Membr Cell Biol*. (2022) 16:115–26. doi: 10.1134/S1990747822030096

9. Dilsiz N. A comprehensive review on recent advances in exosome isolation and characterization: toward clinical applications. *Transl Oncol*. (2024) 50:102121. doi: 10.1016/j.tranon.2024.102121

10. Kumari S, Lausted C, Scherler K, Ng AHC, Lu Y, Lee I, et al. Approaches and challenges in characterizing the molecular content of extracellular vesicles for biomarker discovery. *Biomol Ther*. (2024) 14:1599. doi: 10.3390/biom14121599

11. Chen YF, Luh F, Ho YS, Yen Y. Exosomes: a review of biologic function, diagnostic and targeted therapy applications, and clinical trials. *J Biomed Sci*. (2024) 31:67. doi: 10.1186/s12929-024-01055-0

12. Zhang X, Borg EGF, Liaci AM, Vos HR, Stoorvogel W. A novel three step protocol to isolate extracellular vesicles from plasma or cell culture medium with both high yield and purity. *J Extracell Vesicles*. (2020) 9:1791450. doi: 10.1080/20013078.2020.1791450

13. Jafari N, Llevenes P, Denis GV. Exosomes as novel biomarkers in metabolic disease and obesity-related cancers. *Nat Rev Endocrinol*. (2022) 18:327–8. doi: 10.1038/s41574-022-00666-7

14. Ghosh S, Rajendran RL, Mahajan AA, Chowdhury A, Bera A, Guha S, et al. Harnessing exosomes as cancer biomarkers in clinical oncology. *Cancer Cell Int*. (2024) 24:278. doi: 10.1186/s12935-024-03464-5

15. Zhang Z, Zou Y, Song C, Cao K, Cai K, Chen S, et al. Advances in the study of exosomes in cardiovascular diseases. *J Adv Res*. (2024) 66:133–53. doi: 10.1016/j.jare.2023.12.014

16. Park C, Weerakkody JS, Schneider R, Miao S, Pitt D. CNS cell-derived exosome signatures as blood-based biomarkers of neurodegenerative diseases. *Front Neurosci*. (2024) 18:1426700. doi: 10.3389/fnins.2024.1426700

17. Bano A, Vats R, Verma D, Yadav P, Kamboj M, Bhardwaj R. Exploring salivary exosomes as early predictors of oral cancer in susceptible tobacco consumers: noninvasive diagnostic and prognostic applications. *J Cancer Res Clin Oncol*. (2023) 149:15781–93. doi: 10.1007/s00432-023-05343-4

18. Bozyk N, Tang KD, Zhang X, Batstone M, Kenny L, Vasani S, et al. Salivary exosomes as biomarkers for early diagnosis of oral squamous cell carcinoma. *Oral Oncol Rep*. (2023) 6:100017. doi: 10.1016/j.oor.2023.100017

19. Nijakowski K, Surdacka A. Salivary biomarkers for diagnosis of inflammatory bowel diseases: a systematic review. *Int J Mol Sci*. (2020) 21:7477. doi: 10.3390/ijms21207477

20. Sun IO, Lerman LO. Urinary extracellular vesicles as biomarkers of kidney disease: from diagnostics to therapeutics. *Diagnostics*. (2020) 10:311. doi: 10.3390/diagnostics10050311

21. Lee N, Canagasingham A, Bajaj M, Shanmugasundaram R, Hutton A, Bucci J, et al. Urine exosomes as biomarkers in bladder cancer diagnosis and prognosis: from functional roles to clinical significance. *Front Oncol*. (2022) 12:1019391. doi: 10.3389/fonc.2022.1019391

22. Nilsson J, Skog J, Nordstrand A, Baranov V, Mincheva-Nilsson L, Breakefield XO, et al. Prostate cancer-derived urine exosomes: a novel approach to biomarkers for prostate cancer. *Br J Cancer*. (2009) 100:1603–7. doi: 10.1038/sj.bjc.6605058

23. Zhou B, Xu K, Zheng X, Chen T, Wang J, Song Y, et al. Application of exosomes as liquid biopsy in clinical diagnosis. *Signal Transduct Target Ther*. (2020) 5:144. doi: 10.1038/s41392-020-00258-9

24. Mosquera-Heredia MI, Morales LC, Vidal OM, Barcelo E, Silvera-Redondo C, Velez JI, et al. Exosomes: potential disease biomarkers and new therapeutic targets. *Biomedicines*. (2021) 9:1061. doi: 10.3390/biomedicines9081061

25. Baranyai T, Herczeg K, Onodi Z, Voszka I, Modos K, Marton N, et al. Isolation of exosomes from blood plasma: qualitative and quantitative comparison of ultracentrifugation and size exclusion chromatography methods. *PLoS One*. (2015) 10:e0145686. doi: 10.1371/journal.pone.0145686

26. Chhoy P, Brown CW, Amante JJ, Mercurio AM. Protocol for the separation of extracellular vesicles by ultracentrifugation from in vitro cell culture models. *STAR Protoc*. (2021) 2:100303. doi: 10.1016/j.xpro.2021.100303

27. Plaschke K, Brenner T, Fiedler MO, Holle T, von der Forst M, Wolf RC, et al. Extracellular vesicles as possible plasma markers and mediators in patients with Sepsis-associated delirium-a pilot study. *Int J Mol Sci*. (2023) 24:15781. doi: 10.3390/ijms242115781

28. Soares Martins T, Catita J, Martins Rosa I, Silva OAB, Henriques AG. Exosome isolation from distinct biofluids using precipitation and column-based approaches. *PLoS One*. (2018) 13:e0198820. doi: 10.1371/journal.pone.0198820

29. Tangwattanachuleeporn M, Muanwien P, Teethaisong Y, Somparn P. Optimizing concentration of Polyethelene glycol for exosome isolation from plasma for downstream application. *Medicina (Kaunas)*. (2022) 58:1600. doi: 10.3390/medicina58111600

30. Coughlan C, Bruce KD, Burgy O, Boyd TD, Michel CR, Garcia-Perez JE, et al. Exosome isolation by ultracentrifugation and precipitation and techniques for downstream analyses. *Curr Protoc Cell Biol*. (2020) 88:e110. doi: 10.1002/cpcb.110

31. Liu WZ, Ma ZJ, Kang XW. Current status and outlook of advances in exosome isolation. *Anal Bioanal Chem*. (2022) 414:7123–41. doi: 10.1007/s00216-022-04253-7

32. Abdouh M, Hamam D, Gao ZH, Arena V, Arena M, Arena GO. Exosomes isolated from cancer patients' sera transfer malignant traits and confer the same phenotype of primary tumors to oncosuppressor-mutated cells. *J Exp Clin Cancer Res*. (2017) 36:113. doi: 10.1186/s13046-017-0587-0

33. Fiandaca MS, Kapogiannis D, Mapstone M, Boxer A, Eitan E, Schwartz JB, et al. Identification of preclinical Alzheimer's disease by a profile of pathogenic proteins in neurally derived blood exosomes: a case-control study. *Alzheimers Dement*. (2015) 11:600–7.e1. doi: 10.1016/j.jalz.2014.06.008

34. Jia L, Zhu M, Kong C, Pang Y, Zhang H, Qiu Q, et al. Blood neuro-exosomal synaptic proteins predict Alzheimer's disease at the asymptomatic stage. *Alzheimers Dement*. (2021) 17:49–60. doi: 10.1002/alz.12166

35. Agliardi C, Guerini FR, Zanzottera M, Bolognesi E, Picciolini S, Caputo D, et al. Myelin basic protein in oligodendrocyte-derived extracellular vesicles as a diagnostic and prognostic biomarker in multiple sclerosis: a pilot study. *Int J Mol Sci*. (2023) 24:894. doi: 10.3390/ijms24010894

36. Lee EE, Winston-Gray C, Barlow JW, Rissman RA, Jeste DV. Plasma levels of neuron- and astrocyte-derived Exosomal amyloid Beta1-42, amyloid Beta1-40, and phosphorylated tau levels in schizophrenia patients and non-psychiatric comparison subjects: relationships with cognitive functioning and psychopathology. *Front Psych*. (2020) 11:532624. doi: 10.3389/fpsyt.2020.532624

37. Wang YT, Shi T, Srivastava S, Kagan J, Liu T, Rodland KD. Proteomic analysis of exosomes for discovery of protein biomarkers for prostate and bladder Cancer. *Cancers (Basel)*. (2020) 12:2335. doi: 10.3390/cancers12092335

38. Sun Y, Liu S, Qiao Z, Shang Z, Xia Z, Niu X, et al. Systematic comparison of exosomal proteomes from human saliva and serum for the detection of lung cancer. *Anal Chim Acta*. (2017) 982:84–95. doi: 10.1016/j.aca.2017.06.005

39. Logozzi M, Di Raimo R, Mizzoni D, Fais S. Immunocapture-based ELISA to characterize and quantify exosomes in both cell culture supernatants and body fluids. *Methods Enzymol*. (2020) 645:155–80. doi: 10.1016/bs.mie.2020.06.011

40. Elkommos-Zakhary M, Rajesh N, Beljanski V. Exosome RNA sequencing as a tool in the search for cancer biomarkers. *Noncoding RNA*. (2022) 8:75. doi: 10.3390/ncrna8060075

41. Tang S, Cheng J, Yao Y, Lou C, Wang L, Huang X, et al. Combination of four serum exosomal MiRNAs as novel diagnostic biomarkers for early-stage gastric cancer. *Front Genet*. (2020) 11:237. doi: 10.3389/fgene.2020.00237

42. Gu Z, Yin H, Zhang H, Zhang H, Liu X, Zeng X, et al. Optimization of a method for the clinical detection of serum exosomal miR-940 as a potential biomarker of breast cancer. *Front Oncol*. (2022) 12:956167. doi: 10.3389/fonc.2022.956167

43. Shi J, Kantoff PW, Wooster R, Farokhzad OC. Cancer nanomedicine: progress, challenges and opportunities. *Nat Rev Cancer*. (2017) 17:20–37. doi: 10.1038/nrc.2016.108

44. Azzi S, Hebda JK, Gavard J. Vascular permeability and drug delivery in cancers. *Front Oncol*. (2013) 3:211. doi: 10.3389/fonc.2013.00211

45. Abdelsalam M, Ahmed M, Osaid Z, Hamoudi R, Harati R. Insights into exosome transport through the blood-brain barrier and the potential therapeutical applications in brain diseases. *Pharmaceuticals (Basel)*. (2023) 16:571. doi: 10.3390/ph16040571

46. Fu S, Wang Y, Xia X, Zheng JC. Exosome engineering: current progress in cargo loading and targeted delivery. *Nanoimpact*. (2020) 20:100261. doi: 10.1016/j.impact.2020.100261

47. Cecchin R, Troyer Z, Witwer K, Morris KV. Extracellular vesicles: the next generation in gene therapy delivery. *Mol Ther*. (2023) 31:1225–30. doi: 10.1016/j.ymthe.2023.01.021

48. Lee J. Trends in developing extracellular vesicle-based therapeutics. *Brain Tumor Res Treat*. (2024) 12:153–61. doi: 10.14791/btrt.2024.0027

49. Zhang Z, Wang C, Li T, Liu Z, Li L. Comparison of ultracentrifugation and density gradient separation methods for isolating Tca8113 human tongue cancer cell line-derived exosomes. *Oncol Lett*. (2014) 8:1701–6. doi: 10.3892/ol.2014.2373

50. Lobb RJ, Becker M, Wen SW, Wong CS, Wiegmans AP, Leimgruber A, et al. Optimized exosome isolation protocol for cell culture supernatant and human plasma. *J Extracell Vesicles*. (2015) 4:27031. doi: 10.3402/jev.v4.27031

51. Collino F, Pomatto M, Bruno S, Lindoso RS, Tapparo M, Sicheng W, et al. Exosome and microvesicle-enriched fractions isolated from mesenchymal stem cells by gradient separation showed different molecular signatures and functions on renal tubular epithelial cells. *Stem Cell Rev Rep*. (2017) 13:226–43. doi: 10.1007/s12015-016-9713-1

52. Busatto S, Vilanilam G, Ticer T, Lin WL, Dickson DW, Shapiro S, et al. Tangential flow filtration for highly efficient concentration of extracellular vesicles from large volumes of fluid. *Cells*. (2018) 7:273. doi: 10.3390/cells7120273

53. Luo H, Zhang J, Yang A, Ouyang W, Long S, Lin X, et al. Large-scale isolation of exosomes derived from NK cells for anti-tumor therapy. *Bio Protoc*. (2023) 13:e4693. doi: 10.3389/fimmu.2022.1087689

54. Kim JY, Rhim WK, Seo HJ, Lee JY, Park CG, Han DK. Comparative analysis of MSC-derived exosomes depending on cell culture Media for Regenerative Bioactivity. *Tissue Eng Regen Med*. (2021) 18:355–67. doi: 10.1007/s13770-021-00352-1

55. Kawai-Harada Y, Nimmagadda V, Harada M. Scalable isolation of surface-engineered extracellular vesicles and separation of free proteins via tangential flow filtration and size exclusion chromatography (TFF-SEC). *BMC Methods*. (2024) 1:9. doi: 10.1186/s44330-024-00009-0

56. Ahn SH, Ryu SW, Choi H, You S, Park J, Choi C. Manufacturing therapeutic exosomes: from bench to industry. *Mol Cells*. (2022) 45:284–90. doi: 10.14348/molcells.2022.2033

57. Guerreiro EM, Vestad B, Steffensen LA, Aass HCD, Saeed M, Ovstebo R, et al. Efficient extracellular vesicle isolation by combining cell media modifications, ultrafiltration, and size-exclusion chromatography. *PLoS One*. (2018) 13:e0204276. doi: 10.1371/journal.pone.0204276

58. Nguyen VVT, Witwer KW, Verhaar MC, Strunk D, van Balkom BWM. Functional assays to assess the therapeutic potential of extracellular vesicles. *J Extracell Vesicles*. (2020) 10:e12033. doi: 10.1002/jev2.12033

59. Thery C, Witwer KW, Aikawa E, Alcaraz MJ, Anderson JD, Andriantsitohaina R, et al. Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of

the MISEV2014 guidelines. *J Extracell Vesicles*. (2018) 7:1535750. doi: 10.1080/20013078.2018.1535750

60. Welsh JA, Goberdhan DCI, O'Driscoll L, Buzas EI, Blenkiron C, Bussolati B, et al. Minimal information for studies of extracellular vesicles (MISEV2023): from basic to advanced approaches. *J Extracell Vesicles*. (2024) 13:e12404. doi: 10.1002/jev2.12404

61. Kowal EJK, Ter-Ovanesyan D, Regev A, Church GM. Extracellular vesicle isolation and analysis by Western blotting. *Methods Mol Biol*. (2017) 1660:143–52. doi: 10.1007/978-1-4939-7253-1_12

62. Takizawa K, Nishimura T, Harita Y. Enzyme-linked immunosorbent assay to detect surface marker proteins of extracellular vesicles purified from human urine. *STAR Protoc*. (2023) 4:102415. doi: 10.1016/j.xpro.2023.102415

63. Noble JM, Roberts LM, Vidavsky N, Chiou AE, Fischbach C, Paszek MJ, et al. Direct comparison of optical and electron microscopy methods for structural characterization of extracellular vesicles. *J Struct Biol*. (2020) 210:107474. doi: 10.1016/j.jsb.2020.107474

64. van der Pol E, Coumans FA, Grootemaat AE, Gardiner C, Sargent IL, Harrison P, et al. Particle size distribution of exosomes and microvesicles determined by transmission electron microscopy, flow cytometry, nanoparticle tracking analysis, and resistive pulse sensing. *J Thromb Haemost*. (2014) 12:1182–92. doi: 10.1111/jth.12602

65. Koritzinsky EH, Street JM, Star RA, Yuen PS. Quantification of exosomes. *J Cell Physiol*. (2017) 232:1587–90. doi: 10.1002/jcp.25387

66. Rupert DLM, Claudio V, Lasser C, Bally M. Methods for the physical characterization and quantification of extracellular vesicles in biological samples. *Biochim Biophys Acta Gen Subj*. (1861) 2017:3164–79. doi: 10.1016/j.bbagen.2016.07.028

67. Fortunato D, Mladenovic D, Criscuoli M, Loria F, Veiman KL, Zocco D, et al. Opportunities and pitfalls of fluorescent labeling methodologies for extracellular vesicle profiling on high-resolution single-particle platforms. *Int J Mol Sci*. (2021) 22:10510. doi: 10.3390/ijms221910510

68. Dehghani M, Montange RK, Olszowy MW, Pollard D. An emerging fluorescence-based technique for quantification and protein profiling of extracellular vesicles. *SLAS Technol*. (2021) 26:189–99. doi: 10.1177/2472630320970458

69. Chung IM, Rajakumar G, Venkidasamy B, Subramanian U, Thiruvengadam M. Exosomes: current use and future applications. *Clin Chim Acta*. (2020) 500:226–32. doi: 10.1016/j.cca.2019.10.022

70. Mager E.L.A. S I., X.O. Breakefield, and Wood M.J., Extracellular vesicles: biology and emerging therapeutic opportunities. *Nat Rev Drug Discov* 12 (2013) 347–357. doi: 10.1038/nrd3978

71. Aheget H, Tristan-Manzano M, Mazini L, Cortijo-Gutierrez M, Galindo-Moreno P, Herrera C, et al. Exosome: a new player in translational nanomedicine. *J Clin Med*. (2020) 9:2380. doi: 10.3390/jcm9082380

72. Tienda-Vazquez MA, Hanel JM, Marquez-Arteaga EM, Salgado-Alvarez AP, Scheckhuber CQ, Alanis-Gomez JR, et al. Exosomes: a promising strategy for repair, regeneration and treatment of skin disorders. *Cells*. (2023) 12:1625. doi: 10.3390/cells12121625

73. Zhang K, Cheng K. Stem cell-derived exosome versus stem cell therapy. *Nat Rev Bioeng*. (2023) 1:608–9. doi: 10.1038/s44222-023-00064-2

74. Wang X. Exosome therapy and diagnostics: The path to becoming clinical giants. Washington, USA: CAS (2022).

75. Nooshabadi VT, Khanmohamadi M, Valipour E, Mahdipour S, Salati A, Malekshahi ZV, et al. Impact of exosome-loaded chitosan hydrogel in wound repair and layered dermal reconstitution in mice animal model. *J Biomed Mater Res A*. (2020) 108:2138–49. doi: 10.1002/jbm.a.36959

76. Tan F, Li X, Wang Z, Li J, Shahzad K, Zheng J. Clinical applications of stem cell-derived exosomes. *Signal Transduct Target Ther*. (2024) 9:17. doi: 10.1038/s41392-023-01704-0

77. Prasai A, Jay JW, Jupiter D, Wolf SE, El Ayadi A. Role of exosomes in dermal wound healing: a systematic review. *J Invest Dermatol*. (2022) 142:662–678.e8. doi: 10.1016/j.jid.2021.07.167

78. Shi L, Song D, Meng C, Cheng Y, Wang B, Yang Z. Opportunities and challenges of engineered exosomes for diabetic wound healing. *Giant*. (2024) 18:100251. doi: 10.1016/j.giant.2024.100251

79. Fan MH, Pi JK, Zou CY, Jiang YL, Li QJ, Zhang XZ, et al. Hydrogel-exosome system in tissue engineering: a promising therapeutic strategy. *Bioact Mater*. (2024) 38:1–30. doi: 10.1016/j.bioactmat.2024.04.007

80. Xie Y, Guan Q, Guo J, Chen Y, Yin Y, Han X. Hydrogels for exosome delivery in biomedical applications. *Gels*. (2022) 8:328. doi: 10.3390/gels8060328

81. Khayambashi P, Iyer J, Pillai S, Upadhyay A, Zhang Y, Tran SD. Hydrogel encapsulation of mesenchymal stem cells and their derived exosomes for tissue engineering. *Int J Mol Sci*. (2021) 22:684. doi: 10.3390/ijms22020684

82. Zhang K, Zhao X, Chen X, Wei Y, Du W, Wang Y, et al. Enhanced therapeutic effects of mesenchymal stem cell-derived exosomes with an injectable hydrogel for Hindlimb ischemia treatment. *ACS Appl Mater Interfaces*. (2018) 10:30081–91. doi: 10.1021/acsami.8b08449

83. Yang J, Chen Z, Pan D, Li H, Shen J. Umbilical cord-derived mesenchymal stem cell-derived exosomes combined Pluronic F127 hydrogel promote chronic diabetic wound healing and complete skin regeneration. *Int J Nanomedicine*. (2020) 15:5911–26. doi: 10.2147/IJN.S249129

84. Zhang Y, Zhang P, Gao X, Chang L, Chen Z, Mei X. Preparation of exosomes encapsulated nanohydrogel for accelerating wound healing of diabetic rats by promoting angiogenesis. *Mater Sci Eng C*. (2021) 120:111671. doi: 10.1016/j.msec.2020.111671

85. Wang C, Wang M, Xu T, Zhang X, Lin C, Gao W, et al. Engineering bioactive self-healing antibacterial exosomes hydrogel for promoting chronic diabetic wound healing and complete skin regeneration. *Theranostics*. (2019) 9:65–76. doi: 10.7150/thno.29766

86. Li P, Lv S, Jiang W, Si L, Liao B, Zhao G, et al. Exosomes derived from umbilical cord mesenchymal stem cells protect cartilage and regulate the polarization of macrophages in osteoarthritis. *Ann Transl Med*. (2022) 10:976. doi: 10.21037/atm-22-3912

87. Cao H, Chen M, Cui X, Liu Y, Liu Y, Deng S, et al. Cell-free osteoarthritis treatment with sustained-release of chondrocyte-targeting exosomes from umbilical cord-derived mesenchymal stem cells to rejuvenate aging chondrocytes. *ACS Nano*. (2023) 17:13358–76. doi: 10.1021/acsnano.3c01612

88. Chen P, Zheng L, Wang Y, Tao M, Xie Z, Xia C, et al. Desktop-stereolithography 3D printing of a radially oriented extracellular matrix/mesenchymal stem cell exosome bioink for osteochondral defect regeneration. *Theranostics*. (2019) 9:2439–59. doi: 10.7150/thno.31017

89. Wu KY, Ahmad H, Lin G, Carbonneau M, Tran SD. Mesenchymal stem cell-derived exosomes in ophthalmology: a comprehensive review. *Pharmaceutics*. (2023) 15:1167. doi: 10.3390/pharmaceutics15041167

90. Di G, Du X, Qi X, Zhao X, Duan H, Li S, et al. Mesenchymal stem cells promote diabetic corneal epithelial wound healing through TSG-6-dependent stem cell activation and macrophage switch. *Invest Ophthalmol Vis Sci*. (2017) 58:4344–54. doi: 10.1167/iovs.17-21506

91. Samaeekia R, Rabiee B, Putra I, Shen X, Park YJ, Hematti P, et al. Effect of human corneal mesenchymal stromal cell-derived exosomes on corneal epithelial wound healing. *Invest Ophthalmol Vis Sci*. (2018) 59:5194–200. doi: 10.1167/iovs.18-24803

92. Mead B, Tomarev S. Bone marrow-derived mesenchymal stem cells-derived exosomes promote survival of retinal ganglion cells through miRNA-dependent mechanisms. *Stem Cells Transl Med*. (2017) 6:1273–85. doi: 10.1002/sctm.16-0428

93. Wang T, Zhou Y, Zhang W, Xue Y, Xiao Z, Zhou Y, et al. Exosomes and exosome composite scaffolds in periodontal tissue engineering. *Front Bioeng Biotechnol*. (2023) 11:1287714. doi: 10.3389/fbioe.2023.1287714

94. Lin H, Chen H, Zhao X, Ding T, Wang Y, Chen Z, et al. Advances of exosomes in periodontitis treatment. *J Transl Med*. (2022) 20:279. doi: 10.1186/s12967-022-03487-4

95. Zhai LY, Li MX, Pan WL, Chen Y, Li MM, Pang JX, et al. In situ detection of plasma Exosomal MicroRNA-1246 for breast Cancer diagnostics by a au Nanoflare probe. *ACS Appl Mater Interfaces*. (2018) 10:39478–86. doi: 10.1021/acsami.8b12725

96. Sandfeld-Paulsen B, Jakobsen KR, Baek R, Folkersen BH, Rasmussen TR, Meldgaard P, et al. Exosomal proteins as diagnostic biomarkers in lung Cancer. *J Thorac Oncol*. (2016) 11:1701–10. doi: 10.1016/j.jtho.2016.05.034

97. Akbar S, Raza A, Mohsin R, Kanbour A, Qadri S, Parray A, et al. Circulating exosomal immuno-oncological checkpoints and cytokines are potential biomarkers to monitor tumor response to anti-PD-1/PD-L1 therapy in non-small cell lung cancer patients. *Front Immunol*. (2022) 13:1097117. doi: 10.3389/fimmu.2022.1097117

98. Risha Y, Minic Z, Ghobadloo SM, Berezovski MV. The proteomic analysis of breast cell line exosomes reveals disease patterns and potential biomarkers. *Sci Rep*. (2020) 10:13572. doi: 10.1038/s41598-020-70393-4

99. Hung Y, Wang Y-L, Lin Y-Z, Chiang S-F, Wu W-R, Wang S-C. The exosomal compartment protects epidermal growth factor receptor from small molecule inhibitors. *Biochem Biophys Res Commun*. (2019) 510:42–7. doi: 10.1016/j.bbrc.2018.12.187

100. Hannafon BN, Trigoso YD, Calloway CL, Zhao YD, Lum DH, Welm AL, et al. Plasma exosome microRNAs are indicative of breast cancer. *Breast Cancer Res*. (2016) 18:90. doi: 10.1186/s13058-016-0753-x

101. Gobbo J, Marcion G, Cordonnier M, Dias AMM, Pernet N, Hammann A, et al. Restoring anticancer immune response by targeting tumor-derived exosomes with a HSP70 peptide Aptamer. *JNCI J Natl Cancer Inst*. (2015) 108:djv330. doi: 10.1093/jnci/djv330

102. Melo SA, Luecke LB, Kahlert C, Fernandez AF, Gammon ST, Kaye J, et al. Glypican-1 identifies cancer exosomes and detects early pancreatic cancer. *Nature*. (2015) 523:177–82. doi: 10.1038/nature14581

103. Ueda H, Takahashi H, Kobayashi S, Kubo M, Sasaki K, Iwagami Y, et al. miR-6855-5p enhances Radioresistance and promotes migration of pancreatic Cancer by inducing epithelial-mesenchymal transition via suppressing FOXA1: potential of plasma Exosomal miR-6855-5p as an indicator of radiosensitivity in patients with pancreatic cancer. *Ann Surg Oncol*. (2024) 32:720–35. doi: 10.1245/s10434-024-16115-w

104. Dash S, Wu CC, Wu CC, Chiang SF, Lu YT, Yeh CY, et al. Extracellular vesicle membrane protein profiling and targeted mass spectrometry unveil CD59 and Tetraspanin 9 as novel plasma biomarkers for detection of colorectal cancer. *Cancers*. (2022) 15:177. doi: 10.3390/cancers15010177

105. Overbye A, Skotland T, Koehler CJ, Thiede B, Seierstad T, Berge V, et al. Identification of prostate cancer biomarkers in urinary exosomes. *Oncotarget*. (2015) 6:30357–76. doi: 10.18632/oncotarget.4851

106. Jang JY, Lee JK, Jeon YK, Kim CW. Exosome derived from epigallocatechin gallate treated breast cancer cells suppresses tumor growth by inhibiting tumor-associated macrophage infiltration and M2 polarization. *BMC Cancer*. (2013) 13:421. doi: 10.1186/1471-2407-13-421

107. Whiteside TL. Exosomes and tumor-mediated immune suppression. *J Clin Invest*. (2016) 126:1216–23. doi: 10.1172/JCI81136

108. Hao Q, Wu Y, Wu Y, Wang P, Vadgama JV. Tumor-derived exosomes in tumor-induced immune suppression. *Int J Mol Sci*. (2022) 23:1416. doi: 10.3390/ijms23031461

109. Saccani A, Schioppa T, Porta C, Biswas SK, Nebuloni M, Vago L, et al. p50 nuclear factor-kappaB overexpression in tumor-associated macrophages inhibits M1 inflammatory responses and antitumor resistance. *Cancer Res*. (2006) 66:11432–40. doi: 10.1158/0008-5472.CAN-06-1867

110. Mantovani A, Marchesi F, Porta C, Sica A, Allavena P. Inflammation and cancer: breast cancer as a prototype. *Breast*. (2007) 16:27–33. doi: 10.1016/j.breast.2007.07.013

111. Chanmee T, Ontong P, Konno K, Itano N. Tumor-associated macrophages as major players in the tumor microenvironment. *Cancers (Basel)*. (2014) 6:1670–90. doi: 10.3390/cancers6031670

112. Sica A, Schioppa T, Mantovani A, Allavena P. Tumour-associated macrophages are a distinct M2 polarised population promoting tumour progression: potential targets of anti-cancer therapy. *Eur J Cancer*. (2006) 42:717–27. doi: 10.1016/j.ejca.2006.01.003

113. Ham S, Lima LG, Chai EPZ, Muller A, Lobb RJ, Krumeich S, et al. Breast Cancer-derived exosomes Alter macrophage polarization via gp130/STAT3 signaling. *Front Immunol*. (2018) 9:871. doi: 10.3389/fimmu.2018.00871

114. Piao YJ, Kim HS, Hwang EH, Woo J, Zhang M, Moon WK. Breast cancer cell-derived exosomes and macrophage polarization are associated with lymph node metastasis. *Oncotarget*. (2018) 9:7398–410. doi: 10.18632/oncotarget.23238

115. Nieto MA. The ins and outs of the epithelial to mesenchymal transition in health and disease. *Annu Rev Cell Dev Biol*. (2011) 27:347–76. doi: 10.1146/annurev-cellbio-092910-154036

116. Thiery JP, Acloque H, Huang RY, Nieto MA. Epithelial-mesenchymal transitions in development and disease. *Cell*. (2009) 139:871–90. doi: 10.1016/j.cell.2009.11.007

117. Shen Y, TanTai J. Exosomes secreted by metastatic cancer cells promotes epithelial mesenchymal transition in small cell lung carcinoma: the key role of Src/TGF-beta1 axis. *Gene*. (2024) 892:147873. doi: 10.1016/j.gene.2023.147873

118. Franzen CA, Blackwell RH, Todorovic V, Greco KA, Foreman KE, Flanigan RC, et al. Urothelial cells undergo epithelial-to-mesenchymal transition after exposure to muscle invasive bladder cancer exosomes. *Oncogenesis*. (2015) 4:e163. doi: 10.1038/oncsis.2015.21

119. Christianson HC, Svensson KJ, van Kuppevelt TH, Li JP, Belting M. Cancer cell exosomes depend on cell-surface heparan sulfate proteoglycans for their internalization and functional activity. *Proc Natl Acad Sci USA*. (2013) 110:17380–5. doi: 10.1073/pnas.1304266110

120. Liu D, Li C, Trojanowicz B, Li X, Shi D, Zhan C, et al. CD97 promotion of gastric carcinoma lymphatic metastasis is exosome dependent. *Gastric Cancer*. (2016) 19:754–66. doi: 10.1007/s10120-015-0523-y

121. Rahman MA, Barger JF, Lovat F, Gao M, Otterson GA, Nana-Sinkam P. Lung cancer exosomes as drivers of epithelial mesenchymal transition. *Oncotarget*. (2016) 7:54852–66. doi: 10.18632/oncotarget.10243

122. Cao M, Seike M, Soeno C, Mizutani H, Kitamura K, Minegishi Y, et al. MiR-23a regulates TGF-beta-induced epithelial-mesenchymal transition by targeting E-cadherin in lung cancer cells. *Int J Oncol*. (2012) 41:869–75. doi: 10.3892/ijo.2012.1535

123. Kim J, Kim TY, Lee MS, Mun JY, Ihm C, Kim SA. Exosome cargo reflects TGF-beta1-mediated epithelial-to-mesenchymal transition (EMT) status in A549 human lung adenocarcinoma cells. *Biochem Biophys Res Commun*. (2016) 478:643–8. doi: 10.1016/j.bbrc.2016.07.124

124. Ohshima K, Inoue K, Fujiwara A, Hatakeyama K, Kanto K, Watanabe Y, et al. Let-7 microRNA family is selectively secreted into the extracellular environment via exosomes in a metastatic gastric cancer cell line. *PLoS One*. (2010) 5:e13247. doi: 10.1371/journal.pone.0013247

125. Tanaka S, Hosokawa M, Ueda K, Iwakawa S. Effects of Decitabine on invasion and Exosomal expression of miR-200c and miR-141 in Oxaliplatin-resistant colorectal Cancer cells. *Biol Pharm Bull*. (2015) 38:1272–9. doi: 10.1248/bpb.b15-00129

126. Xiao D, Barry S, Kmetz D, Egger M, Pan J, Rai SN, et al. Melanoma cell-derived exosomes promote epithelial-mesenchymal transition in primary melanocytes through paracrine/autocrine signaling in the tumor microenvironment. *Cancer Lett*. (2016) 376:318–27. doi: 10.1016/j.canlet.2016.03.050

127. Zhou W, Fong MY, Min Y, Somlo G, Liu L, Palomares MR, et al. Cancer-secreted miR-105 destroys vascular endothelial barriers to promote metastasis. *Cancer Cell*. (2014) 25:501–15. doi: 10.1016/j.ccr.2014.03.007

128. Kumar S, Dhar R, Kumar L, Shivji GG, Jayaraj R, Devi A. Theranostic signature of tumor-derived exosomes in cancer. *Med Oncol*. (2023) 40:321. doi: 10.1007/s12032-023-02176-6

129. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. (2022) 74:229–63. doi: 10.3322/caac.21660

130. Shin H, Oh S, Hong S, Kang M, Kang D, Ji YG, et al. Early-stage lung Cancer diagnosis by deep learning-based spectroscopic analysis of circulating exosomes. *ACS Nano*. (2020) 14:5435–44. doi: 10.1021/acsnano.9b09119

131. Hai J, Zhu CQ, Bandarchi B, Wang YH, Navab R, Shepherd FA, et al. L1 cell adhesion molecule promotes tumorigenicity and metastatic potential in non-small cell lung cancer. *Clin Cancer Res*. (2012) 18:1914–24. doi: 10.1158/1078-0432.CCR-11-2893

132. Tischler V, Pfeifer M, Hausladen S, Schirmer U, Bonde AK, Kristiansen G, et al. L1CAM protein expression is associated with poor prognosis in non-small cell lung cancer. *Mol Cancer*. (2011) 10:127. doi: 10.1186/1476-4598-10-127

133. Liu Y, Ding W, Wang J, Ao X, Xue J. Non-coding RNAs in lung cancer: molecular mechanisms and clinical applications. *Front Oncol*. (2023) 13:1256537. doi: 10.3389/fonc.2023.1256537

134. Makler A, Asghar W. Exosomal biomarkers for cancer diagnosis and patient monitoring. *Expert Rev Mol Diagn*. (2020) 20:387–400. doi: 10.1080/14737159.2020.1731308

135. Lee CH, Im EJ, Moon PG, Baek MC. Discovery of a diagnostic biomarker for colon cancer through proteomic profiling of small extracellular vesicles. *BMC Cancer*. (2018) 18:1058. doi: 10.1186/s12885-018-4952-y

136. Xu Y, Lou J, Yu M, Jiang Y, Xu H, Huang Y, et al. Urinary exosomes diagnosis of urological tumors: a systematic review and meta-analysis. *Front Oncol*. (2021) 11:734587. doi: 10.3389/fonc.2021.734587

137. Hiltbrunner S, Mints M, Eldh M, Rosenblatt R, Holmstrom B, Alamdari F, et al. Urinary exosomes from bladder cancer patients show a residual cancer phenotype despite complete pathological downstaging. *Sci Rep*. (2020) 10:5960. doi: 10.1038/s41598-020-62753-x

138. Shusharina N, Yukhnenko D, Botman S, Sapunov V, Savinov V, Kamyshov G, et al. Modern methods of diagnostics and treatment of neurodegenerative diseases and depression. *Diagnostics*. (2023) 13:573. doi: 10.3390/diagnostics13030573

139. Konickova D, Mensikova K, Tuckova L, Henykova E, Strnad M, Friedecky D, et al. Biomarkers of neurodegenerative diseases: biology, taxonomy, clinical relevance, and current research status. *Biomedicines*. (2022) 10:1760. doi: 10.3390/biomedicines10071760

140. Gao P, Li X, Du X, Liu S, Xu Y. Diagnostic and therapeutic potential of exosomes in neurodegenerative diseases. *Front Aging Neurosci*. (2021) 13:790863. doi: 10.3389/fnagi.2021.790863

141. Rastogi S, Sharma V, Bharti PS, Rani K, Modi GP, Nikolajeff F, et al. The evolving landscape of exosomes in neurodegenerative diseases: exosomes characteristics and a promising role in early diagnosis. *Int J Mol Sci*. (2021) 22:440. doi: 10.3390/ijms22010440

142. Geekiyanage H, Jicha GA, Nelson PT, Chan C. Blood serum miRNA: non-invasive biomarkers for Alzheimer's disease. *Exp Neurol*. (2012) 235:491–6. doi: 10.1016/j.expneurol.2011.11.026

143. Manna I, De Benedittis S, Quattrone A, Maisano D, Iaccino E, Quattrone A. Exosomal miRNAs as potential diagnostic biomarkers in Alzheimer's disease. *Pharmaceuticals*. (2020) 13:243. doi: 10.3390/ph13090243

144. Agarwal M, Khan S. Plasma lipids as biomarkers for Alzheimer's disease: a systematic review. *Cureus*. (2020) 12:e12008. doi: 10.7759/cureus.12008

145. Jia L, Qiu Q, Zhang H, Chu L, Du Y, Zhang J, et al. Concordance between the assessment of Abeta42, T-tau, and P-T181-tau in peripheral blood neuronal-derived exosomes and cerebrospinal fluid. *Alzheimers Dement*. (2019) 15:1071–80. doi: 10.1016/j.jalz.2019.05.002

146. Nila IS, Sumsuzzman DM, Khan ZA, Jung JH, Kazema AS, Kim SJ, et al. Identification of exosomal biomarkers and its optimal isolation and detection method for the diagnosis of Parkinson's disease: a systematic review and meta-analysis. *Ageing Res Rev*. (2022) 82:101764. doi: 10.1016/j.arr.2022.101764

147. Barbo M, Ravnik-Glavac M. Extracellular vesicles as potential biomarkers in amyotrophic lateral sclerosis. *Genes*. (2023) 14:325. doi: 10.3390/genes14020325

148. Zafar MN, Abuwatfa WH, Husseini GA. Acoustically-activated liposomal Nanocarriers to mitigate the side effects of conventional chemotherapy with a focus on emulsion-liposomes. *Pharmaceutics*. (2023) 15:421. doi: 10.3390/pharmaceutics15020421

149. Srivastava A, Amreddy N, Razaq M, Towner R, Zhao YD, Ahmed RA, et al. Exosomes as Theranostics for lung Cancer. *Adv Cancer Res*. (2018) 139:1–33. doi: 10.1016/bs.acr.2018.04.001

150. Paskeh MDA, Entezari M, Mirzaei S, Zabolian A, Saleki H, Naghdi MJ, et al. Emerging role of exosomes in cancer progression and tumor microenvironment remodeling. *J Hematol Oncol*. (2022) 15:83. doi: 10.1186/s13045-022-01305-4

151. Maqsood Q, Sumrin A, Saleem Y, Wajid A, Mahnoor M. Exosomes in Cancer: diagnostic and therapeutic applications. *Clin Med Insights Oncol*. (2024) 18:11795549231215966. doi: 10.1177/11795549231215966

152. Cao Y, Xu P, Shen Y, Wu W, Chen M, Wang F, et al. Exosomes and cancer immunotherapy: a review of recent cancer research. *Front Oncol*. (2022) 12:1118101. doi: 10.3389/fonc.2022.1118101

153. Araldi RP, Delvalle DA, da Costa VR, Alievi AL, Teixeira MR, Dias Pinto JR, et al. Exosomes as a Nano-Carrier for chemotherapeutics: a new era of oncology. *Cells*. (2023) 12:2411. doi: 10.3390/cells12172144

154. Hadla M, Palazzolo S, Corona G, Caligiuri I, Canzonieri V, Toffoli G, et al. Exosomes increase the therapeutic index of doxorubicin in breast and ovarian cancer mouse models. *Nanomedicine*. (2016) 11:2431–41. doi: 10.2217/nnm-2016-0154

155. Srivastava A, Amreddy N, Babu A, Panneerselvam J, Mehta M, Muralidharan R, et al. Nanosomes carrying doxorubicin exhibit potent anticancer activity against human lung cancer cells. *Sci Rep*. (2016) 6:38541. doi: 10.1038/srep38541

156. Kim MS, Haney MJ, Zhao Y, Mahajan V, Deygen I, Klyachko NL, et al. Development of exosome-encapsulated paclitaxel to overcome MDR in cancer cells. *Nanomedicine*. (2016) 12:655–64. doi: 10.1016/j.nano.2015.10.012

157. Aspe JR, Diaz Osterman CJ, Jutzy JM, Deshields S, Whang S, Wall NR. Enhancement of gemcitabine sensitivity in pancreatic adenocarcinoma by novel exosome-mediated delivery of the Survivin-T34A mutant. *J Extracell Vesicles*. (2014) 3:1–9. doi: 10.3402/jev.v3.23244

158. Geis-Asteggiante L, Belew AT, Clements VK, Edwards NJ, Ostrand-Rosenberg S, El-Sayed NM, et al. Differential content of proteins, mRNAs, and miRNAs suggests that MDSC and their exosomes may mediate distinct immune suppressive functions. *J Proteome Res*. (2018) 17:486–98. doi: 10.1021/acs.jproteome.7b00646

159. O'Brien KP, Khan S, Gilligan KE, Zafar H, Lalor P, Glynn C, et al. Employing mesenchymal stem cells to support tumor-targeted delivery of extracellular vesicle (EV)-encapsulated microRNA-379. *Oncogene*. (2018) 37:2137–49. doi: 10.1038/s41388-017-0116-9

160. Ding Y, Cao F, Sun H, Wang Y, Liu S, Wu Y, et al. Exosomes derived from human umbilical cord mesenchymal stromal cells deliver exogenous miR-145-5p to inhibit pancreatic ductal adenocarcinoma progression. *Cancer Lett*. (2019) 442:351–61. doi: 10.1016/j.canlet.2018.10.039

161. Song H, Liu B, Dong B, Xu J, Zhou H, Na S, et al. Exosome-based delivery of natural products in Cancer therapy. *Front Cell Dev Biol*. (2021) 9:650426. doi: 10.3389/fcell.2021.650426

162. Mansouri K, Rasoulpoor S, Daneshkhah A, Abolfathi S, Salari N, Mohammadi M, et al. Clinical effects of curcumin in enhancing cancer therapy: a systematic review. *BMC Cancer*. (2020) 20:791. doi: 10.1186/s12885-020-07256-8

163. Osterman CJ, Lynch JC, Leaf P, Gonda A, Ferguson Bennit HR, Griffiths D, et al. Curcumin Modulates Pancreatic Adenocarcinoma Cell-Derived Exosomal Function. *PLoS One*. (2015) 10:e0132845. doi: 10.1371/journal.pone.0132845

164. Zhuang X, Xiang X, Grizzle W, Sun D, Zhang S, Axtell RC, et al. Treatment of brain inflammatory diseases by delivering exosome encapsulated anti-inflammatory drugs from the nasal region to the brain. *Mol Ther*. (2011) 19:1769–79. doi: 10.1038/mt.2011.164

165. Shahraki K, Boroumand PG, Lotfi H, Radnia F, Shahriari H, Sargazi S, et al. An update in the applications of exosomes in cancer theranostics: from research to clinical trials. *J Cancer Res Clin Oncol*. (2023) 149:8087–116. doi: 10.1007/s00432-023-04701-6

166. Besse B, Charrier M, Lapierre V, Dansin E, Lantz O, Planchard D, et al. Dendritic cell-derived exosomes as maintenance immunotherapy after first line chemotherapy in NSCLC. *Onco Targets Ther*. (2016) 5:e1071008. doi: 10.1080/2162402X.2015.1071008

167. Zeissig MN, Ashwood LM, Kondrashova O, Sutherland KD. Next batter up! Targeting cancers with KRAS-G12D mutations. *Trends Cancer*. (2023) 9:955–67. doi: 10.1016/j.trecan.2023.07.010

168. Han Y, Liu D, Li L. PD-1/PD-L1 pathway: current researches in cancer. *Am J Cancer Res*. (2020) 10:727–42.

169. Qiu Y, Yang Y, Yang R, Liu C, Hsu J-M, Jiang Z, et al. Activated T cell-derived exosomal PD-1 attenuates PD-L1-induced immune dysfunction in triple-negative breast cancer. *Oncogene*. (2021) 40:4992–5001. doi: 10.1038/s41388-021-01896-1

170. Banks WA, Sharma P, Bullock KM, Hansen KM, Ludwig N, Whiteside TL. Transport of extracellular vesicles across the blood-brain barrier: brain pharmacokinetics and effects of inflammation. *Int J Mol Sci*. (2020) 21:4407. doi: 10.3390/ijms21124407

171. Fayazi N, Sheykhhasan M, Soleimani Asl S, Najafi R. Stem cell-derived exosomes: a new strategy of neurodegenerative disease treatment. *Mol Neurobiol*. (2021) 58:3494–514. doi: 10.1007/s12035-021-02324-x

172. Luarte A, Batiz LF, Wyneken U, Lafourcade C. Potential therapies by stem cell-derived exosomes in CNS diseases: focusing on the neurogenic niche. *Stem Cells Int*. (2016) 2016:5736059. doi: 10.1155/2016/5736059

173. Mohamed AS, Abdel-Fattah DS, Abdel-Aleem GA, El-Sheikh TF, Elbatch MM. Biochemical study of the effect of mesenchymal stem cells-derived exosome versus L-Dopa in experimentally induced Parkinson's disease in rats. *Mol Cell Biochem*. (2023) 478:2795–811. doi: 10.1007/s11010-023-04700-8

174. Wang X, Yang G. Bone marrow mesenchymal stem cells-derived exosomes reduce Abeta deposition and improve cognitive function recovery in mice with Alzheimer's disease by activating sphingosine kinase/sphingosine-1-phosphate signaling pathway. *Cell Biol Int*. (2021) 45:775–84. doi: 10.1002/cbin.11522

175. Elia CA, Tamborini M, Rasile M, Desiato G, Marchetti S, Swuec P, et al. Intracerebral injection of extracellular vesicles from mesenchymal stem cells exerts reduced Abeta plaque burden in early stages of a preclinical model of Alzheimer's disease. *Cells*. (2019) 8:1059. doi: 10.3390/cells8091059

176. Ma X, Huang M, Zheng M, Dai C, Song Q, Zhang Q, et al. ADSCs-derived extracellular vesicles alleviate neuronal damage, promote neurogenesis and rescue memory loss in mice with Alzheimer's disease. *J Control Release*. (2020) 327:688–702. doi: 10.1016/j.jconrel.2020.09.019

177. Chen YA, Lu CH, Ke CC, Chiu SJ, Jeng FS, Chang CW, et al. Mesenchymal stem cell-derived exosomes ameliorate Alzheimer's disease pathology and improve cognitive deficits. *Biomedicines*. (2021) 9:594. doi: 10.3390/biomedicines9060594

178. Zhai L, Shen H, Sheng Y, Guan Q. ADMSC Exo-MicroRNA-22 improve neurological function and neuroinflammation in mice with Alzheimer's disease. *J Cell Mol Med*. (2021) 25:7513–23. doi: 10.1111/jcmm.16787

179. Cone AS, Yuan X, Sun L, Duke LC, Vreones MP, Carrier AN, et al. Mesenchymal stem cell-derived extracellular vesicles ameliorate Alzheimer's disease-like phenotypes in a preclinical mouse model. *Theranostics*. (2021) 11:8129–42. doi: 10.7150/thno.62069

180. Tsivion-Visbord H, Perets N, Sofer T, Bikovski L, Goldshmit Y, Ruban A, et al. Mesenchymal stem cells derived extracellular vesicles improve behavioral and biochemical deficits in a phencyclidine model of schizophrenia. *Transl Psychiatry*. (2020) 10:305. doi: 10.1038/s41398-020-00988-y

181. Xie X, Song Q, Dai C, Cui S, Tang R, Li S, et al. Clinical safety and efficacy of allogenic human adipose mesenchymal stromal cells-derived exosomes in patients with mild to moderate Alzheimer's disease: a phase I/II clinical trial. *Gen Psychiatr*. (2023) 36:e101143. doi: 10.1136/gpsych-2023-101143

182. Zheng X, Sun K, Liu Y, Yin X, Zhu H, Yu F, et al. Resveratrol-loaded macrophage exosomes alleviate multiple sclerosis through targeting microglia. *J Control Release*. (2023) 353:675–84. doi: 10.1016/j.jconrel.2022.12.026

183. Casella G, Colombo F, Finardi A, Descamps H, Ill-Raga G, Spinelli A, et al. Extracellular vesicles containing IL-4 modulate neuroinflammation in a mouse model of multiple sclerosis. *Mol Ther*. (2018) 26:2107–18. doi: 10.1016/j.ymthe.2018.06.024

184. Borghi C, Agabiti-Rosei E, Johnson RJ, Kielstein JT, Lurbe E, Mancia G, et al. Hyperuricaemia and gout in cardiovascular, metabolic and kidney disease. *Eur J Intern Med*. (2020) 80:1–11. doi: 10.1016/j.ejim.2020.07.006

185. Tedeschi A, Agostoni P, Pezzuto B, Corra U, Scrutinio D, La Gioia R, et al. Role of comorbidities in heart failure prognosis part 2: chronic kidney disease, elevated serum uric acid. *Eur J Prev Cardiol*. (2020) 27:35–45. doi: 10.1177/2047487320957793

186. Wang L, Liu J, Xu B, Liu YL, Liu Z. Reduced exosome miR-425 and miR-744 in the plasma represents the progression of fibrosis and heart failure. *Kaohsiung J Med Sci*. (2018) 34:626–33. doi: 10.1016/j.kjms.2018.05.008

187. Bohatá J, Horváthová V, Pavlíková M, Stibůrková B. Circulating microRNA alternations in primary hyperuricemia and gout. *Arthritis Res Ther*. (2021) 23:186. doi: 10.1186/s13075-021-02569-w

188. Chen Z, Shi J, Huang X, Yang Y, Cheng Y, Qu Y, et al. Exosomal miRNAs in patients with chronic heart failure and hyperuricemia and the underlying mechanisms. *Gene*. (2025) 933:148920. doi: 10.1016/j.gene.2024.148920

189. Zhao Y, Xu J. Synovial fluid-derived exosomal lncRNA PCGEM1 as biomarker for the different stages of osteoarthritis. *Int Orthop*. (2018) 42:2865–72. doi: 10.1007/s00264-018-4093-6

190. Kolhe R, Hunter M, Liu S, Jadeja RN, Pundkar C, Mondal AK, et al. Gender-specific differential expression of exosomal miRNA in synovial fluid of patients with osteoarthritis. *Sci Rep*. (2017) 7:2029. doi: 10.1038/s41598-017-01905-y

191. Kraus VB, Collins JE, Hargrove D, Losina E, Nevitt M, Katz JN, et al. Predictive validity of biochemical biomarkers in knee osteoarthritis: data from the FNIH OA biomarkers consortium. *Ann Rheum Dis*. (2017) 76:186–95. doi: 10.1136/annrheumdis-2016-209252

192. Cao Y, Liao S, Deng C, Qin H, Li Y. A pH-responsive phase-transition bi-affinity nanopolymer-assisted exosome metabolomics for early screening of osteoarthritis. *Talanta*. (2025) 283:127144. doi: 10.1016/j.talanta.2024.127144

193. Zou Z, Li H, Xu G, Hu Y, Zhang W, Tian K. Current knowledge and future perspectives of exosomes as Nanocarriers in diagnosis and treatment of diseases. *Int J Nanomedicine*. (2023) 18:4751–78. doi: 10.2147/IJN.S417422

194. Wang Q, Cheng S, Qin F, Fu A, Fu C. Application progress of RVG peptides to facilitate the delivery of therapeutic agents into the central nervous system. *RSC Adv*. (2021) 11:8505–15. doi: 10.1039/D1RA00550B

195. Cui GH, Guo HD, Li H, Zhai Y, Gong ZB, Wu J, et al. RVG-modified exosomes derived from mesenchymal stem cells rescue memory deficits by regulating inflammatory responses in a mouse model of Alzheimer's disease. *Immun Ageing*. (2019) 16:10. doi: 10.1186/s12979-019-0150-2

196. Liang Y, Duan L, Lu J, Xia J. Engineering exosomes for targeted drug delivery. *Theranostics*. (2021) 11:3183–95. doi: 10.7150/thno.52570

197. Liu L, Li Y, Peng H, Liu R, Ji W, Shi Z, et al. Targeted exosome coating gene-chem nanocomplex as "nanoscavenger" for clearing α-synuclein and immune activation of Parkinson's disease. *Sci Adv*. (2020) 6:eaba3967. doi: 10.1126/sciadv.aba3967

198. Gomes ER, Souza FR, Cassali GD, Sabino AP, Barros ALB, Oliveira MC. Investigation of the antitumor activity and toxicity of tumor-derived exosomes fused with Long-circulating and pH-sensitive liposomes containing doxorubicin. *Pharmaceutics*. (2022) 14:2256. doi: 10.3390/pharmaceutics14112256

199. Li L, He D, Guo Q, Zhang Z, Ru D, Wang L, et al. Exosome-liposome hybrid nanoparticle codelivery of TP and miR497 conspicuously overcomes chemoresistant ovarian cancer. *J Nanobiotechnol*. (2022) 20:50. doi: 10.1186/s12951-022-01264-5

200. Wang X, Li D, Li G, Chen J, Yang Y, Bian L, et al. Enhanced therapeutic potential of hybrid exosomes loaded with paclitaxel for Cancer therapy. *Int J Mol Sci*. (2024) 25:3645. doi: 10.3390/ijms25073645

201. Lv Q, Cheng L, Lu Y, Zhang X, Wang Y, Deng J, et al. Thermosensitive exosome-liposome hybrid nanoparticle-mediated chemoimmunotherapy for improved treatment of metastatic peritoneal cancer. *Adv Sci*. (2020) 7:2000515. doi: 10.1002/advs.202000515

202. Rayamajhi S, Nguyen TDT, Marasini R, Aryal S. Macrophage-derived exosome-mimetic hybrid vesicles for tumor targeted drug delivery. *Acta Biomater*. (2019) 94:482–94. doi: 10.1016/j.actbio.2019.05.054

203. Zhou X, Zhuang Y, Liu X, Gu Y, Wang J, Shi Y, et al. Study on tumour cell-derived hybrid exosomes as dasatinib nanocarriers for pancreatic cancer therapy. *Artif Cells Nanomed Biotechnol*. (2023) 51:532–46. doi: 10.1080/21691401.2023.2264358

204. Zhang W, Song Q, Bi X, Cui W, Fang C, Gao J, et al. Preparation of *Pueraria lobata* root-derived exosome-like Nanovesicles and evaluation of their effects on mitigating alcoholic intoxication and promoting alcohol metabolism in mice. *Int J Nanomedicine*. (2024) 19:4907–21. doi: 10.2147/IJN.S462602

205. Wang D, Zhang H, Liao X, Li J, Zeng J, Wang Y, et al. Oral administration of *Robinia pseudoacacia* L. flower exosome-like nanoparticles attenuates gastric and small intestinal mucosal ferroptosis caused by hypoxia through inhibiting HIF-1α- and HIF-2α-mediated lipid peroxidation. *J Nanobiotechnol*. (2024) 22:479. doi: 10.1186/s12951-024-02663-6

206. Brogyanyi T, Kejík Z, Veselá K, Dytrych P, Hoskovec D, Masařik M, et al. Iron chelators as mitophagy agents: potential and limitations. *Biomed Pharmacother*. (2024) 179:117407. doi: 10.1016/j.biopha.2024.117407

207. Li J, Cao F, Yin H, Huang Z, Lin Z, Mao N, et al. Ferroptosis: past, present and future. *Cell Death Dis*. (2020) 11:88. doi: 10.1038/s41419-020-2298-2

208. Ahima RS, Lazar MA. The health risk of obesity—better metrics imperative. *Science*. (2013) 341:856–8. doi: 10.1126/science.1241244

209. Matarese G. The link between obesity and autoimmunity. *Science*. (2023) 379:1298–300. doi: 10.1126/science.ade0113

210. Blüher M. Obesity: global epidemiology and pathogenesis. *Nat Rev Endocrinol*. (2019) 15:288–98. doi: 10.1038/s41574-019-0176-8

211. Wang J, Zhang T, Gu R, Ke Y, Zhang S, Su X, et al. Development and evaluation of reconstructed Nanovesicles from turmeric for multifaceted obesity intervention. *ACS Nano*. (2024) 18:23117–35. doi: 10.1021/acsnano.4c05309

212. Zheng M, Chavda VP, Vaghela DA, Bezbaruah R, Gogoi NR, Patel K, et al. Plant-derived exosomes in therapeutic nanomedicine, paving the path toward precision medicine. *Phytomedicine*. (2024) 135:156087. doi: 10.1016/j.phymed.2024.156087

213. Hazas MCLDL, Tome-Carneiro J, Del Pozo-Acebo L, Del Saz-Lara A, Chapado LA, Balaguer L, et al. Therapeutic potential of plant-derived extracellular vesicles as nanocarriers for exogenous miRNAs. *Pharmacol Res*. (2023) 198:106999. doi: 10.1016/j.phrs.2023.106999

214. Gao C, Zhou Y, Chen Z, Li H, Xiao Y, Hao W, et al. Turmeric-derived nanovesicles as novel nanobiologics for targeted therapy of ulcerative colitis. *Theranostics*. (2022) 12:5596–614. doi: 10.7150/thno.73650

215. Zipkin M. Big pharma buys into exosomes for drug delivery. *Nat Biotechnol*. (2020) 38:1226–8. doi: 10.1038/s41587-020-0725-7

216. Greening DW, Xu R, Ji H, Tauro BJ, Simpson RJ. A protocol for exosome isolation and characterization: evaluation of ultracentrifugation, density-gradient separation, and immunoaffinity capture methods. *Methods Mol Biol*. (2015) 1295:179–209. doi: 10.1007/978-1-4939-2550-6_15

217. Onodi Z, Pelyhe C, Terezia Nagy C, Brenner GB, Almasi L, Kittel A, et al. Isolation of high-purity extracellular vesicles by the combination of Iodixanol density gradient ultracentrifugation and bind-elute chromatography from blood plasma. *Front Physiol*. (2018) 9:1479. doi: 10.3389/fphys.2018.01479

218. Shu S, Allen CL, Benjamin-Davalos S, Koroleva M, MacFarland D, Minderman H, et al. A rapid exosome isolation using ultrafiltration and size exclusion chromatography (REIUS) method for exosome isolation from melanoma cell lines. *Methods Mol Biol*. (2021) 2265:289–304. doi: 10.1007/978-1-0716-1205-7_22

219. Cheruvanky A, Zhou H, Pisitkun T, Kopp JB, Knepper MA, Yuen PS, et al. Rapid isolation of urinary exosomal biomarkers using a nanomembrane ultrafiltration concentrator. *Am J Physiol Renal Physiol*. (2007) 292:F1657–61. doi: 10.1152/ajprenal.00434.2006

220. Nordin JZ, Lee Y, Vader P, Mager I, Johansson HJ, Heusermann W, et al. Ultrafiltration with size-exclusion liquid chromatography for high yield isolation of extracellular vesicles preserving intact biophysical and functional properties. *Nanomedicine*. (2015) 11:879–83. doi: 10.1016/j.nano.2015.01.003

221. Ansari FJ, Tafti HA, Amanzadeh A, Rabbani S, Shokrgozar MA, Heidari R, et al. Comparison of the efficiency of ultrafiltration, precipitation, and ultracentrifugation methods for exosome isolation. *Biochem Biophys Rep*. (2024) 38:101668. doi: 10.1016/j.bbrep.2024.101668

222. Contreras H, Alarcón-Zapata P, Nova-Lamperti E, Ormazabal V, Varas-Godoy M, Salomon C, et al. Comparative study of size exclusion chromatography for isolation of small extracellular vesicle from cell-conditioned media, plasma, urine, and saliva. *Front Nanotechnol*. (2023) 5:1146772. doi: 10.3389/fnano.2023.1146772

223. Navajas R, Corrales FJ, Paradela A. Serum exosome isolation by size-exclusion chromatography for the discovery and validation of preeclampsia-associated biomarkers. *Methods Mol Biol*. (2019) 1959:39–50. doi: 10.1007/978-1-4939-9164-8_3

224. Jiao R, Sun S, Gao X, Cui R, Cao G, Wei H, et al. A polyethylene glycol-based method for enrichment of extracellular vesicles from culture supernatant of human ovarian Cancer cell line A2780 and body fluids of high-grade serous carcinoma patients. *Cancer Manag Res*. (2020) 12:6291–301. doi: 10.2147/CMAR.S228288

225. Rider MA, Hurwitz SN, Meckes DG Jr. ExtraPEG: a polyethylene glycol-based method for enrichment of extracellular vesicles. *Sci Rep*. (2016) 6:23978. doi: 10.1038/srep23978

226. Sharma P, Ludwig S, Muller L, Hong CS, Kirkwood JM, Ferrone S, et al. Immunoaffinity-based isolation of melanoma cell-derived exosomes from plasma of patients with melanoma. *J Extracell Vesicles*. (2018) 7:1435138. doi: 10.1080/20013078.2018.1435138

Glossary

AD - Alzheimer's disease

ALS - amyotrophic lateral sclerosis

BBB - blood–brain barrier

CHF - chronic heart failure

DGUC - density gradient ultracentrifugation

DLS - dynamic light scattering

DUC - differential ultracentrifugation

EEs - early endosomes

EGFR - epidermal growth factor receptor

ELISA - enzyme-linked immunosorbent assay

EMT - epithelial–mesenchymal transition

EVs - extracellular vehicles

ExoDOX - exosomes loaded with doxorubicin

ExoDOX-GNPs - exosomes loaded with doxorubicin conjugated with gold nanoparticles

Exo-L - large exosomes

ExoPTX - exosomes loaded with paclitaxel

Exo-S - small exosomes

HUA - hyperuricemia

HUCMSCs-EXOs - mesenchymal stem cell exosomes derived from human umbilical cord

ILVs - intraluminal vesicles

ISEV - International Society for Extracellular Vesicles

LC–MS - liquid chromatography–mass spectrometry tandem technique

LEs - late endosomes

MS - multiple sclerosis

MSCs - mesenchymal stem cells

MSCs-EXOs - mesenchymal stem cell-derived exosomes

MVBs - multi-vesicular bodies

nanoFCM - nanoflow cytometry

NDs - neurodegenerative diseases

NGS - next-generation sequencing

NSCLC - non-small cell lung cancer

NTA - nanoparticle tracking analysis

PD - Parkinson disease

PDAC - pancreatic ductal adenocarcinoma

PD-1 - programmed death-1

PD-L1 - programmed death-ligand 1

P-ELNs - plant-derived exosomes

qRT-PCR - quantitative reverse transcription polymerase chain reaction

RPS - resistive pulse sensing

SEC - size exclusion chromatography

SERS - surface-enhanced Raman spectroscopy

TEXs - tumor-derived exosomes

TFF - tangential flow filtration

TME - tumor microenvironment

TNBC - triple-negative breast cancer

OA - osteoarthritis

UF - ultrafiltration

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 21 March 2025
DOI 10.3389/fgene.2025.1532651

Check for updates

OPEN ACCESS

EDITED BY
Lin Zhang,
China University of Mining and Technology, China

REVIEWED BY
Xiaobo Sun,
Zhongnan University of Economics and Law, China
Conghao Wang,
Nanyang Technological University, Singapore

*CORRESPONDENCE
Rui Miao,
miaorui.research@gmail.com

†These authors have contributed equally to this work

RECEIVED 02 December 2024
ACCEPTED 24 February 2025
PUBLISHED 21 March 2025

CITATION
Miao R, Zhong B-J, Mei X-Y, Dong X, Ou Y-D, Liang Y, Yu H-Y, Wang Y and Dong Z-H (2025) A semi-supervised weighted SPCA- and convolution KAN-based model for drug response prediction.
Front. Genet. 16:1532651.
doi: 10.3389/fgene.2025.1532651

COPYRIGHT
© 2025 Miao, Zhong, Mei, Dong, Ou, Liang, Yu, Wang and Dong. 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 semi-supervised weighted SPCA- and convolution KAN-based model for drug response prediction

Rui Miao[1*†], Bing-Jie Zhong[1†], Xin-Yue Mei[2], Xin Dong[2], Yang-Dong Ou[3], Yong Liang[4], Hao-Yang Yu[1], Ying Wang[1] and Zi-Han Dong[1]

[1]Basic Teaching Department, Zhuhai Campus of Zunyi Medical University, Zhu Hai, China, [2]Institute of Systems Engineering, Macau University of Science and Technology, Macau, China, [3]School of Biomedical Engineering, Guangdong Medical University, Dongguan, China, [4]Peng Cheng Laboratory, Shenzhen, China

Motivation: Predicting the response of cell lines to characteristic drugs based on multi-omics gene information has become the core problem of precision oncology. At present, drug response prediction using multi-omics gene data faces the following three main challenges: first, how to design a gene probe feature extraction model with biological interpretation and high performance; second, how to develop multi-omics weighting modules for reasonably fusing genetic data of different lengths and noise conditions; third, how to construct deep learning models that can handle small sample sizes while minimizing the risk of possible overfitting.

Results: We propose an innovative drug response prediction model (NMDP). First, the NMDP model introduces an interpretable semi-supervised weighted SPCA module to solve the feature extraction problem in multi-omics gene data. Next, we construct a multi-omics data fusion framework based on sample similarity networks, bimodal tests, and variance information, which solves the data fusion problem and enables the NMDP model to focus on more relevant genomic data. Finally, we combine a one-dimensional convolution method and Kolmogorov–Arnold networks (KANs) to predict the drug response. We conduct five sets of real data experiments and compare NMDP against seven advanced drug response prediction methods. The results show that NMDP achieves the best performance, with sensitivity and specificity reaching 0.92 and 0.93, respectively—an improvement of 11%–57% compared to other models. Bio-enrichment experiments strongly support the biological interpretation of the NMDP model and its ability to identify potential targets for drug activity prediction.

KEYWORDS

drug response prediction, feature extraction, sparse PCA, Kolmogorov–Arnold networks, data fusion

1 Introduction

Precision oncology aims to leverage genomic information to identify patient groups with similar biological traits, enabling the delivery of the most suitable treatments (Dlamini et al., 2020; Garraway et al., 2013; Hodson, 2020; Prasad, 2016; Prasad et al., 2016). In clinical applications, this approach generally involves choosing targeted therapies based on the individual genomic profiles of patients (Ballester and Carmona, 2021). However, research reveals that only approximately 9% of patients experience effective outcomes from such targeted treatments, which greatly restricts the broad applicability of precision oncology (Barretina et al., 2012a; Rubio-Perez et al., 2015). Moreover, limited drug response prediction models for non-specific therapies mean that many patients miss out on the benefits of precision oncology and may even receive ineffective treatments. Fortunately, data from extensive pharmacogenomic screenings have shown that nearly all cancer cell lines and patient-derived xenografts (PDXs) respond to some form of targeted therapy or non-specific chemotherapy (Barretina et al., 2012b; Gao et al., 2015; Garnett et al., 2012). Thus, a primary challenge now is accurately aligning cancer patients with treatments that match their unique drug response profiles.

Currently, a significant research focus is predicting drug responses in cancer patients using single genomics data (Adam et al., 2020; Dong et al., 2015; Firoozbakht et al., 2022; Sheng et al., 2015). For instance, as demonstrated by Geeleher et al., a ridge regression model that utilizes gene expression data from the Genomics of Drug Sensitivity in Cancer (GDSC) database has shown effective application to clinical trial datasets for drugs including erlotinib, cisplatin, docetaxel, and bortezomib. The study also found that incorporating data from cancer cell lines other than breast cancer can improve the predictive performance of the docetaxel drug response model (Geeleher et al., 2014). Moreover, our preliminary research indicates that combining statistical methods based on individual genomic information from patients with machine learning techniques can construct highly performant drug response prediction models (Miao et al., 2020; Zheng et al., 2024; Sharma et al., 2024).

Recently, the increasing availability of multi-omics datasets for drug response has opened new avenues for machine learning models, enabling a deeper understanding of biological processes. Multi-omics data have shown notable success across various bioinformatics tasks, including survival prediction, cancer subtype classification, and target gene identification (Xu et al., 2024). As deep learning continues to progress rapidly, constructing predictive models that utilize multi-omics data through deep learning techniques becomes a primary research focus. Several multi-omics drug response models have been developed (Ballester et al., 2022; Baptista et al., 2021; Chen and Zhang, 2021; Zhou et al., 2024; Rashid, 2024; Baptista and Ferreira, 2023). For instance, Chiu et al. developed a deep learning model that utilizes autoencoders to combine diverse omics features for drug response prediction (Chiu et al., 2019). Similarly, Hossein et al. proposed a model that employs deep neural network fusion, combining hidden layer representations from different multi-omics networks to synthesize feature information effectively (Sharifi-Noghabi et al., 2019). Peng et al. proposed a two-space graph convolutional neural network (TSGCNN) that combines cell line and drug feature spaces to predict drug responses by leveraging both homogeneous and heterogeneous relationships (Peng et al., 2023). Similarly, Trac et al. proposed a GCN-based drug response prediction model for acute myeloid leukemia (AML), highlighting the versatility of graph-based neural networks in oncology research (Trac et al., 2023). Wang et al. proposed MOICVAE, a deep learning model that integrates multi-omics data using a variational autoencoder to improve drug sensitivity prediction (Wang et al., 2023). Meanwhile, Sharma et al. proposed DeepInsight-3D, architecture to fuse multi-omics data for anticancer drug response prediction, offering an advanced deep learning perspective for modeling complex interactions in diverse biological datasets (Sharma et al., 2023).

Currently, the multi-omics drug response prediction model faces three major challenges. First, genomic data typically involve small sample sizes, which increases the likelihood of overfitting in existing models (Deng et al., 2023). Developing an efficient and biologically interpretable feature selection method to select key genomic data is the first major challenge currently faced (Deng et al., 2023). Second, most genomic datasets for drug response prediction contain multiple independent genomic data types (Munquad et al., 2024). The data lengths and noise levels of these genomic datasets vary significantly, making the rational design of the multi-omics fusion method the second major challenge in constructing high-performance drug response medical models. Third, considering that drug response prediction is a complex biological problem and the dataset has only limited training samples, constructing a sufficiently high-performance prediction model based on a small sample of data remains the third major challenge.

In this study, we introduce an innovative model for predicting drug response (NMDP, Figure 1). The NMDP model is composed of four main modules. 1) Key genome selection module: we propose an interpretable, semi-supervised weighted sparse PCA to identify essential biological features. 2) Similarity network construction module: this module addresses the challenge of aligning data across different omics. 3) Data fusion module: we introduce a weighted similarity network fusion approach, incorporating the dip test method and variance information. 4) Drug response prediction module: we integrate one-dimensional convolutional neural networks (CNNs) and the Kolmogorov–Arnold network (KAN) method.

2 Materials and methods

2.1 Datasets

In this study, we use publicly available datasets to extract drug response and genomic data from cell lines. The first dataset is Genomics of Drug Sensitivity in Cancer (GDSC), which provides extensive data on drug response measurements. The second is the Gene Expression Omnibus (GEO) and Cell Model Passports, along with the European Bioinformatics Institute (EMBL-EBI) datasets. These two datasets provide the genomic data needed for this experiment.

It is important to note that the GDSC dataset comprises 624 unique drugs, 576,758 IC_50 values, and 978 cell lines.

10.3389/fgene.2025.1532651

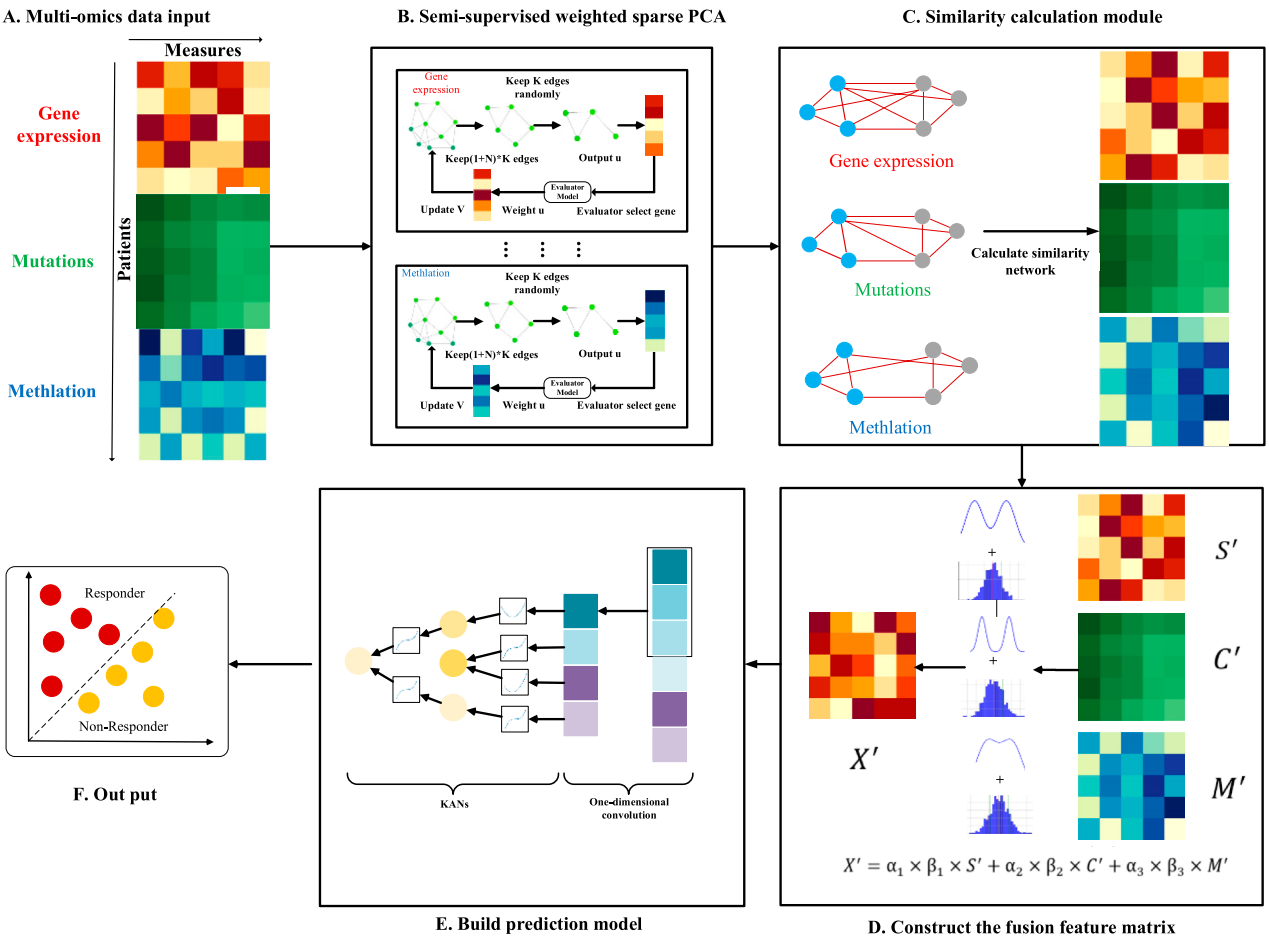


FIGURE 1
Overview of the NMDP model workflow. **(A)** Multi-omics data input. **(B)** Semi-supervised weighted sparse PCA. **(C)** Similarity calculation module. **(D)** Construct of the fusion feature matrix. **(E)** Prediction model. **(F)** Output.

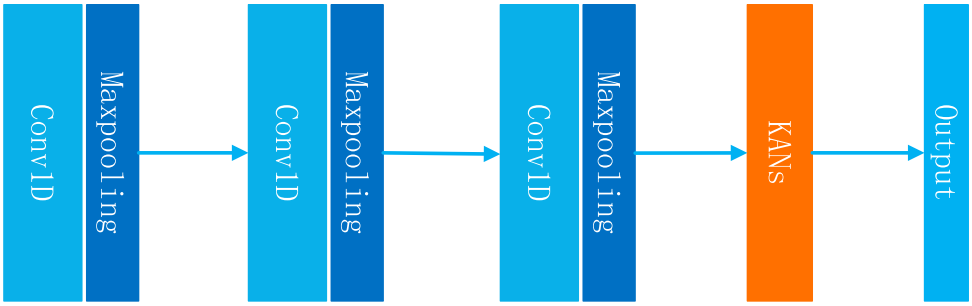


FIGURE 2
Structure of the deep learning model used in the NMDP model.

Genomic characteristics for each cell line include somatic copy number alterations (SCNAs) across 21,878 genes, RNA-Seq expression levels for 44,421 probes, and methylation levels for 365,860 CpG sites. For our study, we select 68 drugs: 14 FDA-approved targeted therapies, 49 drugs with known target genes not yet FDA-approved, and 5 nonspecific treatments (Supplementary Tables S1–S3).

2.2 Dataset of gene pathway data

The pathway data used in this study are sourced from the Pathway Commons database, which contains commonly used pathway datasets such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO).

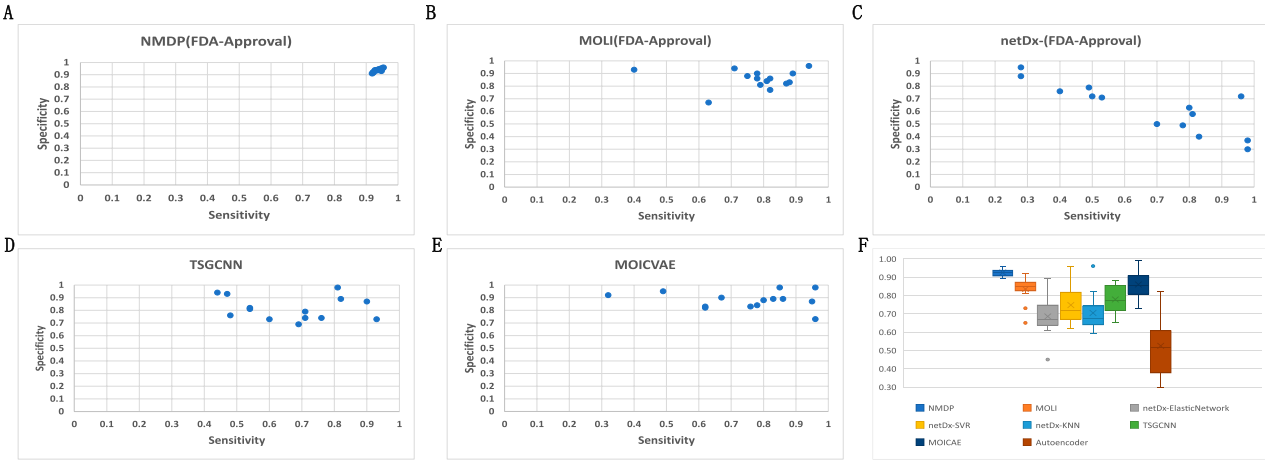


FIGURE 3
Results of 14 FDA-approved drugs for each model. **(A)** Sensitivity and specificity of the NMDP model; **(B)** sensitivity and specificity of the MOLI model; **(C)** sensitivity and specificity of the netDx model; **(D)** sensitivity and specificity of the TSGCNN model; **(E)** sensitivity and specificity of the MOICVAE model; and **(F)** accuracy of each model.

2.3 Drug response data

In addition to the preprocessing already performed by the provider of this dataset, we also perform additional preprocessing. The following are the steps and criteria of our preprocessing: first, we remove samples with certain missing data, such as samples with a missing rate of more than 10%; second, we remove drugs with limited IC_50 test information, requiring the amount of IC_50 test data for each drug to be no less than 200 samples. Third, we use waterfall distribution to divide drug response data (Ding et al., 2018). Waterfall distribution is a method that sorts drugs based on their IC_50 values and uses a linear model to fit the data, which is used to determine whether a drug is effective. Specifically, the drugs are sorted according to the true IC_50 information. A linear model is then constructed to fit the distribution, and Pearson's coefficient is used to evaluate the degree of fit of the model. If the fit is higher than 0.95, then the median is chosen as the cut-off point. If the fit is less than 0.95, a new monadic linear function is created, and the parameters of the function are determined by the smallest and largest points of IC_50. Finally, the point furthest away from the unary linear function in the IC_50 curve is calculated as the demarcation point. Ultimately, we classify divide drugs into two categories: responsive and non-responsive. In addition, to ensure that the data are balanced, we ensure that the response group constitutes at least 25% of the total data.

2.4 Methods

In this section, we provide a detailed overview of the architecture and algorithm flow of the NMDP model. This NMDP method transforms the sparse PCA model from a non-supervised to a semi-supervised approach, improving the ability of feature selection (key genome selection module). Second: Similarity network building module: in this module, we construct a sample similarity network based on the Spearman and Kendall correlation coefficients. Third: Data fusion module: we develop a data fusion algorithm based on the dip and variance tests. Fourth: Drug response prediction module: in this module, we propose a drug response prediction model based on one-dimensional convolution and KANs.

2.4.1 Key genome selection module

2.4.1.1 ESPCA method

Before introducing the NMDP model, we first define the sparse PCA (SPCA) and edge sparse PCA (ESPCA) models. Suppose we have an $m \times n$ feature matrix $X \in R^{m,n}$, where n represents the number of samples and m represents the number of gene probes. The definition of SPCA is given by Formula 1:

$$\underset{\|u\|_2 \le 1}{maximize}\, u^T X X^T u, s.t. \|u\|_0 \le s. \tag{1}$$

Here, $\|^*\|_2$ and $\|^*\|_0$ represent L_2 and L_0 norms, respectively. u represents principal component (PC) loading, which has the dimension as the number of gene probes. s represents the retention number of gene probes. In most cases, the SVD method is used to solve Formula 1. Therefore, the formula can also be written as Formula 2:

$$\underset{\|u\|_2 \le 1, \|v\|_2 \le 1}{maximize}\, u^T X v, s.t. \| u \|_0 \le s. \tag{2}$$

In this case, v represents the weight information corresponding to the sample, with dimensions matching the number of samples. ESPCA builds upon SPCA by incorporating improvements. Its main contribution is the integration of pathway structure information from the genome as *a priori* knowledge. Suppose that the known pathway structure information (edge set) is represented as $\mathcal{G} = \{e_1, \ldots, e_l\}$. At this point, the researcher introduces $\|u\|_{ES}$ regulon, which is represented as Formula 3:

$$\|u\|_{ES} = \underset{\forall \mathcal{G}' \in \mathcal{G}, support(u) \le V(\mathcal{G}')}{minimize} \left|\mathcal{G}'\right|. \tag{3}$$

Here, $\mathcal{G}' \in \mathcal{G}$ and $V(\mathcal{G}')$ represents the vertex set derived from the $\|u\|_{ES}$ regulon. Therefore, ESPCA can also be represented as Formula 4 (Min et al., 2018):

10.3389/fgene.2025.1532651

TABLE 1 Results of each model of the 14 drugs approved by the FDA.

		NMDP	MOLI	Deep autoencoder	netDx−KNN	netDx−ElasticNet	netDx−SVR	TSGCNN	MOICVAE
Accuracy	All	0.93	0.84	0.64	0.70	0.71	0.74	0.78	0.86
F1 score	Responsive	0.92	0.83	0.48	0.72	0.67	0.71	0.74	0.74
	Non-responsive	0.92	0.83	0.44	0.56	0.58	0.66	0.78	0.86
	All	0.92	0.83	0.46	0.64	0.63	0.68	0.77	0.80

$$\underset{\|u\|_2 \le 1, \|v\|_2 \le 1}{maximize}\, u^T Xv, s.t. \| u \|_{ES} \le s. \tag{4}$$

2.4.1.2 Semi-supervised weighted edge sparse PCA

The existing SPCA and ESPCA methods are pure non-supervised methods; this method has a great advantage in data analysis with small samples and high dimensions. However, two primary issues arise: first, the method cannot utilize existing grouping information, which may reduce its effectiveness. Second, for the problem of drug response, the existing sparse PCA method selects the exact same key gene probe for all types of drugs; it obviously does not accord with the common sense of biology. In this study, we propose a novel semi-supervised weighted edge sparse PCA. This method mainly includes a weighted parameter t, which is calculated using a machine learning model. The parameter t leverages known grouping information on drug responses. Each time the model completes a cycle, we calculate t based on the currently selected key gene probes and weight u. Finally, we can select different key gene probes for each drug. The specific steps are shown in Formulas 5–12.

In general, the semi-supervised weighted edge sparse PCA method proposed in this paper can be expressed as Formula 5:

$$\underset{\|u\|_2 \le 1, \|v\|_2 \le 1}{maximize}\, u^T Xv, s.t. \| u \|_{NM} \le k. \tag{5}$$

Here, $\|u\|_{NM}$ is a sparse regulon representing the edge group proposed by ESPCA and k is the regularization parameter. The regulon is given by Formula 6:

$$\|u\|_{NM} = \underset{\forall \mathcal{G}_w^{'} \in \mathcal{G}_w, support(u) \subseteq V(\mathcal{G}_w^{'})}{minimize} \left|\mathcal{G}_w^{'}\right|. \tag{6}$$

Here, $\mathcal{G}_w^{'}$ represents a subset of vertices selected from the edge set, with $|\mathcal{G}_w^{'}|$ representing the count of vertices within this subset. Additionally, support (u) represents the collection of non-zero elements in the sparse vector u. Then, we specifically explain how to calculate $\mathcal{G}_w^{'}$, supposing $e_h = (u_i, u_j) \in \mathcal{G}, u_i, u_j \in R^m$. At the beginning of the algorithm, v is randomly initialized. We use $u = Xv$ to calculate the weight of u. Based on u, we use Formula 7 to calculate the edge weight w_h corresponding to e_h:

$$w_h = \sqrt{u_i^2 + u_j^2}. \tag{7}$$

Finally, the edge weight can be represented as $\mathcal{G}_w = \{w_h\}_1^l$. In this paper, we used a greedy principle based on the random sampling method, previously developed by our team, to sparsify u, as represented in Formula 8 (Miao et al., 2022):

$$\left[\mathcal{P}_{\mathcal{G}_w}(z,k)\right]_i = \begin{cases} z_i, if\ \mathcal{G}_w(i) \cap sample(I,k) \neq \varnothing \\ 0, otherwise \end{cases}. \tag{8}$$

Here, $\mathcal{P}_{\mathcal{G}}(z,k)$ represents the sparse projection, with $[\mathcal{P}_{\mathcal{G}_p}(z,k)]_i (i = 1, \ldots, m)$. $I = supp(norm_{\mathcal{G}_p}^{DM}(e'), (1+\omega) \times k)$. $sample(I,k)$ represents the random selection of k elements from the set I. k denotes the number of non-zero elements selected in the sparsification process. If gene i is selected, then $[\mathcal{P}_{\mathcal{G}}(z,k)]_i = z_i$; otherwise, $[\mathcal{P}_{\mathcal{G}}(z,k)]_i = 0$.

In this case, we can obtain a sparse gene weight vector $\hat{u} = \mathcal{P}_{\mathcal{G}_w}(z,k)$. Since existing sparse PCA models are non-supervised, identical input gene expression information results in the same $\hat{u}$ for each drug. In order to find a more suitable key gene

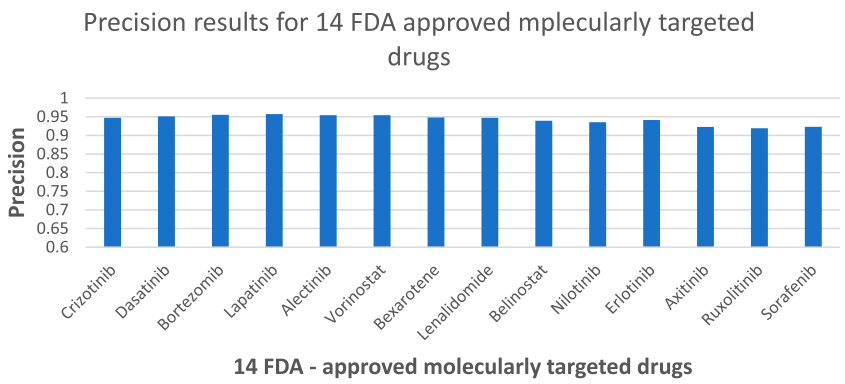


FIGURE 4
NMDP model precision results of 14 FDA-approved drugs.

TABLE 2 External independent validation results.

	Precision	Recall	F1-score
Responsive	0.72	0.74	0.73
Non-responsive	0.83	0.81	0.82
Accuracy			0.77
Macro average	0.73	0.77	0.74
Weighted average	0.8	0.78	0.79

set for different drugs, we design a linear evaluator based on machine learning, denoted as $\hat{y} = f_\theta(x)$. For example, linear models or random forests can be used as evaluators. Here, θ represents the parameterized model, while the classification label corresponds to the drug response grouping information. Each time the sparse PCA model completes a cycle, we extract a new genome key expression matrix $\hat{X} \in R^{p,n}$ and $\hat{X} \in X$ based on $\hat{u}$, where p represents the number of non-zero gene probes contained in $\hat{u}$ at that time. Next, $\hat{X}$ is input into $\hat{y}$, as represented in Formula 9:

$$\theta^* = \arg\min_\theta \mathcal{L}\left(\hat{X}; \theta, \omega\right). \tag{9}$$

Here, $\mathcal{L}$ represents the loss function. ω represents the optimizer of the model. θ^* represents the parameter after model training. Once training is complete, an importance score t is calculated for each gene probe associated with $\hat{X}$. Finally, we obtain $t = \{t_1, \ldots, t_p\}$. In order to ensure the stability in the weighting process, we perform a normalization step on t,scaling the values to the range 0–2. Finally, we update $\hat{u}$ based on t, as represented in Formula 10:

$$\hat{u} = \{t_1\hat{u}_1, \ldots, t_m\hat{u}_m\}. \tag{10}$$

In this case, if the gene probe corresponding to t_i is not included in the set of t, then $t_i = 0$. We use Formulas 11, 12 to cross-update u:

$$u \leftarrow \frac{\hat{u}}{\|\hat{u}\|}, \tag{11}$$

$$v \leftarrow \frac{\hat{v}}{\|v\|}, where\ \hat{v} = X^T u. \tag{12}$$

1: $u = Xv$
2: for any weight of edge e in $\mathcal{G}_w$ do
3: $w_h' = \sqrt{u_i^2 + u_j^2}$ *Generate a dynamic network.*
4: *update* $\mathcal{G}_w' = w_h'$
5: ***end for***
6: Let $\text{norm}_{\mathcal{G}_w'}^{NM}(e') = (\| e_1' \|, \cdots, \| e_1' \|)^T$
7: $I = \text{supp}\,(\text{norm}_{C_n}^{NM}(e'), (1+\omega) \times k)$ Extract $(1+\omega) \times k$ edges
8: $J_k = \text{sample}\,(I, k)$
9: if $\omega > 0$ then $\omega = \omega - \rho$
10: $V_{\mathcal{G}_w'} = V(\mathcal{G}_w')$
11: for any gene i in $V_{\mathcal{G}_w'}$ do
12: $\hat{u}_i = u_i$
13: end for
14: $\hat{X} = X[\hat{u}]$
15: Class = RandomForestClassifier ()
16: Class . fit (Class, Y)\#Train and test the classifier
17: t = Class . feature importances
18: $\hat{u} = t * \hat{u}$ # Weight $\hat{u}$
19: return $\hat{u}$
20: $u_{update} = \frac{\hat{u}}{\|\hat{u}\|}$
21: $v \leftarrow \frac{\hat{v}}{\|\hat{v}\|}, where\ \hat{v} = X^T u_{update}$
22: $loss = \|u - u_{update}\|_2, if\ loss < 0.0001, end, then\ return\ step\ 1$

Algorithm 1. Semi-supervised weighted edge SPCA.

2.4.2 Similarity network building modules

For the same drug, we can get at least three different genomics data. The experiments in this paper mainly include gene expression data, copy volume data, and methylation data. Each omics performs sparse PCA operations independently. Finally, we can obtain three key feature matrices, namely, $S \in R^{k,n}$, $C \in R^{p,n}$, and $M \in R^{h,n}$. k, p, h represents the number of key gene probes retained by each of the three omics.

Because of the inconsistency of the data lengths for each omics, data cannot be aligned. Therefore, we calculate a sample similarity subnet for each omics based on the concept of the sample similarity network. In this paper, we use two similarity measurement methods. We use the Spearman correlation coefficient to calculate the sample similarity subnet of gene expression and methylation omics, as represented in Formula 13:

$$\rho_2 = \frac{\sum_1^k (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_1^k (x_i - \bar{x})^2 \sum_1^k (y_i - \bar{y})^2}}. \quad (13)$$

Here, for the *x and y*th sample, x_i, y_i represent the expression information on the *i*th gene expression in each sample, with k representing the total number of gene probes. The symbols $\bar{x}, \bar{y}$ indicate the average gene expression levels for each sample.

The Kendall correlation coefficient is used for the copy number dataset, as provided in Formula 14:

$$Tau = \frac{C - D}{\frac{1}{2}k(k-2)}. \quad (14)$$

Here, the *x and y*th sample can be showed as a set of two elements containing p gene probe. C represents the number of consistent elements. D represents the number of inconsistent elements. k denotes the total number of gene probes in each sample.

2.4.3 Data fusion module

After completing the construction of the sample similarity subnet, we can obtain three feature matrices, namely, $S' \in R^{n,n}$, $C' \in R^{n,n}$, and $M' \in R^{n,n}$. Then, we propose a subnet fusion algorithm based on the dip test and variance estimation using Formula 15:

$$X' = \alpha_1 \times \beta_1 \times S' + \alpha_2 \times \beta_2 \times C' + \alpha_3 \times \beta_3 \times M', \quad (15)$$

where X' represent the feature representation after fusion. α and β are defined as the amount of statistical information corresponding to the genomics data matrix. Theoretically, our goal is to retain as much of the feature matrix as possible, prioritizing genomics with higher statistical significance for drug response prediction. To achieve this, we use two statistical methods to assess the amount of information in the data. The first method is the single peak test, which aims to retain similarity matrices that exhibit more typical bimodal distributions. A bimodal distribution is a statistical concept that represents a dataset in more than two regions. In gene expression analysis, if the data show a bimodal part, it indicates a significant statistical difference within the sample. In this paper, we assess the bimodal property of data using the dip test method, originally proposed by Hartigan et al. (1985). We assume that $\rho(F, G)$ follows Formula 16 for any bounded functions F, G. Let μ be the class of unimodal distribution functions.

$$\rho(F, G) = sup_x |F(x) - G(x)|. \quad (16)$$

We define μ as a typical unimodal distribution function and F as a dip distribution function. We can obtain Formulas 17, 18 as follows:

$$D(F) = \rho(F, \mu), \quad (17)$$

$$D(F_1) \leq D(F_2) + \rho(F_1, F_2). \quad (18)$$

It is important to note that $D(F) = 0$ for $F \in \mu$, indicating that the dip quantifies deviation from unimodality. Assume that the result of the dip function is p, as shown in Equation 19:

$$\begin{matrix} p > 0.95: \ significant\ unimodality \\ p < 0.05: \ significant\ bimodality \end{matrix}. \quad (19)$$

Another statistical method is the variance test. In addition to information about the probability distribution of the samples, our goal is to retain a matrix of sample similarity features that preserves as much discrete information as possible. The formula for the variance S information is provided in Formula 20:

$$S = \frac{\sum (X - \bar{X})^2}{n - 1}, \quad (20)$$

where X is the variable, $\bar{X}$ is the sample mean, and n denotes the sample size. Suppose that the result of the variance of the *i* feature of the i_{th} histology is S_{ij}. Then, β_i of i_{th} histology can be expressed as Formula 21:

$$\beta_i = \frac{1}{n} \sum_1^n S_{ij}. \quad (21)$$

The computed $\beta = \{\beta_1, \beta_2, \beta_3\}$ accounts for the possibility that β having a large parameter. Therefore, we normalize β using Formulas 22, 23 as follows:

$$w_i = a + p(k_i - Min), \quad (22)$$

$$p = (b - a) / (Max - Min). \quad (23)$$

Here, a and b are user-defined parameters, representing the normalized range of data. p represents a scaling factor used to normalize the raw data β to a user-defined range. Max and Min represent the maximum and minimum values of β, respectively.

2.4.4 Drug response prediction module

Finally, we obtain the feature matrix X'. Although the problem of the high dimensionality of data has been largely alleviated after genomic feature extraction and similarity network computation, researchers still need a powerful enough deep learning model to achieve high performance and avoid overfitting. However, considering the limitation of the number of samples, researchers still need a sufficiently powerful drug response prediction deep learning model to avoid model overfitting. In this study, we construct a deep learning model based on one-dimensional convolution and KANs to predict drug response (Figure 2). One-dimensional convolution can further localize the features of the samples and remove potential noise. Experimental results indicate that one-dimensional convolution significantly enhances the model's prediction performance. KANs, proposed by Liu et al. (2024), aim to replace the traditional fully connected neural network layer. The network is based on the Kolmogorov–Arnold theorem, which states that any continuous function $f(x)$ in n-dimensional real space, where $x = (x_1, \ldots, x_n)$ can be represented as a combination of a single-variable continuous function h and a series of continuous bivariate functions g_i and $g_{i,j}$. Specifically, the theorem is expressed in Formula 24:

$$f(x_1, \ldots, x_n) = \sum_{q=0}^{2n} h\left(\sum_{i=1}^{n} g_{q,i}(x_i) \right). \quad (24)$$

The theorem shows that even a complex function in a high-dimensional space can be reconstructed using a series of lower-dimensional function operations. Specifically, a KAN layer with n_{in} dimensional inputs and n_{out} dimensional outputs can be defined as a matrix of one-dimensional functions, as represented in Formula 25:

$$KANs = \{\phi_{q,p}\}, p = 1, 2, \ldots, n_{in}, q = 1, 2, \ldots, n_{out}, \quad (25)$$

where the function ϕ is defined as shown in Formula 26 and consists of a B-spline curve and a residual activation function $b(x)$, all multiplied by a learnable parameter w. The function ϕ is defined as shown in Formula 26:

$$\phi = w_1 \times spline(x) + w_2 b(x). \quad (26)$$

The main advantage of KANs is that they can achieve results beyond fully connected neural networks while using fewer parameters. This is especially important for the drug response prediction problem. Due to the limitation of the sample size, it is unlikely that we can construct a deep learning model that contains a huge neural network. To summarize, the module can be expressed using Formulas 27, 28 as follows:

$$\acute{X} = One - Dimensional\, Convolution(X'), \quad (27)$$

$$out = KANs(\acute{X}). \quad (28)$$

3 Results

The procedure in this article consists of six distinct steps. Initially, experiments were performed using 14 FDA-approved targeted therapy drugs already authorized for clinical use. In the second step, we broadened the model evaluation by testing it with 49 targeted therapy drugs not approved by the FDA. In the third step, we conducted experiments on five chemotherapeutic agents (non-targeted therapeutics) in order to verify that the NMDP model has good scalability. We used seven state-of-the-art AI models for comparison, namely, TSGCNN, MOICVAE, MOLI, netDx, netDx–elastic network, deep autoencoder, and netDx–SVR. Five evaluation indicators were used, namely, sensitivity, specificity, precision, accuracy, and F1 score. The details of comparison models are provided in Supplementary Materials.

In the fourth step, we selected the GDSC1 dataset for training and testing and the GDSC2 dataset for validation. We selected 14 FDA-approved drugs to perform and calculate the mean value. The consistency and reliability of the results were ensured by calculating the mean value.

The fifth step included conducting ablation experiments to determine the importance of each sub-module of the NMDP model. We randomly selected 10 drugs for analysis and averaged the results. All experiments were performed using the GDSC2 dataset (Supplementary Table S7). We conducted four independent experiments: the first experiment was conducted to remove the multi-omics weighting module; the second experiment, to remove the convolution module; the third experiment, to remove the sample similarity network; and the last experiment, to replace KANs with MLPs.

Ultimately, we used the Metascape platform to examine the biological pathways associated with the gene probes chosen by the NMDP model (Zhou et al., 2019). The details of the indicators are provided in Supplementary Materials.

3.1 FDA-approved targeted therapy drugs

Based on the results presented in Figure 3 and Supplementary Table S4, experiments show that the NMDP model is much better than the advanced deep learning model. Notably, the NMDP model achieves an average sensitivity of 0.92 and a specificity of 0.93. Among the models for comparison, the MOICVAE model ranks highest, with a sensitivity of 0.77 and a specificity of 0.91. The deep autoencoder model, however, performs the lowest, with sensitivity and specificity values of 0.53 and 0.44, respectively.

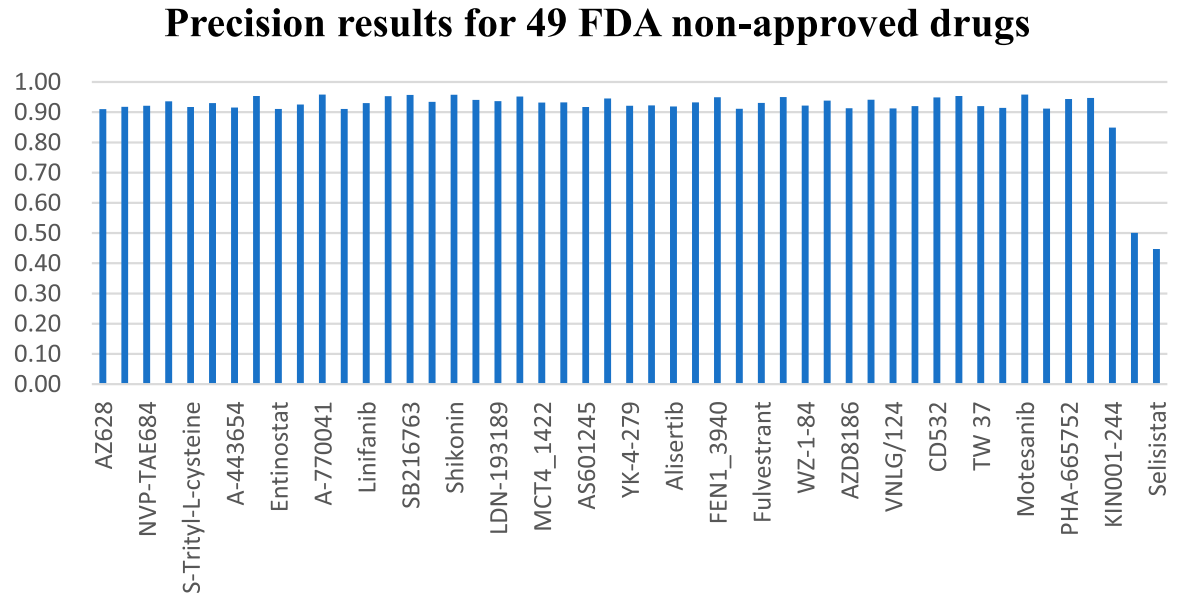


FIGURE 5
NMDP model precision results of five non-specific therapeutic drugs.

Based on Figure 3C, it is evident that the deep autoencoder model exhibits overfitting across multiple drugs. Moreover, the NMDP model demonstrates minimal fluctuation across 14 drugs, indicating its superior stability (Figure 3F). Compared to other models, all except the MOLI model show relatively high levels of fluctuation, suggesting weaker stability in those models. The NMDP model achieves values of 0.93, 0.92, 0.92, and 0.92 for average accuracy and F1 score, outperforming the MOICVAE model by 10% in each metric (Table 1). When compared to the deep autoencoder model, it shows improvements of 31%, 48%, 52%, and 50%, respectively.

According to Figure 4, the NMDP model demonstrated excellent performance in predictive accuracy. It is worth mentioning that out of these 14 drugs, the prediction accuracy for 13 of them exceeded 90%.

3.2 FDA non-approved targeted therapy drugs

Across the 49 drugs not approved by the FDA, we observe similar outcomes. Experimental findings indicate that the NMDP model achieves average sensitivity and specificity values of 0.92 and 0.93, respectively, outperforming the comparison models by 11%–57% (Supplementary Figure S1; Supplementary Table S5). Supplementary Figure S1F illustrates the NMDP model's high stability, with only 5 out of the 49 drugs showing precision below 0.9 (see Figure 5). In addition, the average precision of the NMDP method can reach 0.95. F1 score can reach 0.92, surpassing the MOLI model by 14%, 10%, and 13% (Supplementary Table S1). Among the seven models compared, MOLI achieves the highest performance. Nevertheless, its sensitivity, specificity, precision, and accuracy for response, non-response, and all samples are only 0.80, 0.88, and 0.86, respectively, with F1 scores of 0.83, 0.83, 0.83, 0.82, and 0.84.

3.3 Non-specific therapeutic drugs

In the study of five non-specific therapeutic drugs, we achieve optimal outcomes in three experiments. Results indicate that the NMDP model achieves 0.93 for average sensitivity, surpassing the comparison models by 19%–42% (Supplementary Figure S2; Supplementary Table S6). Additionally, the NMDP model demonstrates a precision close to 0.95 across the five drugs (Figure 6). The model also shows outstanding performance in F1 score and accuracy, reaching 0.93 (Supplementary Table S2). Compared to the other models, the MOLI model achieves the best results, but its average F1 score and accuracy reach only 0.78. The experimental results show that the NMDP model has good expansibility.

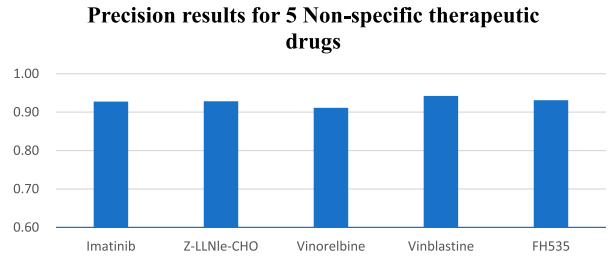


FIGURE 6
NMDP model precision results of five non-specific therapeutic drugs.

3.4 External independent validation results

To evaluate the model's performance and test its generalization ability, we design this external independent validation experiment. The experimental outcomes demonstrate that the NMDP model exhibits superior generalization capabilities. Specifically, the NMDP model achieves an overall prediction accuracy of 0.77 and precision, recall, F1 score, and accuracy of 0.73, 0.77, 0.74, and 0.77, respectively (Table 2). It is worth noting that a slight decrease in accuracy is observed on the validation set compared to the results on the test set. This may be due to the noise difference between the datasets. Overall, the NMDP model exhibits robustness and reliability, with the capability for widespread application.

3.5 Ablation experiment

The experimental results show that the model feature extraction effect is weakened by removing the multi-omics weighting module and the convolution module, but the convolution module has a greater impact on the model. The sample similarity network module has the greatest impact, further verifying the importance of similarity across samples. When KANs are replaced with MLPs, the performance of the model improves but still does not surpass that of the original model. This indicates that KANs have a unique advantage in capturing complex relationships, especially when dealing with multi-omics data. Taken together, the results of the ablation experiments fully indicate that the sample similarity network and convolution module are the key factors in improving the model performance. Among them, the sample similarity network module has the greatest impact, and we believe that the main reason is that, even after the feature filtering of the sparse PCA model, the three modules are still able to save more than 9,000 gene probes collectively, and the excessively high data dimensions make it easy for the model to fall into an overfitting state.

3.6 Enrichment analysis

To validate the biological interpretability of the NMDP model, we conduct bio-enrichment analysis using gene selection results for erlotinib across different omics types obtained from the first principal component (PC) in the NMDP model. Erlotinib is an FDA-approved non-small cell lung cancer drug, with EGFR as its primary target (Tsao et al., 2005; Zhou et al., 2021). The analysis yields promising results as the NMDP model successfully identifies lung cancer-related pathways and gene probes. For instance, in copy number omics, we discover pathways corresponding to genes like *EGFR*, *CDK6*, *RASSF5*, *BRAF*, and *CCND1*, which are associated with non-small cell lung cancer (Chen et al., 2018; Xue et al., 2019; Zhao et al., 2018) (Figure 7A). In methylation omics, we observe the developmental process pathway GO:0032502, involving genes such as *EGFR*, *FGFR2*, *GATA6*, *ASCL1*, *BMP4*, and *FOXA1*, indicating relevance to

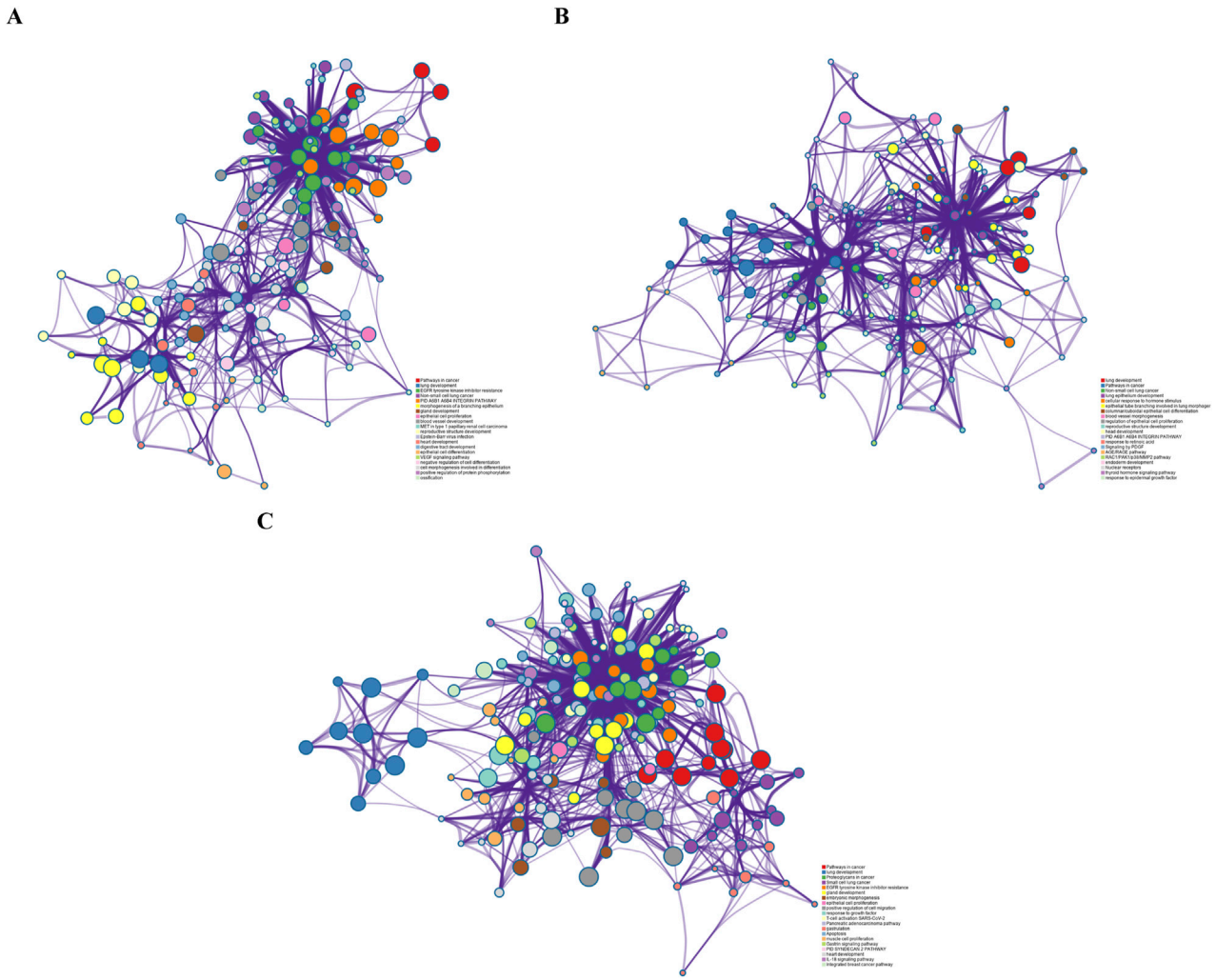


FIGURE 7
(A) Pathway results from the first PC of copy number; **(B)** pathway results from the first PC of methylation data; and **(C)** pathway results from the first PC of sequencing data.

lung development (Bach et al., 2018; Ju et al., 2019; Murai et al., 2015) (Figure 7B). In Seq omics, we identify pathways related to the KEGG pathway, which include genes like *EGFR*, *MAPK1*, *KRAS*, *CCND1*, *HRAS*, *NRAS*, *PLCG1*, and *GRB2*, which show strong connections to non-small cell lung cancer (Betticher et al., 1996; Park et al., 2020; Pązik et al., 2021) (Figure 7C). Remarkably, EGFR, the target gene of erlotinib, is consistently identified across these three omics types.

Overall, the NMDP model demonstrates superior performance compared to other models across all metrics.

4 Discussion

With advancements in bioassay technology, a growing number of large-scale drug response datasets are being released, creating new possibilities for building drug response prediction models. In recent years, researchers have proposed a number of AI-based drug response prediction models (Chiu et al., 2020). However, drug response prediction data are mostly characterized by the typical features of multi-omics, small samples, high dimensionality, and high noise. Designing feature selection and multi-omics fusion methods based on regularization ideas becomes very important. In addition, considering the potential overfitting problem, it is difficult for researchers to build AI-based drug response prediction models with many parameters.

In light of the aforementioned issues, we propose the NMDP model, which integrates semi-supervised weighted SPCA, similarity networks, dip tests, and KANs. Unlike the traditional unsupervised sparse PCA model, the NMDP model proposes an independent evaluator that converts the sparse PCA model from a traditional unsupervised to a semi-supervised model. This improvement allows the NMDP model to use known dataset grouping information, ultimately allowing the model to stably select different potential target genes for different target drugs. The experimental results show that the NMDP model inherits the advantages of the sparse PCA model, such as good biological interpretability and strong denoising ability, further enhances the feature selection ability of the model in multi-omics gene data, and greatly strengthens the stability of the model in high-dimensional small-sample cases. Sample similarity networks further address the

dimensionality challenge of the samples while helping the model perform multi-omics data alignment. We introduce a fusion algorithm that utilizes both dip test and variance data with weighted integration, which allows the model to focus on important histological information, thus improving prediction accuracy. Finally, we propose a one-dimensional convolution combined with KANs for drug response prediction modeling. The model achieves efficient prediction with a small number of parameters, thereby effectively avoiding the overfitting problem.

To enhance the validation of the model, we also conduct external validation experiments to assess the generalization capability of the model using independent datasets. The experimental findings indicate that the NMDP model performs consistently on different datasets, validating its robustness and reliability. In addition, we conduct ablation experiments to evaluate the contribution of each component to the model performance. The results of the ablation experiments show that removing the multi-omics weighting module and the convolution module significantly degrades the model's performance, and in particular, the sample similarity network module plays the most crucial role in influencing the model's effectiveness. This further emphasizes the importance of inter-sample similarity and the unique advantage of KANs in capturing complex relationships. Bioenrichment experiments fully validate the biointerpretability of the model, suggesting that the NMDP model could help researchers in drug development.

We also acknowledge some limitations to this study: our research is confined to predicting the response to a single drug, without considering the effects of combination drug therapies. Moreover, the weighted edge sparse PCA method has high time complexity, which leads to slower model computations. In future work, we plan to improve the model's ability to predict responses to drug combinations and optimize its computational efficiency.

Data availability statement

The data presented in the study are deposited in the European Genome-phenome Archive (EGA) repository, accession numbers EGAD00001001039, EGAD00001004201, EGAD00010000644; and in the Gene Expression Omnibus (GEO) repository, accession number GSE68379.

Author contributions

RM: data curation, methodology, software, writing–original draft, and writing–review and editing. B-JZ: writing–original draft and writing–review and editing. X-YM: methodology, software, and writing–original draft. XD: software and writing–original draft. Y-DO: software and writing–original draft. YL: writing–review and editing. H-YY: writing–review and editing. YW: writing–review and editing. Z-HD: writing–review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. The study is supported by the Guangdong Provincial Department of Education youth innovative talent project (no. 2023KQNCX155), Post-doctoral training project of Zunyi Medical University (no. 2023F-ZH- 019), the Key Construction Discipline of Immunology and Pathogen Biology Fund, Zunyi Medical University Zhuhai Campus, China (no. ZHGF 2024-1), and the Science and Technology Project of Guizhou Province (Project Number: Guizhou Science and Technology Support [2021] General 446).

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.

Generative AI statement

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2025.1532651/full#supplementary-material

References

Adam, G., Rampášek, L., Safikhani, Z., Smirnov, P., Haibe-Kains, B., and Goldenberg, A. (2020). Machine learning approaches to drug response prediction: challenges and recent progress. *NPJ Precis. Oncol.* 4 (1), 19. doi:10.1038/s41698-020-0122-1

Bach, D.-H., Kim, D., An, Y. J., Park, S., Park, H. J., Lee, S. K., et al. (2018). Bmp4 upregulation is associated with acquired drug resistance and fatty acid metabolism in egfr-mutant non-small-cell lung cancer cells. *Mol. Therapy-Nucleic Acids* 12, 817–828. doi:10.1016/j.omtn.2018.07.016

Ballester, P. J., and Carmona, J. (2021). Artificial intelligence for the next generation of precision oncology. *NPJ Precis. Oncol.* 5 (1), 79. doi:10.1038/s41698-021-00216-w

Ballester, P. J., Stevens, R., Haibe-Kains, B., Huang, R. S., and Aittokallio, T. (2022). *Artificial intelligence for drug response prediction in disease models, bbab450*. Oxford University Press.

Baptista, D., and Ferreira, P. G. (2023). A systematic evaluation of deep learning methods for the prediction of drug synergy in. *Cancer* 19 (3), e1010200. doi:10.1371/journal.pcbi.1010200

Baptista, D., Ferreira, P. G., and Rocha, M. (2021). Deep learning for drug response prediction in cancer. *Briefings Bioinforma.* 22 (1), 360–379. doi:10.1093/bib/bbz171

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., et al. (2012a). The Cancer Cell Line Encyclopedia enables predictive modelling of

anticancer drug sensitivity. *Cancer Cell Line Encycl. enables Predict. Model. anticancer drug Sensit. Nat.* 483, 603–607. doi:10.1038/nature11003

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., et al. (2012b). The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483 (7391), 603–607. doi:10.1038/nature11003

Betticher, D. C., Heighway, J., Hasleton, P. S., Altermatt, H. J., Ryder, W. D. J., Cerny, T., et al. (1996). Prognostic significance of Ccnd1 (cyclin D1) overexpression in primary resected non-small-cell lung cancer. *Br. J. cancer* 73 (3), 294–300. doi:10.1038/bjc.1996.52

Chen, C., Zhang, Z., Li, J., and Sun, Y. (2018). Snhg8 is identified as a key regulator in non-small-cell lung cancer progression sponging to mir-542-3p by targeting ccnd1/cdk6. *OncoTargets Ther.* 11, 6081–6090. doi:10.2147/OTT.S170482

Chen, J., and Zhang, L. (2021). A survey and systematic assessment of computational methods for drug response prediction. *Briefings Bioinforma.* 22 (1), 232–246. doi:10.1093/bib/bbz164

Chiu, Y.-C., Chen, H.-I. H., Gorthi, A., Mostavi, M., Zheng, S., Huang, Y., et al. (2020). Deep learning of pharmacogenomics resources: moving towards precision oncology. *Briefings Bioinforma.* 21 (6), 2066–2083. doi:10.1093/bib/bbz144

Chiu, Y.-C., Chen, H.-I. H., Zhang, T., Zhang, S., Gorthi, A., Wang, L.-Ju, et al. (2019). Predicting drug response of tumors from integrated genomic profiles by deep neural networks. *BMC Med. genomics* 12 (1), 18–55. doi:10.1186/s12920-018-0460-9

Deng, T., Chen, S., Zhang, Y., Xu, Y., Feng, Da, and Wu, H. (2023). A cofunctional grouping-based approach for non-redundant feature gene selection in unannotated single-cell rna-seq analysis. *Brief. Bioinform.* 24, bbad042. doi:10.1093/bib/bbad042

Ding, M. Q., Chen, L., Cooper, G. F., Young, J. D., and Lu, X. (2018). Precision oncology beyond targeted therapy: combining omics data with machine learning matches the majority of cancer cells to effective therapeutics. *Mol. cancer Res.* 16 (2), 269–278. doi:10.1158/1541-7786.MCR-17-0378

Dlamini, Z., Francies, F. Z., Hull, R., and Marima, R. (2020). Artificial intelligence (ai) and big data in cancer and precision oncology. *Comput. Struct. Biotechnol. J.* 18, 2300–2311. doi:10.1016/j.csbj.2020.08.019

Dong, Z., Zhang, N., Li, C., Wang, H., Fang, Y., Wang, J., et al. (2015). Anticancer drug sensitivity prediction in cell lines from baseline gene expression through recursive feature selection. *BMC cancer* 15, 489–512. doi:10.1186/s12885-015-1492-6

Firoozbakht, F., Yousefi, B., and Schwikowski, B. (2022). An overview of machine learning methods for monotherapy drug response prediction. *Briefings Bioinforma.* 23 (1), bbab408. doi:10.1093/bib/bbab408

Gao, H., Korn, J. M., Ferretti, S., Monahan, J. E., Wang, Y., Singh, M., et al. (2015). High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat. Med.* 21 (11), 1318–1325. doi:10.1038/nm.3954

Garnett, M. J., Edelman, E. J., Heidorn, S. J., Greenman, C. D., Dastur, A., Lau, K. W., et al. (2012). Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* 483 (7391), 570–575. doi:10.1038/nature11005

Garraway, L. A., Verweij, J., and Ballman, K. V. (2013). Precision oncology: an overview. *J. Clin. Oncol.* 31 (15), 1803–1805. doi:10.1200/JCO.2013.49.4799

Geeleher, P., Cox, N. J., and Huang, R. S. (2014). Clinical drug response can Be predicted using baseline gene expression levels and *in vitro* drug sensitivity in cell lines. *Genome Biol.* 15, R47–R12. doi:10.1186/gb-2014-15-3-r47

Hartigan, J. A., and Pamela, M. J. (1985). "The annals of statistics hartigan," in *The dip test of unimodality*, 70–84.

Hodson, R. (2020). Precision oncology. *Nature* 585 (7826), S1. doi:10.1038/d41586-020-02673-y

Ju, F.-J., Meng, F.-Q., Hu, H.-L., and Liu, J. (2019). Association between Bmp4 expression and pathology, ct characteristics and prognosis of non-small cell lung cancer. *Eur. Rev. Med. and Pharmacol. Sci.* 23, 5787–5794. doi:10.26355/eurrev_201907_18317

Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., et al. (2024). Kan: Kolmogorov-arnold networks. arXiv preprint arXiv:2404.19756.

Miao, R., Chen, H.-H., Qi, D., Xia, L.-Y., Yang, Z.-Yi, He, M.-F., et al. (2020). Beyond the limitation of targeted therapy: improve the application of targeted drugs combining genomic data with machine learning. *Pharmacol. Res.* 159, 104932. doi:10.1016/j.phrs.2020.104932

Miao, R., Dong, X., Liu, X.-Y., Lo, S.-L., Xin-Yue, M., Qi, D., et al. (2022). Dynamic meta-data network sparse pca for cancer subtype biomarker screening. *Front. Genet.* 13, 869906. doi:10.3389/fgene.2022.869906

Min, W., Liu, J., and Zhang, S. (2018). Edge-group sparse pca for network-guided high dimensional data analysis. *Bioinformatics* 34 (20), 3479–3487. doi:10.1093/bioinformatics/bty362

Munquad, S., Bikas, A., and Das, J. H. (2024). Uncovering the subtype-specific disease module and the development of drug response prediction models for glioma. *Heliyon* 10 (5), e27190. doi:10.1016/j.heliyon.2024.e27190

Murai, F., Koinuma, D., Shinozaki-Ushiku, A., Fukayama, M., Miyaozono, K., and Ehata, S. (2015). Ezh2 promotes progression of small cell lung cancer by suppressing the tgf-B-smad-ascl1 pathway. *Cell Discov.* 1 (1), 1–17. doi:10.1038/celldisc.2015.26

Park, S., Kim, T. M., Cho, S.-Y., Kim, S., Oh, Y., Kim, M., et al. (2020). Combined blockade of polo-like kinase and pan-raf is effective against nras-mutant non-small cell lung cancer cells. *Cancer Lett.* 495, 135–144. doi:10.1016/j.canlet.2020.09.018

Pązik, M., Michalska, K., Żebrowska-Nawrocka, M., Zawadzka, I., Łochowski, M., and Balcerczak, E. (2021). Clinical significance of hras and kras genes expression in patients with non–small-cell lung cancer-preliminary findings. *BMC cancer* 21, 130–213. doi:10.1186/s12885-021-07858-w

Peng, W., Chen, T., Liu, H., Dai, W., Yu, N., and Lan, W. (2023). Improving drug response prediction based on two-space graph convolution. *Comput. Biol. Med.* 158, 106859. doi:10.1016/j.compbiomed.2023.106859

Prasad, V. (2016). Perspective: the precision-oncology illusion. *Nature* 537 (7619), S63. doi:10.1038/537S63a

Prasad, V., Fojo, T., and Brada, M. (2016). Precision oncology: origins, optimism, and potential. *Lancet Oncol.* 17 (2), e81–e86. doi:10.1016/S1470-2045(15)00620-8

Rashid, Md M. (2024). Advancing drug-response prediction using multi-modal and-omics machine learning integration (momlin): a case study on breast cancer clinical data. *Brief. Bioinform.* 25 (4), bbae300. doi:10.1093/bib/bbae300

Rubio-Perez, C., Tamborero, D., Schroeder, M. P., Antolín, A. A., Deu-Pons, J., Perez-Llamas, C., et al. (2015). *In silico* prescription of anticancer drugs to cohorts of 28 tumor types reveals targeting opportunities. *Cancer Cell* 27 (3), 382–396. doi:10.1016/j.ccell.2015.02.007

Sharifi-Noghabi, H., Zolotareva, O., Collins, C. C., and Ester, M. (2019). Moli: multi-omics late integration with deep neural networks for drug response prediction. *Bioinformatics* 35 (14), i501–i509. doi:10.1093/bioinformatics/btz318

Sharma, A., Lysenko, A., and Boroevich, K. A. (2023). DeepInsight-3D architecture for anti-cancer drug response prediction with deep-learning on multi-omics. *Deepinsight-3d Archit. Anti-Cancer Drug Response Predict. Deep-Learning Multi-Omics* 13 (1), 2483. doi:10.1038/s41598-023-29644-3

Sharma, A., Lysenko, A., Jia, S., and Boroevich, K. A. (2024). *Advances in ai and machine learning for predictive medicine*, 1–11.

Sheng, J., Li, F., and Wong, S. T. C. (2015). Optimal drug prediction from personal genomics profiles. *IEEE J. Biomed. Health Inf.* 19 (4), 1264–1270. doi:10.1109/JBHI.2015.2412522

Trac, Q. T., Pawitan, Y., Mou, T., Erkers, T., Östling, P., Bohlin, A., et al. (2023). Prediction model for drug response of acute myeloid leukemia patients. *NPJ Precis. Oncol.* 7 (1), 32. doi:10.1038/s41698-023-00374-z

Tsao, M.-S., Sakurada, A., Cutz, J.-C., Zhu, C.-Qi, Kamel-Reid, S., Squire, J., et al. (2005). Erlotinib in lung cancer—molecular and clinical predictors of outcome. *N. Engl. J. Med.* 353 (2), 133–144. doi:10.1056/NEJMoa050736

Wang, C., Zhang, M., Zhao, J., Li, B., Xiao, X., and Zhang, Y. (2023). The prediction of drug sensitivity by multi-omics fusion reveals the heterogeneity of drug response in pan-cancer. *Comput. Biol. Med.* 163, 107220. doi:10.1016/j.compbiomed.2023.107220

Xu, K., Lu, Y., Hou, S., Liu, K., Du, Y., Huang, M., et al. (2024). Detecting anomalous anatomic regions in spatial transcriptomics with stands. *Nat. Commun.* 15 (1), 8223. doi:10.1038/s41467-024-52445-9

Xue, Y., Meehan, B., Fu, Z., Wang, X. Q. D., Fiset, P. O., Rieker, R., et al. (2019). Smarca4 loss is synthetic lethal with cdk4/6 inhibition in non-small cell lung cancer. *Nat. Commun.* 10 (1), 557. doi:10.1038/s41467-019-08380-1

Zhao, M., Xu, P., Liu, Z., Zhen, Y., Chen, Y., Liu, Y., et al. (2018). Retracted article: dual roles of mir-374a by modulated C-jun respectively targets ccnd1-inducing pi3k/akt signal and pten-suppressing wnt/B-catenin signaling in non-small-cell lung cancer. *Cell Death and Dis.* 9 (2), 78. doi:10.1038/s41419-017-0103-7

Zheng, X., Wang, M., and Huang, K. (2024). Global and cross-modal feature aggregation for multi-omics data classification and application on drug response prediction. *Inf. Fusion* 102, 102077. doi:10.1016/j.inffus.2023.102077

Zhou, L., Wang, N., Zhu, Z., Gao, H., Lu, N., Su, H., et al. (2024). Multi-omics fusion based on attention mechanism for survival and drug response prediction in digestive system tumors, *Neurocomputing* 572, 127168. doi:10.1016/j.neucom.2023.127168

Zhou, Q., Xu, C.-R., Cheng, Y., Liu, Y.-P., Chen, G.-Y., Cui, J.-W., et al. (2021). Bevacizumab plus erlotinib in Chinese patients with untreated, egfr-mutated, advanced nsclc (Artemis-Ctong1509): a multicenter phase 3 study. *Cancer Cell* 39 (9), 1279–1291.e3. doi:10.1016/j.ccell.2021.07.005

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* 10 (1), 1523–1610. doi:10.1038/s41467-019-09234-6

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 03 April 2025
DOI 10.3389/fmed.2025.1539467

Check for updates

OPEN ACCESS

EDITED BY
HaiHui Huang,
Shaoguan University, China

REVIEWED BY
Xingli Fan,
Hangzhou Medical College, China
Linzi Liu,
China Medical University, China

*CORRESPONDENCE
Haiyan Wang
✉ wanghaiyan@qhu.edu.cn

RECEIVED 04 December 2024
ACCEPTED 10 March 2025
PUBLISHED 03 April 2025

CITATION
Wang L, Wang T, Lei Y, Su Y, Lin Y, Wu Z, Wu Q, Zhang S and Wang H (2025) Unraveling the mechanisms of propofol-induced psychological dependence: a multi-omics approach linked to gut microbiota in hippocampal function.
Front. Med. 12:1539467.
doi: 10.3389/fmed.2025.1539467

COPYRIGHT
© 2025 Wang, Wang, Lei, Su, Lin, Wu, Wu, Zhang 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.

Unraveling the mechanisms of propofol-induced psychological dependence: a multi-omics approach linked to gut microbiota in hippocampal function

Li Wang[1], Tangyi Wang[2], Yadian Lei[1,2], Yudong Su[1], Yuxin Lin[1], Zhijing Wu[1], Qiong Wu[1,3,4], Shoude Zhang[2] and Haiyan Wang[1,3,4]*

[1]Department of Basic Medical Sciences, Qinghai University Medical College, Xining, Qinghai, China, [2]State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University, Xining, Qinghai, China, [3]Research Center for High Altitude Medicine, Qinghai University, Xining, Qinghai, China, [4]Key Laboratory of the Ministry of High Altitude Medicine, Qinghai University, Xining, Qinghai, China

Introduction: Drug abuse is becoming a global public health crisis. According to the United Nations, the number of drug users worldwide has increased dramatically over the past decade, with a surge in the number of drug abusers. The problem was exacerbated by the expanding market for illicit drugs and the increasing availability of synthetic drugs such as fentanyl. Clinical drug abuse is a problem that requires particular attention, and the potential addictive properties of some drugs and their mechanisms of action are currently unknown, which limits the development and implementation of drug addiction intervention strategies.

Methods: Eight-week-old C57BL/6J mice were used as study subjects. A mental dependence model was established using the conditional position preference experiment (CPP), and the hippocampal tissues of the model mice were subjected to RNA-seq transcriptome sequencing, LC–MS non-targeted metabolome sequencing, and intestinal macro-genome sequencing in order to discover propofol mental dependence signature genes. Correlation analyses of transcriptomics and metabolomics were performed using the Spearman method, and gene-metabolite networks were mapped using Cytoscape software. Real-time fluorescence quantitative PCR and immunoprotein blotting (Western blotting) methods were used to validate the characterized genes.

Results: After the conditioned position preference experiment, the conditioned preference scores of the 75 mg/kg propofol and 2 g/kg alcohol groups were significantly higher than those of the control saline group. 152 differential genes and 214 differential metabolites were identified in the 75 mg/kg group. Cluster analysis revealed that changes in the neuroactive ligand receptor pathway were most pronounced. Gut microbiomics assays revealed significant changes in five differential enterobacterial phyla (*Campylobacter phylum*, *Thick-walled phylum*, *Anaplasma phylum*, *Actinobacteria phylum*, and *Chlorella verticillata phylum*) in the 75 mg/kg propofol group, which may be related to changes in the differential expression of dopamine.

Discussion: These findings suggest that 75 mg/kg propofol has a significant mind-dependent effect on the biology of drug addiction through neuroactive ligand-receptor interaction pathways in conjunction with the tricarboxylic acid cycle, and the metabolic pathways of alanine, aspartate, and glutamate that may influence intestinal microbial changes through bidirectional signaling.

KEYWORDS

propofol, psychiatric dependence, gut microbes, transcriptomics, metabolomics, hippocampus

1 Introduction

Propofol, as a fast-acting short-acting anesthetic, is widely used in a variety of clinical practices, including short-term anesthesia for abortion, gastroscopy, and the induction and maintenance of general anesthesia (1), but also for conscious sedation in critically ill patients, as well as in the treatment of refractory agitated delirium and antiemetic. Recent studies have also shown that propofol has therapeutic and anti-inflammatory effects (2). Its main mechanism of action is to inhibit neural signaling by promoting chloride inward flow via GABA-A type receptors (3). However, the risk of propofol abuse and addiction cannot be ignored, especially among medical professionals (4, 5), long-term or overdose use of propofol may lead to dependence, increased tolerance, and withdrawal symptoms, and overdose may even be life-threatening (6).

Drug addiction is considered a chronic and relapsing brain disorder characterized by persistent craving for and use of drugs regardless of negative consequences (7). Underlying this craving and use behavior are long-term gene expression changes, neuronal adaptations, and changes in synaptic plasticity triggered by repeated drug ingestion. The hippocampus, a key brain region for learning and memory, plays a central role in drug addiction (8). Drug addiction can lead to significant changes in neuroplasticity in the hippocampus that include changes in neuronal excitability, neurotransmission, morphological changes in dendrites and axons, and synapse formation or elimination (9). Neurotransmitter systems in the hippocampus, including dopamine, glutamate, and GABA, are closely linked to the neural mechanisms of drug addiction (10). Current research on the mechanisms of drug addiction has focused on changes in brain neurotransmitters (dopamine, glutamate, and GABA, among others) (11). The hippocampal region, a key brain area for learning, memory, and spatial navigation, plays a central role in the development and maintenance of drug addiction (8). Addictive drugs such as cocaine, opioids, and nicotine alter the structure and gene expression of the region by modulating synaptic plasticity in the hippocampus. In addition, drug abuse-induced changes in neurotransmitter systems, including serotonin and endorphins, have been strongly associated with the development of addictive behaviors by modulating mood, memory, and reward behaviors. Most addictive drugs are directly linked to reward effects by increasing the release of dopamine in the brain, particularly in the hippocampus, leading to craving and dependence on the drug (12). In neurobiological models of addiction, changes in dopamine receptor expression in the hippocampus have received widespread attention, e.g., chronic drug exposure (e.g., cocaine, endogenous cannabinoid analogs) may lead to down-regulation of dopamine D2 receptors in the hippocampus, which may be related to reduced sensitivity to drugs and disinterest in non-pharmacological rewards in addicted individuals (13). Dysregulated dopamine signaling in the hippocampus has also been associated with the risk of relapse in drug addiction, as these alterations may affect an individual's response to drug-related cues and decision-making processes. Thus, the hippocampus and dopamine system play a critical role in the development, maintenance, and relapse of drug addiction, and long-term alterations in these brain regions provide a neurobiological basis for the persistence of addiction and the complexity of treatment.

Multi-omics technology refers to the integrated application of various genomics techniques such as genomics, transcriptomics, proteomics, metabolomics, etc., to comprehensively analyze the changes in biological samples at different biological levels, and in the case of drug addiction, many researchers have used metabolomics to find that metabolites such as (inositol-1-phosphate, free fatty acids, and metabolites related to tricarboxylic acid cycle, etc.) are increased in the brain of rats after heroin addiction (14). Microbiome and metabolomics approaches to study methamphetamine users identify microbial metabolic pathways involved in addiction (15). The effects of chronic methamphetamine exposure on the neural proteome in the hippocampus and olfactory bulb region of rats were also investigated by proteomic approaches, revealing significant changes in the expression of 18 proteins related to addiction such as (synaptic vesicle glycoprotein 2A, myelin proteolipoproteins, etc.) (16). These technologies enable us to probe deeply into the biological basis of drug addiction at the molecular level, revealing the underlying gene expression regulation, protein function changes, metabolic pathway remodeling, and complex networks of inter-cellular interactions. For example, the development of single-cell sequencing technology and spatial transcriptomics provides a powerful tool to study the cellular heterogeneity and tissue microenvironment of drug addiction (17, 18). Dysregulation of gut flora in alcohol addiction and modification of addiction using gut flora modification (19). The application of these

Abbreviations: CPP, Conditional position preference; TCA, Tricarboxylic acid cycle; LC–MS, Liquid chromatography–tandem mass spectrometry; QPCR, Quantitative polymerase chain reaction; OPLS-DA, Orthogonal partial Least Squares-Discriminant Analysis; PCA, Principal component analysis; NMDS, Non-metric multidimensional scaling; LEfSe, Linear Discriminant Analysis Effect Size; RIPA, Radio Immunoprecipitation Assay; SYBR, Synergetic Binding Reagent; ANOVA, Analysis of variance; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; SCFA, Short-chain fatty acids.

technologies not only greatly broadens our understanding of the pathogenesis of drug addiction, but also provides a solid scientific basis for the development of new treatment strategies. At present, the combined application of propofol and multi-omics technology is still in its infancy, and the related research results are limited. Transcriptome studies have found that propofol can change the structure and function of the developing heart, suggesting its potential cardiotoxic effect (20). Metabolomics studies have shown that propofol can significantly increase the level of inflammatory marker glycoprotein acetylation (GlycA) (21).

In recent years, the interaction between gut microbiota and host metabolism has become a hot topic in biomedical research, such as the brain-gut axis and the gut-liver axis. The gut microbiome not only plays a key role in maintaining the host's nutritional metabolism, immune regulation and intestinal barrier function, but also affects the host's systemic metabolic status through its metabolites (22). More and more evidence suggests that the interaction between gut microbiome and host metabolism may play an important role in the pathogenesis of drug addiction. On the one hand, gut microbes can affect the behavior and mood of the host by regulating neurotransmitter levels and immune responses. On the other hand, metabolic changes may reflect the systemic effects of drug abuse on the body.

The purpose of this study is to explore the mechanism of propophenol-induced mental dependence in mice by network pharmacology, transcriptomtics, metabolomics and metagenomics, reveal the molecular mechanism of propofol addiction, explore the changes of hippocampal and intestinal flora in mice after propofol addiction, and provide therapeutic strategies for clinical treatment of propofol addiction and prevention of propofol abuse. The flow chart of the experiment is shown in Figure 1A.

2 Materials and methods

2.1 Chemical reagents

Propofol was purchased from McLean Biotechnology (D806979), HiScript II Q RT SuperMix for qPCR (R223-01, vazyme, Nanjing, China), primers were purchased from Xianghong Bio-technology Co., Ltd., and antibodies to DRD1 and DRD2 were purchased from Proteintech (17934-1-AP, Wuhan, China).

2.2 Animal models and experimental design

Healthy male C57BL/6J mice (6 weeks old, weighing 20–23 g) were purchased from Xi'an Fraser Biotechnology Co. The mice were acclimatized to the laboratory environment for 2 weeks before the experiment. Standard diet and water were ad libitum. 72 mice were randomly divided into 6 groups of 12 mice each, including saline control group, propofol-treated group (0 mg/kg, 50 mg/kg, 75 mg/kg, 100 mg/kg) and alcohol control group. Propofol was injected intraperitoneally and 100 mg/kg was chosen as the highest dose based on the literature that the 114 mg/kg dose resulted in the loss of the flip reflex in mice. The normal saline group and alcohol group were used as negative and positive controls, respectively. The experiment was divided into adaptation stage, training stage and testing stage. The adaptation period lasted for 3 days. Every day, the mice were placed in the middle channel, opened the channel door, and explored freely for 30 min to help the mice adapt to the experimental environment. On the fourth day, the basic time test was performed with a 15 min limit time, and the mouse's residence time in the preference box and the non-preference box was recorded as the baseline data of behavioral preferences. Then enter the 20 day training phase. In the morning, mice were intraperitoneally injected with normal saline, propofol and alcohol of different concentrations, and placed in the preference box to close the door of the middle channel; In the afternoon, all mice received the same volume of normal saline injection as the control group, placed in the non preference box, closed the channel door, and then cycled to the 21st day, entered the test phase, reopened the door of the middle channel, the mice explored freely for 15 min, and recorded the residence time in the preference box and the non preference box. The preference score was calculated by the residence time before and after the experiment to evaluate the effect of propofol and alcohol on the behavioral preference of mice (23, 24). At the end of the test, mice were executed and hippocampal tissues were separated on ice and stored in liquid nitrogen for subsequent experiments. All experiments involving propofol drug addiction in this study were approved by the Ethics Committee of Qinghai University School of Medicine (2022–01). Animal experiments were conducted in accordance with the European Guidelines for the Care and Use of Laboratory Animals (2010/63/EU).

2.3 Metabolomics

A 50–100 mg sample was taken and added to a methanol-acetonitrile mixture for low temperature sonication extraction. Centrifuge at 12000 rpm for 10 min, take the supernatant and add 200 μL of 30% acetonitrile solution to re-dissolve and centrifuge at 14000 rpm for 15 min. Samples were analyzed using Vanquish UPLC (Thermo, USA). Samples were separated on a Waters HSS T3 column with electrospray ionization source detection. Raw data were pre-processed using Progenesis QI software (Waters Corporation, Milford, USA), normalized and imported into R software. The metabolic abundance of each group of samples was standardized to eliminate the technical variation between different samples. On this basis, the average abundance of each group of metabolites was calculated, and the FC value was further calculated. The p value was calculated by Student's t-test, and the VIp value was calculated by multivariate statistical analysis method OPLS-DA. When screening differential metabolites, strict thresholds were set: P value 1, and FC > 1.5 or < 0.667. The metabolites that met these conditions were identified as differential metabolites, and the significance level of metabolite enrichment in each pathway was analyzed by Fisher's exact test.

2.4 Transcriptomics

Total RNA was extracted by Trizol reagent and evaluated for quantity and purity. High-quality RNA samples were selected for construction of sequencing libraries. The mRNA was enriched using the magnetic bead method, followed by fragmentation, reverse transcription and PCR amplification. DESeq2 software was used to identify

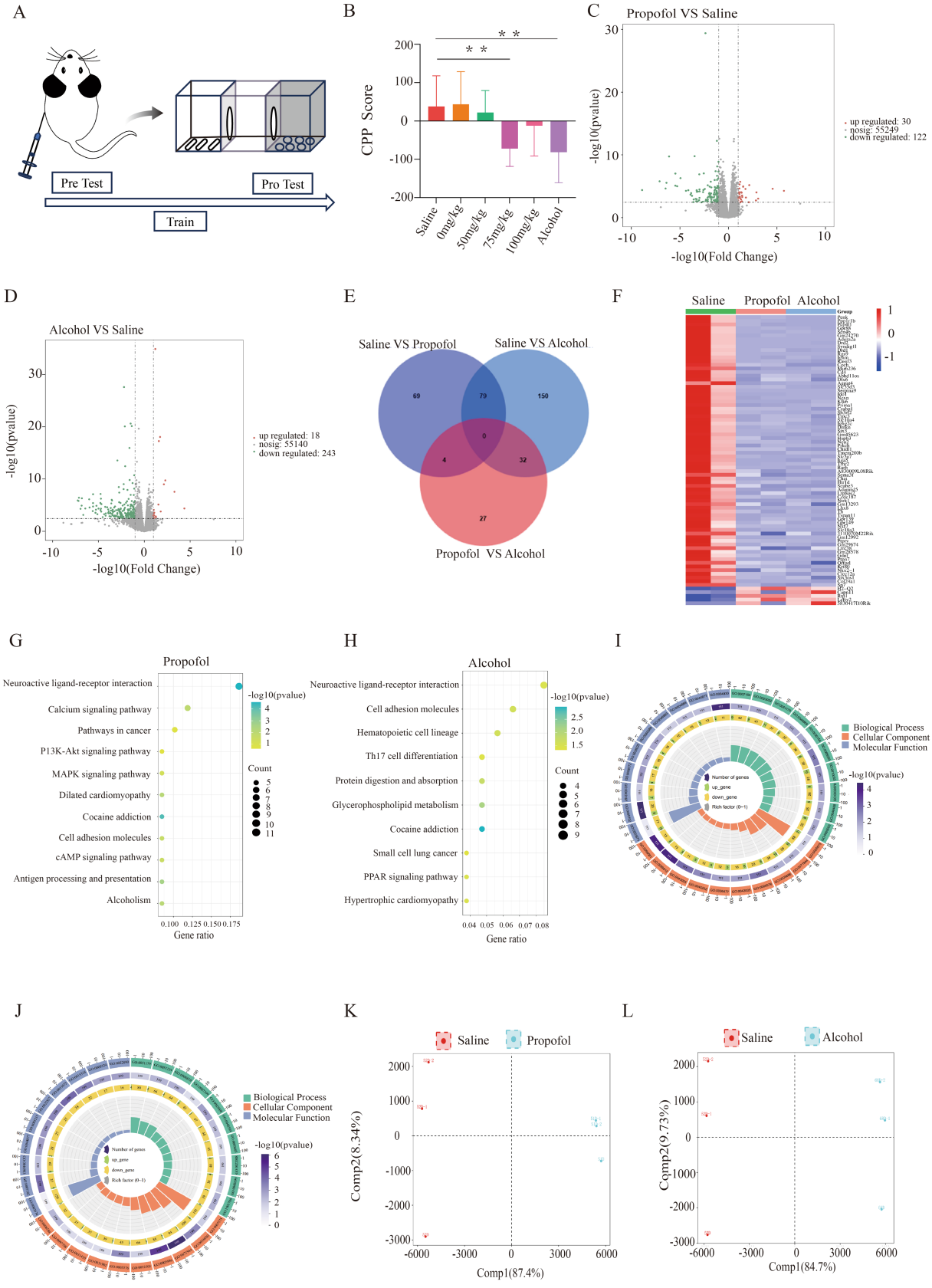


FIGURE 1
Hippocampal transcriptome changes in 75 mg/kg propofol-addicted mice. **(A)** Study design. **(B)** Change in preference scores for each group after the conditional location preference experiment (preference preference score = time before the experiment - time after the experiment) $p < 0.05$, $^{*}p < 0.01$. **(C)** Volcano map of the propofol group. **(D)** Volcanic map of the alcohol group. **(E)** Differential gene intersections between the propofol group and the

(Continued)

FIGURE 1 (Continued)
saline and alcohol groups. **(F)** Heat map of gene expression in saline, propofol group and alcohol group. **(G)** KEGG-enriched bubble map of the propofol group. **(H)** KEGG-enriched bubble plot for the alcohol group. **(I)** GO analysis enrichment circle plot for the propofol group. **(J)** GO analysis enrichment circle plot for the alcohol group. **(K)** Metabolite principal component analysis plot for the propofol group. **(L)** Metabolite principal component analysis plot for the alcohol group.

differentially expressed transcripts and genes, setting $|\log_2(\text{fold change})| > 1$ as the threshold, and KEGG and GO enrichment analyses were performed. p value corrected $p < 0.05$ was used as the screening criterion to further analyze the GO function and KEGG pathway enrichment.

2.5 Macrogenomics

Samples of intestinal contents were collected and total DNA was extracted from the samples and tested for DNA purity and integrity. DNA samples were broken and libraries were constructed for high-throughput sequencing. Raw data were pre-processed in Illumina fastq format to remove host contamination. The obtained sequences were spliced and assembled, and gene prediction, annotation and classification were performed. Finally, the samples were subjected to similarity clustering, sequencing tests and statistical comparison of differences. For the analysis of intestinal microbiota, principal component analysis (PCA) was used to preliminarily distinguish the significant differences between different groups. Subsequently, the differences in the composition of intestinal microbial communities under different groups were shown by species composition histograms, and these differences were further quantified by linear discriminant analysis (LDA). In order to identify microbial species with significant differences, STAMP analysis tool and Wilcoxon rank sum test were used to analyze the differences between groups, and the significance level was also set as $p < 0.05$. This method can not only identify different species, but also visually display their distribution in different groups.

2.6 Network pharmacology and molecular docking

Potential targets of propofol were collected from four databases, PharmMapper, SwissTargetPrediction, Drugbank and BATMAN-TCM, and from NCBI, GeneCard, Therapeutic Targets Database, Pharmacogenomics Knowledge Base four databases to collect relevant targets for drug addiction (25–28). The Wayne diagrams of propofol targets and drug addiction targets were drawn using the Microbiotics Online Platform.[1] Protein interaction networks were constructed using the STRING database. The DAVID database was used to perform GO and KEGG pathway enrichment analyses and visualization of potential targets. The 3D structural data of propofol were downloaded from the PubChem database, while the protein structural data were obtained from the PDB database. Molecular docking was performed through the CB-Dock2 platform.[2]

2.7 Comprehensive analysis of transcriptomics and metabolomics

On the basis of transcriptomics and metabolomics sequencing analysis, the correlation analysis of OPLS-DA was performed on the differential genes and differential metabolites of propofol at a concentration of 75 mg / kg, and the load map was drawn. Subsequently, the differential metabolites and differential genes were analyzed by Pearson correlation analysis for clustering heat map drawing, and the correlation was set to $p < 0.05$ ($p < 0.05$, '*') to explore the correlation between genes and metabolites. After that, Cytoscape (Cytoscape v3.9.0) was used to draw a gene-enzyme reaction-metabolite network diagram to explore the relationship between genes and metabolites.

2.8 Real-time fluorescence quantitative PCR

RNA was extracted from hippocampus, converted to cDNA by reverse transcription kit, and amplified by PCR using SYBR method. Through the PCR instrument, the fluorescence signal changes were monitored and collected to obtain the Ct value (cycling threshold), and the relative quantitative analysis was performed by the $2^{\wedge-\Delta\Delta Ct}$ method.

2.9 Western blotting

Hippocampal tissue was mixed with RIPA lysate, protease and phosphatase inhibitor (100:1:1) and ground, and the supernatant was subjected to polypropylene gel electrophoresis, incubated overnight at 4°C with primary antibody against DRD1 and DRD2, and then the secondary antibody was incubated and developed, and the images were collected.

2.10 Statistical analysis

Behavioral data were calculated by the preference score formula, plotted and analyzed using GraphPad Prism software v10.1.2. Comparisons between multiple groups were analyzed using one-way ANOVA with $p < 0.05$ as the criterion for significance. All experiments were repeated three times with $p < 0.05$ as the criterion for statistical significance.

3 Results

3.1 Behavioral analysis

The results of CPP showed that the preference scores of mice in the 75 mg/kg Propofol group (Propofol group) changed

1 https://www.bioinformatics.com.cn/
2 https://cadd.labshare.cn/cb-dock2/php/index.php

significantly and at $p < 0.05$ compared to the saline control group, the preference scores of mice in the Alcohol group changed significantly and at $p < 0.05$ compared to the saline group, and the preference scores of mice in the Propofol group showed the same trend of change compared to the Alcohol group which indicated that isoPropofol did form a mental dependence, see (Figure 1B).

3.2 Transcriptomic analysis of propofol group and alcohol group

In the transcriptomic analyses performed in the propofol and alcohol groups, we used $|\log_2 FC| > 1$ and $p < 0.05$ as the criteria for screening differentially expressed genes. The results showed that there were 152 genes with significant changes in expression in the propofol group, of which 30 were up-regulated and 122 were down-regulated. The alcohol group, on the other hand, had 261 genes with significant changes in expression, including 18 up-regulated and 243 down-regulated genes (Figures 1C,D). Cross-tabulation analysis of gene expression revealed 79 common differentially expressed genes. Clustering heatmap analysis further revealed that the propofol and alcohol groups were similar in differential gene expression patterns (Figures 1E,F). KEGG pathway enrichment analysis pointed out that samples from both groups exhibited significant changes in cocaine addiction and neuroactive ligand-receptor interaction pathways (Figures 1G,H). In addition, GO functional enrichment analysis identified 10 significantly enriched biological processes covering signal transduction, dopamine receptor activity, and lipid and organic acid binding functions (Figures 1I,J).

3.3 Metabolomic changes in propofol and alcohol groups

Alterations in hippocampal metabolic profiles by propofol addiction were assessed from metabolite expression levels, and OPLS-DA analysis of the propofol and alcohol groups showed that samples within the propofol and alcohol groups clustered together, whereas the samples between the groups tended to be significantly separated, suggesting that there were significant differences between the groups (Figures 1K,L). Volcano plots of the differential metabolites screened in the propofol group and alcohol group are shown (Figures 2A,B). The cross-differential metabolite heatmaps of the saline, propofol, and alcohol groups clearly showed the differences in metabolites and 17 cross-differential metabolites between the propofol and alcohol groups and the control saline group (Figure 2C), which indicated that the metabolite trends were similar in the propofol and alcohol groups. Next, metabolite-related metabolic pathways were analyzed using the KEGG pathway library and plotted as bar graphs (Figures 2D,E), which showed that the metabolites in the propofol group and the alcohol group were mainly concentrated in the citric acid cycle, 2-oxocarboxylic acid metabolism, and alanine, aspartate, and glutamate metabolic pathways. In addition to the common metabolic compounds such as glycerophospholipids, carboxylic acids and their derivatives, which were found in the propofol and alcohol groups, there were also antioxidant and neuroprotective compounds such as benzothiazoles, coumarins and their derivatives, organic oxides, and purine nucleosides, as shown in the graphs (Figures 2F,G).

3.4 Differences in gut microbes between propofol and saline groups

To assess the effect of propofol addiction on gut microbial diversity, we analyzed species evenness and richness in the propofol group versus the saline group using Simpson's index and Shannon's index. The results showed that species evenness and richness were significantly higher in the propofol group than in the saline group (Figures 2H,I). Principal component analysis (PCA) and non-metric multidimensional scaling (NMDS) further revealed a high degree of clustering of the samples in the propofol group, indicative of a high diversity of community composition (Figure 3A). Linear discriminant analysis of LEfSe software identified potential biometabolic pathways (Figure 3B). Histogram analysis of species abundance revealed significant changes in microbial composition at the phylum level in the propofol group, including the disappearance of the Campylobacter phylum, a decrease in the thick-walled phylum, and an increase in the Anaplasma phylum, Actinobacteria phylum, Pseudomonas phylum, and Micrococcus wartyi phylum. At the genus level, *H. pylori* disappeared, Streptococcus decreased, and Lactobacillus spp. and Akkobacter spp. increased in abundance, changes indicative of key biomarker flora in the propofol group (Figures 3C,D). Comparative analysis of the macrogenomic data using Stamp software gave us information on the species composition abundance, functional prediction and their differences in the propofol group. KEGG and eggNOG functional prediction analyses revealed a significant increase in substance-dependent, neural and drug-resistance-associated metabolic pathways in the propofol group (Figures 3E,F).

3.5 Integrated analysis of metabolomics and transcriptomics data

To gain a deeper understanding of the biological changes in the hippocampal region of the propofol group, we constructed a correlation matrix heat map of 79 common differential genes and 17 common differential metabolites among the saline, propofol and alcohol groups by Spearman correlation analysis (Figure 3G), which revealed the expression patterns of key genes and metabolites. The analysis revealed significant correlations ($p < 0.05$) between 1,062 pairs of differential genes and metabolites, e.g., the dopamine receptor was positively correlated with cis-aconitine and (R)-ergosterol-5-pyrophosphate, and negatively correlated with docosahexaenoic acid (DHA), a metabolite of lysophosphatidylethanolamine, and lysophosphatidylcholine. In addition, OPLS-DA analysis of Top25 differential genes and metabolites further revealed the correlation between them (Figure 3H). The compound-reaction-enzyme-gene network diagram constructed using Cytoscape 3.9.0 software (Figure 4A) revealed the potential interactions and pathway regulation between metabolites and genes. These findings not only elucidated the biological changes in the propofol group, but

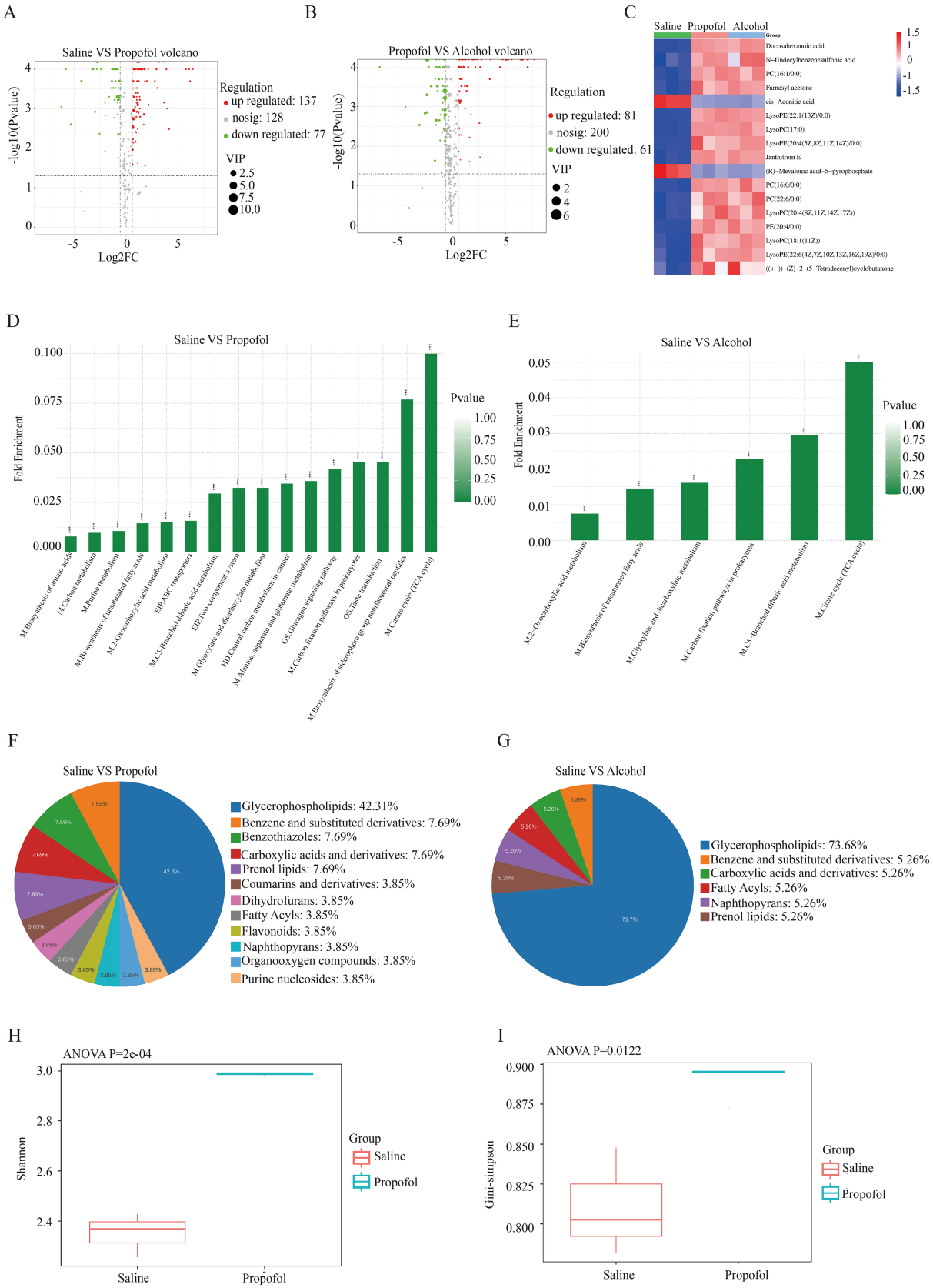


FIGURE 2
Changes in metabolomics in the hippocampus in the 75 mg/kg group. **(A)** Metabolite volcano plot of the propofol group. **(B)** Metabolite volcano plot of the propofol group versus the alcohol group. **(C)** Gene expression heatmap of the saline, propofol and alcohol groups. **(D)** KEGG pathway enrichment histogram for the propofol group. **(E)** KEGG pathway enrichment histogram for the alcohol group. **(F)** Pie chart of chemical classification of differential metabolites in the propofol group. **(G)** Pie chart of chemical classification of differential metabolites in the alcohol group. **(H)** Shannon index box plot of alpha diversity of gut microbes in the propofol group. **(I)** Box plot of the Gini-simpson index of alpha diversity of gut microbes in the propofol group.

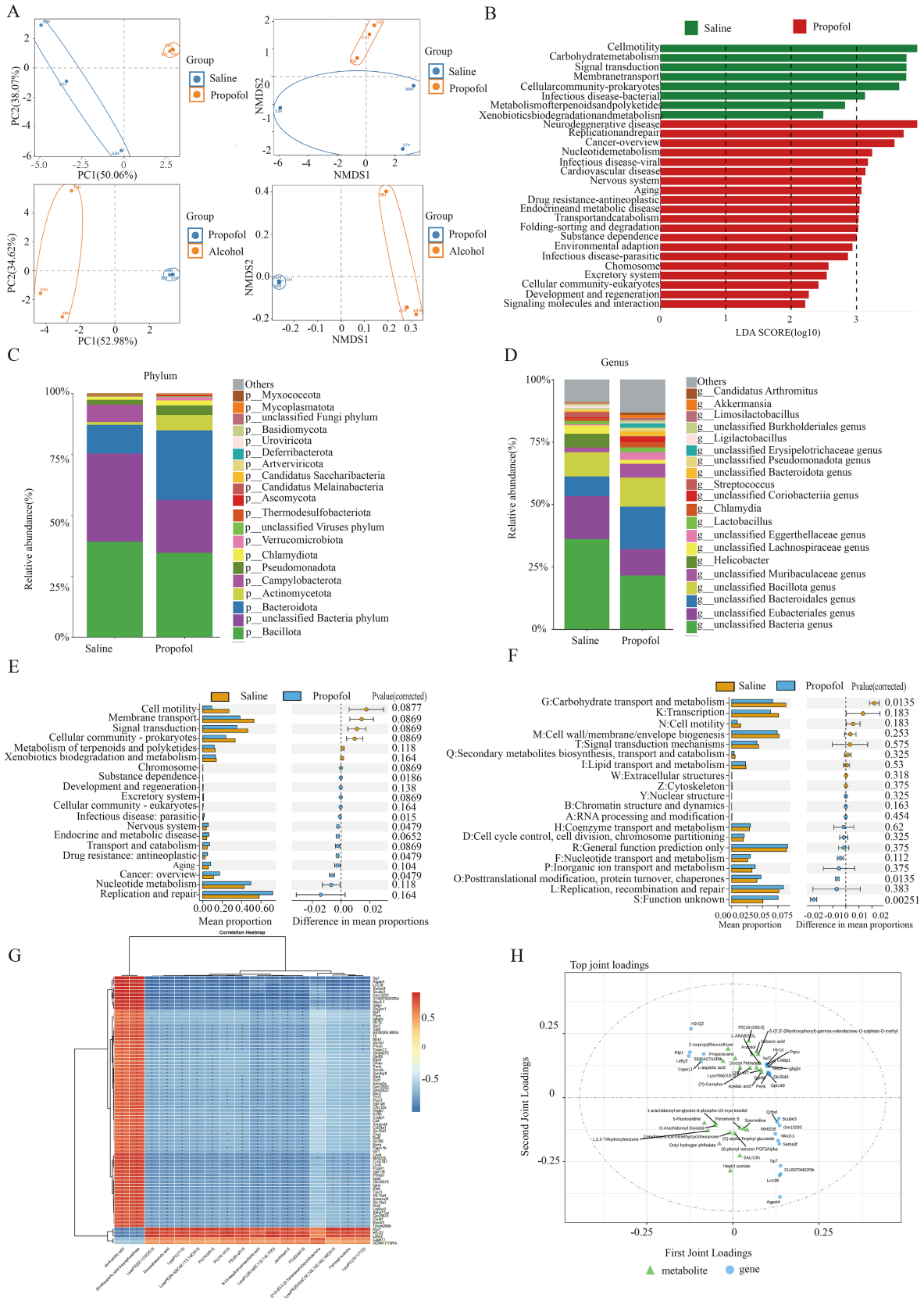


FIGURE 3
Characteristic gut microbiological changes in the propofol group and the control group. **(A)** PCA and NMDS analysis of gut microbes in the propofol group versus the saline and alcohol groups. **(B)** Linear discriminant analysis of gut microbial metabolic pathways in the propofol group. **(C)** Changes in species composition at the phylum level in the propofol group. **(D)** Species composition changes at the genus level in the propofol group. **(E)** KEGG pathway maps for significant differences in microbial abundance in the propofol group. **(F)** Functional annotation maps of eggNOG for significant differences in microbial abundance in the propofol group. **(G)** Correlation matrix heatmaps of differential genes with differential metabolites in the propofol group. *$p < 0.05$, **$p < 0.01$. **(H)** OPLS-DA analyses of differential genes with differential metabolites in the Top20.

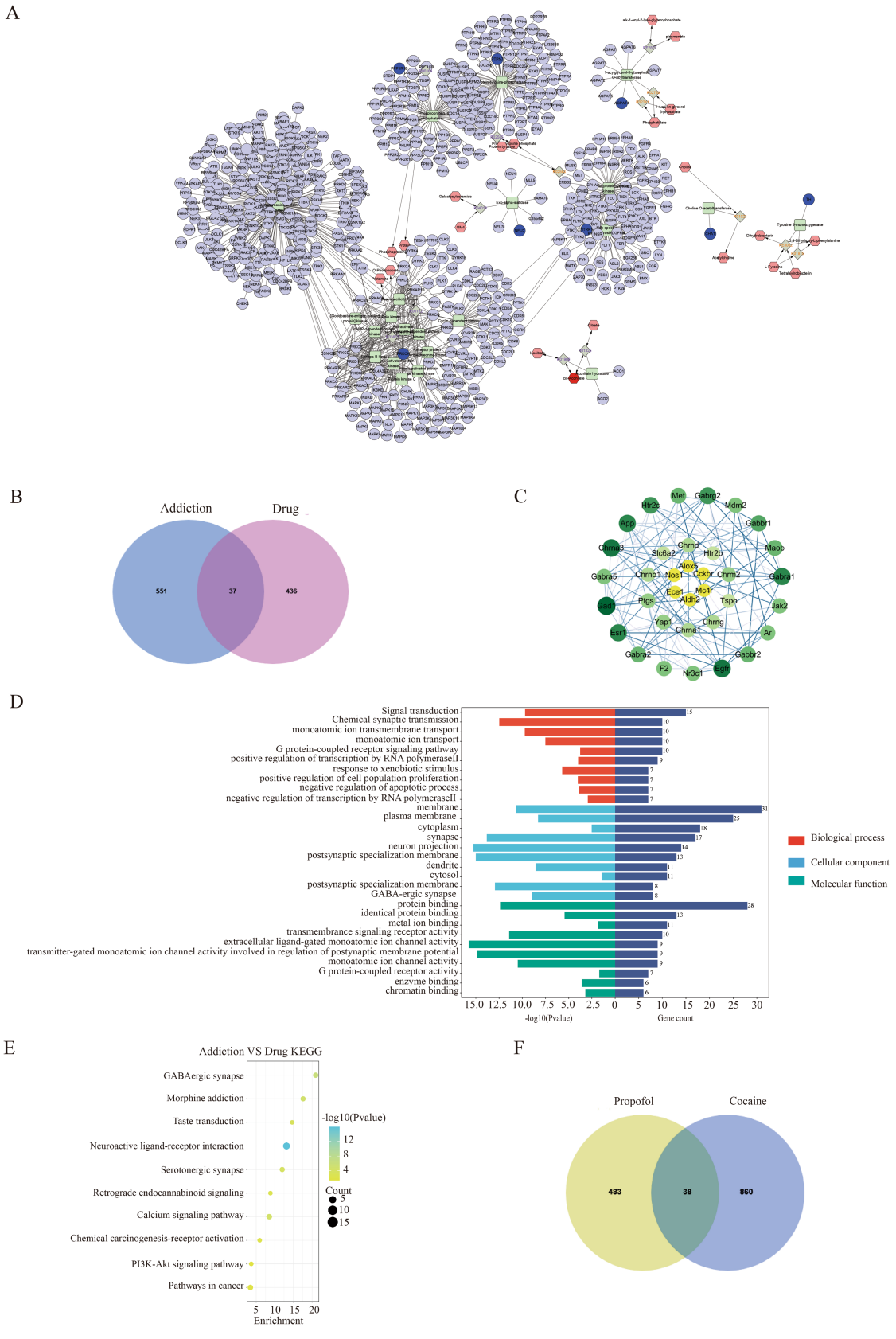


FIGURE 4
Network pharmacology and differential gene-metabolite association analysis of propofol and addiction. **(A)** Gene-metabolite network mapping with cytoscape. **(B)** Venn of targets of action of propofol versus targets of action of addiction. **(C)** Protein interactions map between propofol targets of action and targets of addiction. **(D)** GO enrichment analysis of the common targets of action of propofol and addiction. **(E)** KEGG pathway-enriched bubble map of the intersection of propofol targets of action and targets of addictive action. **(F)** Venn of propofol addiction targets versus cocaine addiction targets.

also provided important data for the discovery of therapeutic targets and biomarkers for propofol addiction.

3.6 Network pharmacology and molecular docking analysis

In this study, 473 propofol action targets and 588 addiction targets were obtained by database screening, and Venn diagram analysis showed that 37 targets were shared between the two (Figure 4B). These targets were collated and visualized by a protein–protein interaction (PPI) network constructed from the STRING database and using Cytoscape 3.9.0 software (Figure 4C), with node color shades indicating the strength of the protein–protein interactions. The GO enrichment analysis identified 143 annotated pathways for biological processes, 51 cellular components and 56 molecular functions, and the analyses of each class The top 10 pathways were visualized (Figure 4D), involving multiple key biological processes such as signal transduction, chemical synaptic transmission, transmembrane ion transport, etc. KEGG pathway enrichment analysis screened out 23 signaling pathways, of which the top 10 were mainly involved in GABAergic synapses, morphine addiction, taste transmission, etc. (Figure 4E), and the neuroactive ligand-receptor interaction pathway was particularly prominent. Transcriptomics sequencing focused on pathways related to neurological ligand-active receptor interactions and cocaine addiction. Cross-analysis with transcriptomic data from cocaine addiction models, seen in dataset GSE108836 (29) identified 38 common targets of action (Figure 4F), and KEGG enrichment analysis further confirmed the importance of the neuroactive ligand-receptor pathway (Figure 5A). Combined with network pharmacological analysis, we screened the neuroactive receptor-ligand pathway and mapped the network of genes in the pathway, and found that the dopamine receptors DRD1 and DRD2 had a high degree of interactions (Figure 5B); therefore, molecular docking analyses of dopamine receptors DRD1 and DRD2 were performed, and the results showed that isoproterenol binds with binding energies of −6.9 and − 6.6 to DRD1 and DRD2, respectively (Figures 5C,D).

3.7 qPCR validation of differential genes

We screened eight differential genes from key pathways involved in neural ligand-receptor interactions for real-time fluorescence quantitative qPCR validation. These genes included DRD1, DRD2, TH, TRH, PPP1R1B, CHAT, RGS9, and GPR6. qPCR validation showed that the expression changes of these genes in the hippocampal region were consistent with the transcriptome sequencing results, thus validating the accuracy of the transcriptome data. This result is detailed in Figure 5E, which provides a solid experimental basis for further exploring the molecular mechanism of propofol addiction.

3.8 Western blot to verify the expression of dopamine receptor

In order to verify the expression of dopamine receptors in the hippocampus, protein immunoblotting was performed for verification, and the results showed that the expression of dopamine DRD1 and DRD2 in the hippocampus of mice in the propofol and Alcohol groups was decreased compared with that in the Saline group, which was consistent with the results of the sequencing analysis (Figures 5F,G).

4 Conclusion

Through this study, it was found that propofol had mental dependence, and the addictive effect was the highest at the dose of 75 mg/kg, and the addictive effect decreased after a certain dose, which provided a basis for the future use of propofol. Transcriptome and network pharmacology showed that propofol addiction caused significant expression of neuroactive ligand receptor pathway, and the tricarboxylic acid cycle and alanine, aspartate, and glutamate pathways in the hippocampus were significantly up-regulated. At the same time, *Campylobacter*, *Bacteroidetes*, *Actinobacteria*, and *Verrucomicrobia* in the intestinal flora were significantly increased. Therefore, it is inferred that the increase of SCFAS in the gut interacts with the activity of dopamine neurotransmitters in the hippocampal fatty acid metabolism, amino acid metabolism and neuroactive receptor ligand pathway.

5 Discussion

In this study, we combined transcriptomic, metabolomic, and metagenomic analyses of the hippocampus to provide new insights into the molecular mechanisms of propofol abuse addiction. These findings reveal the central role of the neural ligand-receptor interaction pathway in propofol addiction, especially the changes of dopamine neurotransmitters in the neuroreceptor pathway, and provide new insights into the detailed mechanism of propofol addiction.

In the analysis of transcriptomics results, significant changes in gene expression in the hippocampus following propofol addiction were observed, and these changes were mainly focused on the neuroactive ligand receptor pathway. Specifically, we found significant changes in the expression of the dopamine receptors DRD1 and DRD2, GPR6, and RGS9, which are closely related to neuroadaptive changes, synaptic plasticity, signaling, and other functions. These findings echo the neurobiological model proposed by Koob and Le Moal et al., which posits that addiction is a vicious cycle driven by a decline in the function of the brain reward system and activation of the anti-reward system, where chronic drug exposure leads to a decline in the function of the reward neurotransmitter system and concomitant activation of the anti-reward system, which induces down-regulation of dopamine receptor expression, thereby increasing the risk of drug craving and relapse (30, 31), causing downregulation of dopamine receptor expression and a high risk of drug craving and relapse. In addition, GPR6, a G protein-coupled receptor, has been identified as a novel therapeutic molecular target for cannabidiol, which provides new therapeutic perspectives (32), RGS9 and its specific splice variant RGS9-2 play roles in the regulation of morphine reward and dependence (33). Studies of the neuroactive ligand pathway have revealed that the pathway contains a variety of neurotransmitter systems including dopamine, endorphins, glutamate, norepinephrine, 5-hydroxytryptamine, and gamma-aminobutyric acid. Among these systems, the dopamine system plays a central role in the regulation of motor, emotional, and

10.3389/fmed.2025.1539467

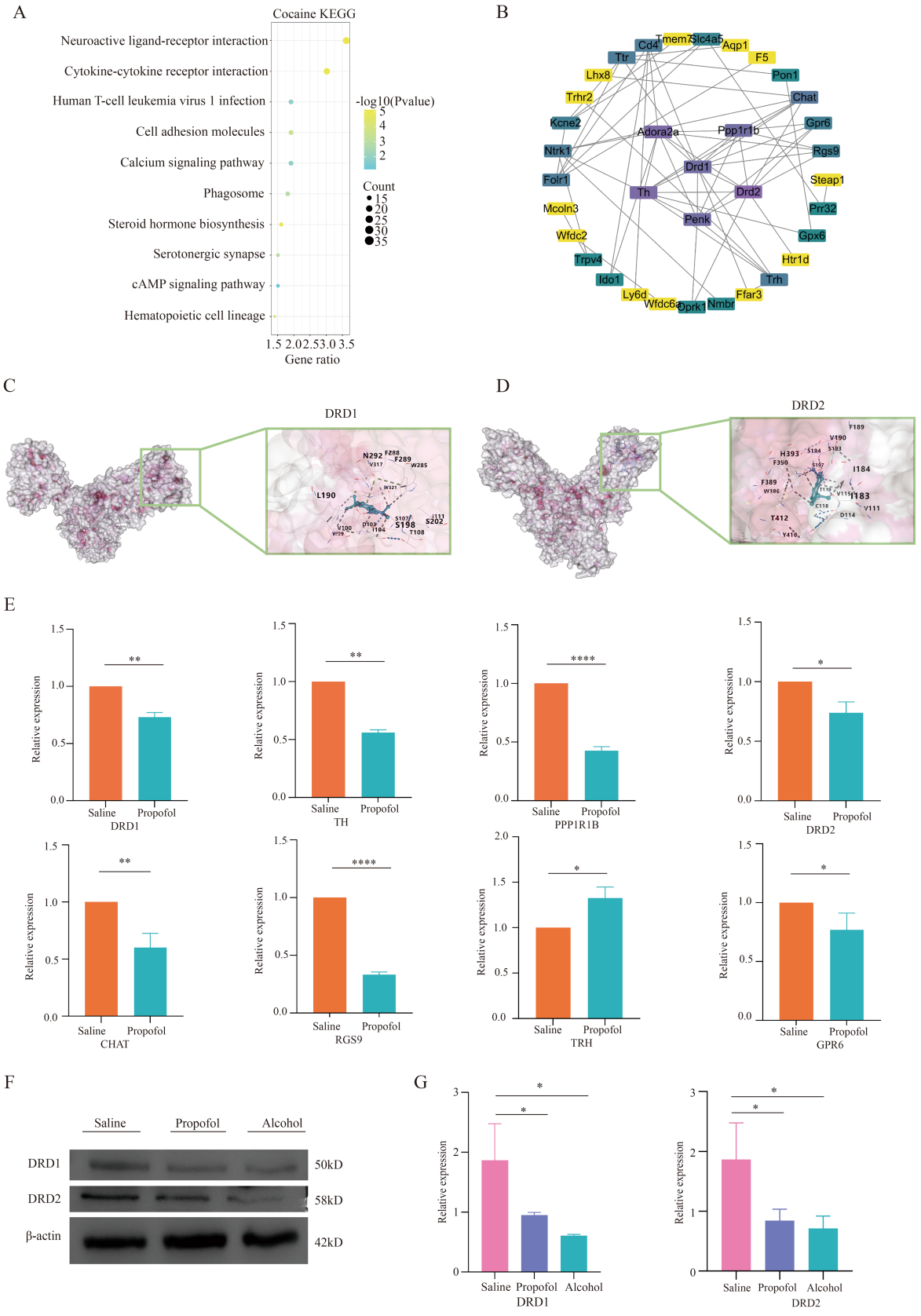


FIGURE 5
Validation of differential gene correlations in the neuroactive receptor-ligand interaction pathway. **(A)** KEGG-enriched bubble plot of cocaine addiction targets. **(B)** Protein interaction plots of genes on the neuroactive receptor-ligand pathway. **(C)** Diagram of molecular docking models for dopamine receptor 1. **(D)** Diagram of molecular docking models for dopamine receptor 2. **(E)** QPCR validation of differential genes in significant pathways regarding propofol addiction. *$p < 0.05$, $p < 0.01$ and $p < 0.001$ and ***$p < 0.0001$. **(F,G)** Western blotting validation of dopamine receptor 1 and dopamine receptor 2 regarding propofol addiction. *$p < 0.05$, $p < 0.01$ and $p < 0.001$ and ***$p < 0.0001$.

reward-related behaviors. Therefore, we believe that the reciprocal regulation of neural ligand-receptor interactions between genes such as DRD1, DRD2, GPR6, and RGS9 is closely linked to the reward effects and addictive behaviors of propofol.

Metabolomics further analyzes the effects of propofol addiction on the hippocampal region, which produces metabolites including citric acid, lysophosphatidylcholine, lysophosphatidylethanolamine, methylcoumarin, and docosahexaenoic acid, all of which are involved in cellular signaling, and energy metabolism related to the hippocampus. Especially dominated by citric acid, which breaks down into a variety of short-chain fatty acids, and the citric acid cycle also produces a variety of short-chain fatty acids. Current research suggests that short-chain fatty acids have anti-inflammatory effects, are involved in G protein-coupled receptors, neurotransmitter synthesis, neuroprotection, and the brain-gut axis (34), an increase in neuroprotection, signaling, and energy metabolism was found in the chemical classification of metabolites. In addition, significant cell membrane metabolic markers found in animal models of nicotine addiction and methamphetamine addiction were phosphatidylcholine (35, 36), and significant changes in energy metabolism-related metabolites such as citric acid cycle products and intermediates were similarly found in human serum as well as in the hippocampus of methamphetamine addicts (37), thus changes in metabolites in the hippocampal region of propofol addiction reflect the effects of propofol on energy metabolism pathways, neural signaling pathways involved in the production, activation, and functioning of neurotransmitters in the neural ligand-receptor interaction pathway to provide energetic substances, and reflect adaptive changes in neuronal cells in response to chronic exposure to propofol, which may further affect neurotransmitter function and neural network stability. This is consistent with existing findings.

In addition to transcriptomics and metabolomics, intestinal microbial macro-genomics sequencing was performed, and the results revealed that propofol addiction resulted in significant changes in the species abundance and composition of microorganisms in the intestine, such as a decrease in *Thick-walled phyla* and an increase in *Anopheles* and *Actinomycetes*, which were found to be the main phyla of the intestinal tract, and that the *Thick-walled phyla* and the *Anopheles* produce, by different means, SCFA, which can affect the brain by acting on G protein-coupled receptors expressed by cells in the intestine. For example, short-chain fatty acids act through G protein-coupled receptors such as FFAR2 and FFAR3, or by inhibiting histone deacetylase activity (38), and an increase in the *actinomycete phylum* improves host resistance to disease and maintains immune stability in the intestinal environment. Gut microbes can affect the immune system, including influencing the activation of immune cells and the production of cytokines. These cytokines can cross the blood–brain barrier and affect neuroinflammation and the activation state of microglia in the brain, which in turn affects neurotransmitter homeostasis. For example, anaerobic bacteria of the phylum *Actinobacteria* such as *Bifidobacteria*, *Propionibacteria*, *Corynebacteria*, and *Streptomyces* modulate the immune-inflammatory response by inducing regulatory T cells (39). In addition *Actinobacteria phylum* has the ability to produce antibiotics, which helps to inhibit the growth of pathogenic microorganisms, and is also involved in the synthesis of vitamins in the intestinal tract and maintenance of intestinal barrier function (40). Additionally the increase in *Lactobacillus* gates suggests that there may be vagal involvement in the brain-gut connection, and that certain specific gut microbes, such as *Lactobacillus rohita*, can transmit signals to the microbe-gut-brain axis via the vagus nerve, thereby affecting neuroendocrine metabolism and altering neurotrophic proteins, neurotransmitters in the hippocampus (41, 42). The gut microbiota communicates bi-directionally with the brain via the gut-brain axis. For example, changes in the metabolism of tryptophan, a precursor for the synthesis of the neurotransmitter 5-hydroxytryptophan, may affect mood and behavior (43). The mechanisms by which the gut microbiota influences mood and behavior through the gut-brain axis are multifaceted and involve complex interactions between the nervous, endocrine and immune systems.

These findings further underscore the pivotal role of the brain-gut axis in drug addiction and elucidate the intricate regulatory mechanisms between gut microbiota and hippocampal neural function. A key innovation of this study lies in the selection of propofol, a widely used clinical anesthetic, as the research subject, combined with neuroomics and gut microbiome analysis to explore the molecular mechanisms underlying drug addiction. However, we acknowledge several limitations, including the constrained scope of experimental data and the absence of large-scale dataset validation, which may restrict the generalizability and long-term applicability of our findings. Moreover, the precise mechanisms governing the interactions between gut microbiota and the host nervous system remain to be fully elucidated. Future studies should employ larger sample sizes and longitudinal designs to comprehensively unravel the molecular mechanisms of propofol addiction, thereby providing a more robust theoretical foundation and practical insights for clinical interventions.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: Transcriptomics data can be found at the NCBI database under the accession number: BioProject: PRJNA1237536. Metagenomics data has also been uploaded to the NCBI database with the accession number: PRJNA1237534. Metabolomics data has been submitted to The MetaboLights Team, EMBL-EBI, with the project ID: MTBLS12337. Further inquiries can be directed to the corresponding author/s.

Ethics statement

The animal study was approved by the Ethics Committee of Qinghai University School of Medicine (2022-01). The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

LW: Conceptualization, Data curation, Formal analysis, Methodology, Validation, Visualization, Writing – original draft. TW: Formal analysis, Investigation, Writing – review & editing. YaL: Conceptualization, Formal analysis, Writing – review & editing. YS: Resources, Software, Writing – review & editing. YuL: Formal analysis, Visualization, Writing – review & editing. ZW: Investigation, Writing – review & editing. QW: Resources, Software, Writing – review & editing. SZ: Investigation, Writing – review & editing. HW:

Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by research funding from the National Natural Science Foundation of China (32260187 to HW), the Applied Basic Research Project Fund of the Qinghai Provincial Department of Science and Technology (2023-ZJ-729 to HW), and the Qinghai University Team Research Project Fund (2023-QYY-6 to HW).

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.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2025.1539467/full#supplementary-material

References

1. Lundström S, Twycross R, Mihalyo M, Wilcock A. Propofol. *J Pain Symptom Manag*. (2010) 40:466–70. doi: 10.1016/j.jpainsymman.2010.07.001

2. Kim ED, Wu X, Lee S, Tibbs GR, Cunningham KP, di Zanni E, et al. Propofol rescues voltage-dependent gating of HCN1 channel epilepsy mutants. *Nature*. (2024) 632:451–9. doi: 10.1038/s41586-024-07743-z

3. Kotani Y, Shimazawa M, Yoshimura S, Iwama T, Hara H. The experimental and clinical pharmacology of propofol, an anesthetic agent with neuroprotective properties. *CNS Neurosci Ther*. (2008) 14:95–106. doi: 10.1111/j.1527-3458.2008.00043.x

4. Burnett GW, Taree A, Martin L, Bryson EO. Propofol misuse in medical professions: a scoping review. *Canadian J. Anesthesia*. (2022) 70:395–405. doi: 10.1007/s12630-022-02382-2

5. Xiong M, Shiwalkar N, Reddy K, Shin P, Bekker A. Neurobiology of propofol addiction and supportive evidence: what is the new development? *Brain Sci*. (2018) 8:36. doi: 10.3390/brainsci8020036

6. Yang M, Zhang Y. Propofol addiction: the mechanism issues we need to know. *Anesthesiol. Perioperat. Sci.* (2024) 2:1–5. doi: 10.1007/s44254-023-00046-y

7. Kalin NH. Substance use disorders and addiction: mechanisms, trends, and treatment implications. *Am J Psychiatry*. (2020) 177:1015–8. doi: 10.1176/appi.ajp.2020.20091382

8. Berke JD, Eichenbaum HB. Drug addiction and the hippocampus. *Science*. (2001) 294:1235. doi: 10.1126/science.294.5545.1235a

9. Peyton L, Oliveros A, Choi DS, Jang MH. Hippocampal regenerative medicine: neurogenic implications for addiction and mental disorders. *Exp Mol Med*. (2021) 53:358–68. doi: 10.1038/s12276-021-00587-x

10. Kutlu MG, Gould TJ. Effects of drugs of abuse on hippocampal plasticity and hippocampus-dependent learning and memory: contributions to development and maintenance of addiction. *Learn Mem*. (2016) 23:515–33. doi: 10.1101/lm.042192.116

11. Lingford-Hughes A, Nutt D. Neurobiology of addiction and implications for treatment. *Br J Psychiatry*. (2003) 182:97–100. doi: 10.1192/bjp.182.2.97

12. Wise RA, Robble MA. Dopamine and addiction. *Annu Rev Psychol*. (2020) 71:79–106. doi: 10.1146/annurev-psych-010418-103337

13. Volkow ND, Michaelides M, Baler R. The neuroscience of drug reward and addiction. *Physiol Rev*. (2019) 99:2115–40. doi: 10.1152/physrev.00014.2018

14. Ghanbari R, Sumner S. Using metabolomics to investigate biomarkers of drug addiction. *Trends Mol Med*. (2018) 24:197–205. doi: 10.1016/j.molmed.2017.12.005

15. Wang D, Feng Y, Yang M, Sun H, Zhang Q, Wang R, et al. Variations in the oral microbiome and metabolome of methamphetamine users. *mSystems*. (2023) 9:e0099123. doi: 10.1128/msystems.00991-23

16. Zhu R, Yang T, Kobeissy F, Mouhieddine TH, Raad M, Nokkari A, et al. The effect of chronic methamphetamine exposure on the hippocampal and olfactory bulb neuroproteomes of rats. *PLoS One*. (2016) 11:e0151034. doi: 10.1371/journal.pone.0151034

17. Lei Y, Tang R, Xu J, Wang W, Zhang B, Liu J, et al. Applications of single-cell sequencing in cancer research: progress and perspectives. *J Hematol Oncol*. (2021) 14:91. doi: 10.1186/s13045-021-01105-2

18. Yao Z, van Velthoven CTJ, Kunst M, Zhang M, McMillen D, Lee C, et al. A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. *Nature*. (2023) 624:317–32. doi: 10.1038/s41586-023-06812-z

19. Wang S-C, Chen YC, Chen SJ, Lee CH, Cheng CM. Alcohol addiction, gut microbiota, and alcoholism treatment: a review. *Int J Mol Sci*. (2020) 21:6413. doi: 10.3390/ijms21176413

20. Qian S, Liu H, Wei H, Liu J, Li X, Luo X. Toxic effects of prolonged propofol exposure on cardiac development in zebrafish larvae. *BMC Anesthesiol*. (2025) 25:81. doi: 10.1186/s12871-025-02942-1

21. Nummela AJ, Laaksonen LT, Laitio TT, Kallionpää RE, Långsjö JW, Scheinin JM, et al. Effects of dexmedetomidine, propofol, sevoflurane and S-ketamine on the human metabolome: a randomised trial using nuclear magnetic resonance spectroscopy. *Eur J Anaesthesiol*. (2022) 39:521–32. doi: 10.1097/EJA.0000000000001591

22. Petrut S-M, Bragaru AM, Munteanu AE, Moldovan A-D, Moldovan C-A, Rusu E. Gut over mind: exploring the powerful gut–brain Axis. *Nutrients*. (2025) 17:842. doi: 10.3390/nu17050842

23. Zheng R, Chen Y, Zhang J, Liu Q, Zheng Y, Wang Z. Prelimbic cortex is involved in the regulation of morphine-induced conditioned place preference in both resistant and sensitive mice. *Sci Rep*. (2025) 15:5596. doi: 10.1038/s41598-025-87084-7

24. Barattini AE, Pahng AR. Interactions of pain and opioids on conditioned place preference in rodents. *Psychopharmacology*. (2025) 242:1–26. doi: 10.1007/s00213-024-06719-1

25. Zhai Y, Liu L, Zhang F, Chen X, Wang H, Zhou J, et al. Network pharmacology: a crucial approach in traditional Chinese medicine research. *Chin Med*. (2025) 20:8. doi: 10.1186/s13020-024-01056-z

26. Zhu H, Wang R, Hua H, Cheng Y, Guo Y, Qian H, et al. Network pharmacology exploration reveals gut microbiota modulation as a common therapeutic mechanism for anti-fatigue effect treated with Maca compounds prescription. *Nutrients*. (2022) 14:1533. doi: 10.3390/nu14081533

27. Hou F, Yu Z, Cheng YH, Liu Y, Liang S, Zhang F. Deciphering the pharmacological mechanisms of Scutellaria baicalensis Georgi on oral leukoplakia by combining network pharmacology, molecular docking and experimental evaluations. *Phytomedicine*. (2022) 103:154195. doi: 10.1016/j.phymed.2022.154195

28. Shang L, Wang Y, Li J, Zhou F, Xiao K, Liu Y, et al. Mechanism of Sijunzi decoction in the treatment of colorectal cancer based on network pharmacology and experimental validation. *J Ethnopharmacol*. (2023) 302:115876. doi: 10.1016/j.jep.2022.115876

29. Wang Y, Teng H, Sapozhnikov DM, du Q, Zhao M. Transcriptome sequencing reveals candidate NF-KB target genes involved in repeated cocaine administration. *Int J Neuropsychopharmacol*. (2018) 21:697–704. doi: 10.1093/ijnp/pyy031

30. Koob GF, Moal ML. Neurobiological mechanisms for opponent motivational processes in addiction. *Philos. Trans. Royal Society B Biol. Sci.* (2008) 363:3113–23. doi: 10.1098/rstb.2008.0094

31. Koob GF, Moal ML. Addiction and the brain antireward system. *Annu Rev Psychol*. (2007) 59:29–53. doi: 10.1146/annurev.psych.59.103006.093548

32. Laun AS, Shrader SH, Brown KJ, Song ZH. GPR3, GPR6, and GPR12 as novel molecular targets: their biological functions and interaction with cannabidiol. *Acta Pharmacol Sin*. (2018) 40:300–8. doi: 10.1038/s41401-018-0031-9

33. Gaspari S, Papachatzaki MM, Koo JW, Carr FB, Tsimpanouli ME, Stergiou E, et al. Nucleus Accumbens-specific interventions in RGS9-2 activity modulate responses to morphine. *Neuropsychopharmacology*. (2014) 39:1968–77. doi: 10.1038/npp.2014.45

34. O'Riordan KJ, Collins MK, Moloney GM, Knox EG, Aburto MR, Fülling C. Short chain fatty acids: microbial metabolites for gut-brain axis signalling. *Mol Cell Endocrinol*. (2022) 546:111572. doi: 10.1016/j.mce.2022.111572

35. Li H, Chen B, Shao X, Hu Z, Deng Y, Zhu R, et al. 1H-nuclear magnetic resonance-based metabolomic analysis of brain in mice with nicotine treatment. *BMC Neurosci*. (2014) 15:32. doi: 10.1186/1471-2202-15-32

36. Xu J, Zhu Z, Jin Y, Wei C, Wang Y, Li X. Effect of aerobic exercise on brain metabolite profiles in the mouse models of methamphetamine addiction: LC-MS-based metabolomics study. *BMC Psychiatry*. (2023) 23:852. doi: 10.1186/s12888-023-05351-1

37. Lin M, Xu J, Liu X, Dai Z, Liu Z, Zhao X, et al. Metabolomics profiling of methamphetamine addicted human serum and three rat brain areas. *RSC Adv*. (2019) 9:41107–19. doi: 10.1039/C9RA08096A

38. Silva YP, Bernardi A, Frozza RL. The role of short-chain fatty acids from gut microbiota in gut-brain communication. *Front Endocrinol*. (2020) 11:25. doi: 10.3389/fendo.2020.00025

39. Noronha NY, Noma IHY, Fernandes Ferreira R, Rodrigues GS, Martins LS, Watanabe LM, et al. Association between the relative abundance of phyla actinobacteria, vitamin C consumption, and DNA methylation of genes linked to immune response pathways. *Front Nutr*. (2024) 11:1373499. doi: 10.3389/fnut.2024.1373499

40. Zothanpuia N, Passari AK, Leo VV, Chandra P, Kumar B, Nayak C, et al. Bioprospection of actinobacteria derived from freshwater sediments for their potential to produce antimicrobial compounds. *Microb Cell Factories*. (2018) 17:68. doi: 10.1186/s12934-018-0912-0

41. Desbonnet L, Clarke G, Shanahan F, Dinan TG, Cryan JF. Microbiota is essential for social development in the mouse. *Mol Psychiatry*. (2013) 19:146–8. doi: 10.1038/mp.2013.65

42. Agirman G, Yu KB, Hsiao EY. Signaling inflammation across the gut-brain axis. *Science*. (2021) 374:1087–92. doi: 10.1126/science.abi6087

43. Waclawiková B, Aidy SE. Role of microbiota and tryptophan metabolites in the remote effect of intestinal inflammation on brain and depression. *Pharmaceuticals*. (2018) 11:63. doi: 10.3390/ph11030063

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 23 April 2025
DOI 10.3389/fmed.2025.1510793

Check for updates

OPEN ACCESS

EDITED BY
HaiHui Huang,
Shaoguan University, China

REVIEWED BY
Jayanta Mondal,
University of Texas MD Anderson Cancer Center, United States
Priyadarshini Mamindla,
University of Pittsburgh, United States

*CORRESPONDENCE
Zhengyong Li
lizydd@sina.com
Zhenyu Zhang
zhangzhenyu@scu.edu.cn

†These authors have contributed equally to this work

RECEIVED 13 October 2024
ACCEPTED 31 March 2025
PUBLISHED 23 April 2025

CITATION
Huang Y, Chen L, Zhang Z, Liu Y, Huang L, Liu Y, Liu P, Song F, Li Z and Zhang Z (2025) Integration of histopathological image features and multi-dimensional omics data in predicting molecular features and survival in glioblastoma.
Front. Med. 12:1510793.
doi: 10.3389/fmed.2025.1510793

COPYRIGHT
© 2025 Huang, Chen, Zhang, Liu, Huang, Liu, Liu, Song, Li 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.

Integration of histopathological image features and multi-dimensional omics data in predicting molecular features and survival in glioblastoma

Yeqian Huang[1†], Linyan Chen[2†], Zhiyuan Zhang[3†], Yu Liu[1], Leizhen Huang[1], Yang Liu[1], Pengcheng Liu[1], Fengqin Song[1], Zhengyong Li[1,4*] and Zhenyu Zhang[1,4*]

[1]Department of Burn and Plastic Surgery, West China Hospital, Sichuan University, Chengdu, China, [2]Department of Biotherapy, Cancer Center and State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, Chengdu, China, [3]West China School of Medicine, West China Hospital, Sichuan University, Chengdu, China, [4]Department of Plastic Reconstructive and Aesthetic Surgery, West China Tianfu Hospital, Sichuan University, Chengdu, China

Objectives: Glioblastoma (GBM) is a highly malignant brain tumor with complex molecular mechanisms. Histopathological images provide valuable morphological information of tumors. This study aims to evaluate the predictive potential of quantitative histopathological image features (HIF) for molecular characteristics and overall survival (OS) in GBM patients by integrating HIF with multi-omics data.

Methods: We included 439 GBM patients with eligible histopathological images and corresponding genetic data from The Cancer Genome Atlas (TCGA). A total of 550 image features were extracted from the histopathological images. Machine learning algorithms were employed to identify molecular characteristics, with random forest (RF) models demonstrating the best predictive performance. Predictive models for OS were constructed based on HIF using RF. Additionally, we enrolled tissue microarrays of 67 patients as an external validation set. The prognostic histopathological image features (PHIF) were identified using two machine learning algorithms, and prognosis-related gene modules were discovered through WGCNA.

Results: The RF-based OS prediction model achieved significant prognostic accuracy (5-year AUC = 0.829). Prognostic models were also developed using single-omics, the integration of HIF and single-omics (HIF + genomics, HIF + transcriptomics, HIF + proteomics), and all features (multi-omics). The multi-omics model achieved the best prediction performance (1-, 3- and 5-year AUCs of 0.820, 0.926 and 0.878, respectively).

Conclusion: Our study indicated a certain prognostic value of HIF, and the integrated multi-omics model may enhance the prognostic prediction of GBM, offering improved accuracy and robustness for clinical application.

KEYWORDS

glioblastoma, histopathological image, genomics, transcriptomics, proteomics, prognosis

1 Introduction

Glioma is the most prevalent primary malignant tumor of the brain, accounting for 40–50% of intracranial tumors (1). Glioblastoma (GBM), classified as a WHO grade IV glioma, is the most common (57.3% of all gliomas) and aggressive form of glioma in adults (2, 3). The age-adjusted incidence rate of GBM is 3.22 per 100,000 population, with a median overall survival (OS) of 12–15 months with standard treatment, while population studies suggest a median survival of 8–10 months (4, 5). Approximately 7% of GBM patients live for at least 5 years after diagnosis, defined as long-term survivors (LTS) in previous research (6–8). Conventional treatments of GBM include maximal surgical resection, postoperative radiotherapy and chemotherapy; however, complete tumor resection is often unattainable due to the tumor's invasive nature and high recurrence rate (9). Prognostic factors such as tumor stage, age, pathological grade, KPS, extent of resection and certain molecular markers have been identified as key indicators of GBM prognosis (10, 11). Therefore, as a cancer characterized by multiple genetic and pathway alterations, further investigation into comprehensive prognostic markers is critical for guiding risk stratification, clinical treatment decisions and survival prediction in GBM patients.

GBM derives from glial cells and neurons and exhibits a complex gene expression profile with various molecular alterations that drive its oncogenesis and progression (12). Notably, isocitrate dehydrogenase-1 (IDH-1) and IDH-2 mutations are observed in primary (6%) and secondary (70%) GBMs (13). Compared with IDH1 wild-type, the survival of IDH1 mutant high-grade glioma patients is significantly prolonged (14). The O6-methylguanine-DNA methyltransferase (MGMT) coded protein involved in methylated bases and DNA repair and the methylation status of MGMT promoter may be a significant predictor for sensitivity to chemotherapy or radiotherapy (15, 16). Telomerase reverse transcriptase (TERT) can activate telomerase to keep the telomeres intact and promote cell proliferation. IDH1 mutant gliomas with mutations in TERT promoter have exhibited better prognosis (17). Alpha thalassemia/X-linked intellectual disability (ATRX) is also discovered as a mutational cancer driver in GBM (18). GBM can be classified into subtypes based on molecular features, including transcriptional profiles (classical, mesenchymal, neural, proneural), genetic mutations (e.g., IDH1 mutations), and epigenetic alterations (e.g., CpG island methylator phenotype, CIMP) and so on (19, 20). Therefore, establishing a comprehensive and effective biomarker will be of great benefit to prognostic prediction and therapeutic strategies for GBM patients.

In clinical practice, in addition to imaging examinations such as CT and MRI, the final diagnosis is confirmed through histopathological biopsy following tumor resection. Histopathological images obtained from H&E-stained tumor tissue slides are routinely used in definite diagnosis and staging of different cancers. The development of computer-assisted medical image processing and analysis systems is increasingly employed in digital pathological image assessment. These systems can accurately and reproducibly capture morphological, structural, and compositional changes in tissues and cells, reducing the subjectivity associated with traditional pathologist assessments (21). Commonly extracted histopathological image features such as texture structure, gray level distribution and morphological features including the size and shape of cell and nuclei, have demonstrated potential in pathological diagnosis, classification and prognosis of human cancers such as breast cancer (22), colorectal cancer (23) and lung cancer (24). In addition to histopathological images, omics profiles such as genomics, transcriptomics and proteomics have also been applied to patient stratification and prognostic prediction. Integrating histopathological image features with multi-omics data has shown promise in various cancers, including renal cancer (25), lung cancer (26) and head and neck squamous cell carcinoma (27). Therefore, exploring the integration of histopathological image features with omics data holds significant potential for prognostic prediction in clinical settings.

In this study, we focused on the analyses of histopathological image features (HIF) and their correlation with genomic and transcriptomic profiles, which has not been explicitly demonstrated in GBM. We first assessed the overall capacity of HIF in classifying somatic mutations, molecular and methylation subtypes of GBM via different machine learning approaches. Subsequently, we identified the prognosis-related histopathological image features and evaluated the underlying correlation with gene expression profiles. Finally, we constructed survival prediction models based on various omics profiles and their integration. We validate these models with both an internal test cohort and an external validation cohort, expecting to enhance the accuracy of prognostic prediction for GBM patients.

2 Materials and methods

2.1 Study design and data acquisition

The overall framework of the study is illustrated in Figure 1, and the specific process is described in the following sections. We obtained a cohort of GBM samples with accessible clinical information, genomics and transcriptomics data from The Cancer Genome Atlas (TCGA) data portal[1] and matched proteomics profile from The Cancer Proteome Atlas (TCPA) repository.[2] The corresponding H&E histopathological images were obtained from The Cancer Imaging Archive (TCIA).[3] A total of 439 GBM patients were selected from TCGA based on the completeness of clinical records and image availability of high-quality histopathological images in TCIA, excluding cases with incomplete data. All included patients had corresponding genomic, transcriptomic, and proteomic data for a comprehensive multi-omics analysis. The GBM tissue microarrays (TMA) of 67 patients with clinical and follow-up data were purchased from Shanghai Outdo Biotech Co., Ltd. (Shanghai, China). Clinical information of patients involved in TMA and TCGA cohorts is provided in Supplementary materials 2, 3.

2.2 Image processing and feature extraction

To extract the quantitative features from whole-slide histopathological images, we applied the Openslide Python library (28) to segment the images into 1,000 × 1,000 pixel sub-images.

1 https://portal.gdc.cancer.gov
2 http://tcpaportal.org/tcpa/
3 http://www.cancerimagingarchive.net/

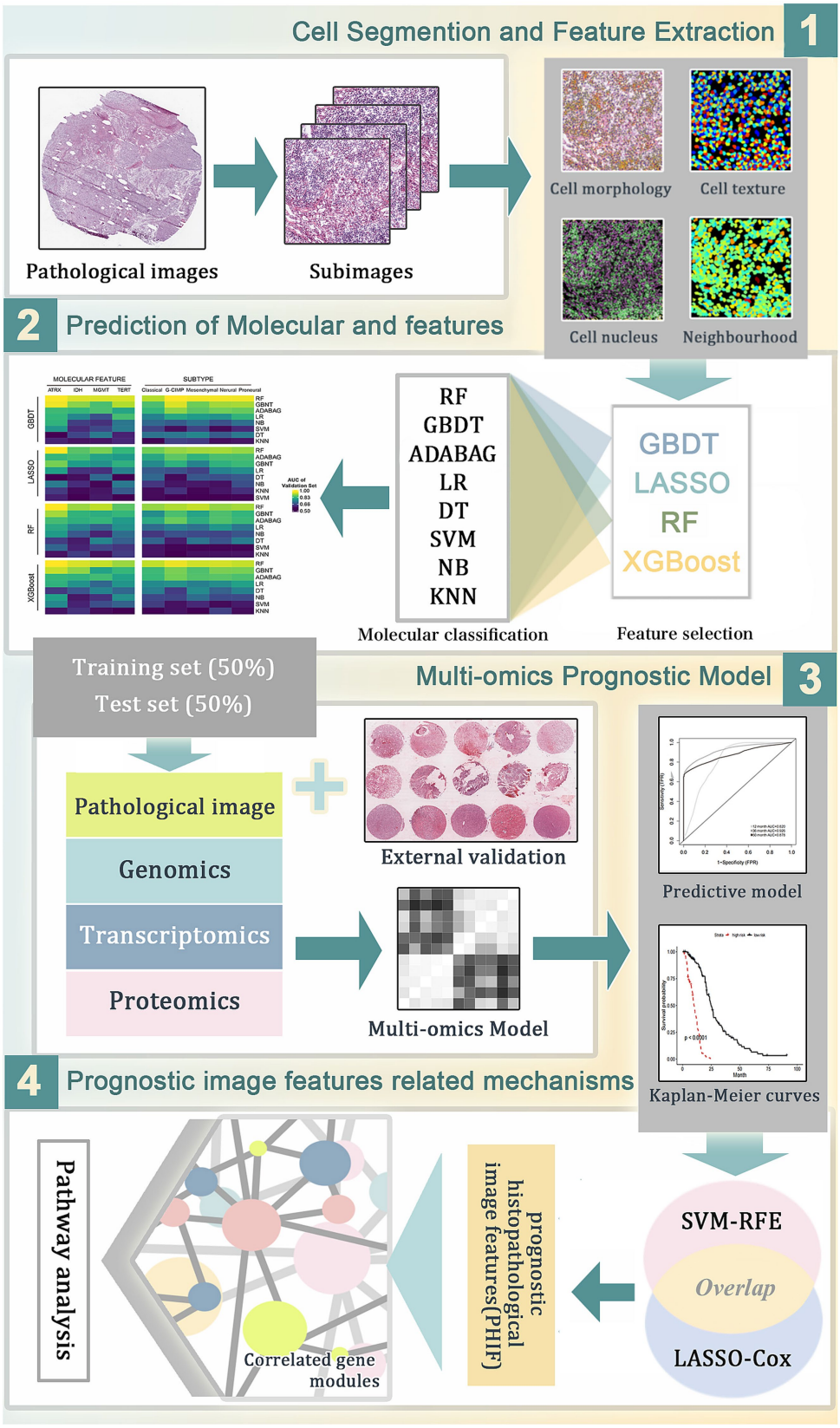


FIGURE 1
The workflow of data analysis and prognostic model construction. (1) The whole-slide histopathological images of GBM were segmented into sub-images of 1,000 × 1,000 pixels. Through CellProfiler the histopathological image features (HIF) were extracted for subsequent analyses. (2) Image feature selection and molecular features prediction based on HIF using different combinations of machine learning algorithms. (3) Construction of prognostic models for overall survival in TCGA training set based on HIF genomics, transcriptomics and proteomics data. (4) Selection of prognostic histopathological image features (PHIF) by two machine learning methods. Identification of prognostic gene modules and gene pathway analysis were performed subsequently.

Furthermore we randomly selected 50 sub-images on behalf of each patient to minimize selection bias and reduce computational load. Image feature extraction was conducted by CellProfiler (29),[4] an open-source tool for biological-image analysis. The H&E-stained images were converted to grayscale for the extraction of features, which can be specifically categorized into 10 aspects including correlation, image area occupied, image granularity, image intensity, image quality, object intensity, object neighbors, object radial distribution, object size shape and texture. In particular, the textural features were calculated by CellProfiler to quantitatively present the perceived textures of histopathological images, thereby measuring the extent and nature of textures within objects in grayscale images. Through automatic identification and segmentation, these quantitative features objectively interpret the size, shape, spatial distribution, the texture of nucleus and the relationship of pixel intensities, etc. Afterwards, each sub-image was screened to exclude irrelevant features. Eventually, a total of 550 image features were extracted, with the average feature values of 50 representative sub-images of each slide calculated for subsequent analysis.

2.3 Statistical analysis

2.3.1 Mutations and subtypes prediction

Initially, we randomly assigned the GBM samples into a training set and a test set by a ratio of 1:1 using R package "randomizr." In order to reduce overfitting caused by the large number of features, we initially employed four machine learning algorithms for feature selection to extract the most informative histopathological image features (HIFs), including least absolute shrinkage and selection operator (LASSO) (30), random forest (RF) (31), gradient boosting decision tree (GBDT) (32), and extreme gradient boosting (XGBoost) (33). Subsequently, we evaluated eight classifiers including RF, GBDT, adaptive boosting (AdaBoost) (34), logistic regression (LR) (34), decision tree (DT) (35), support vector machine (SVM) (36), naive Bayesian (NB) (37) and K-nearest neighbor (KNN) (38) to determine the optimal classification algorithm through the prediction of frequent somatic mutations (i.e., ATRX, IDH, MGMT, and TERT) and molecular subtypes defined by transcription profiles and epigenetics (i.e., classical, mesenchymal, neural, proneural, and G-CIMP) based on the selected imaging features and evaluated with 5-fold cross-validation. By applying multiple approaches, we intended to verify the feasibility and stability of the method in different algorithms. Based on the test set, the performances of trained classifiers were validated and compared respectively, among which RF demonstrated the highest predictive accuracy, as evidenced in Supplementary material 1 and Figure 2.

2.3.2 Survival analysis

For survival analysis, we divided patients in the training cohort into two groups based on the median value of individual HIFs, which was used for Kaplan–Meier survival analysis and log-rank test to compare overall survival (OS) between high-risk and low-risk groups, with $p < 0.05$ considered statistically significant. Univariate Cox regression was conducted based on all HIFs as continuous variables to determine the hazard ratio (HR) and 95% confidence interval (CI) and identify features significantly associated with overall survival.

2.3.3 Data pre-processing and feature selection

To synthetically evaluate the prognostic value of various omics data types, we included independent omics data (HIF, genomics, transcriptomics and proteomics) and integration of multiple features (HIF + genomics, HIF + transcriptomics, HIF + proteomics and HIF + omics) for further analysis. Patients were randomly distributed into training and validation sets on a ratio of 1:1, ensuring a balanced subset size for model training and independent evaluation to assess generalizability. In the training set, we first included the 100 most frequent somatic mutations to reduce the dimensionality in genomics profile for subsequent analyses. Patients with an overall survival (OS) of over 60 months were categorized into the long-term survival group, while those with an OS of 1–12 months were placed in the short-term survival group. Differentially expressed genes (DEGs) between the two groups were conducted using the limma package in R, and the top 100 significant DEGs were used for survival prediction. Additionally, Metascape[5] was employed for enrichment analysis based on the genomic profile.

2.3.4 Prognostic models construction and validation

Based on the training set, we employed the random forest (RF) algorithm with 1,000 decision trees and 5-fold cross-validation to construct prognostic models via R randomForestSRC package. The RF algorithm is a dimension reduction method that has preferable performance in accessing vast amounts of input data and gives estimates of the importance of variables. It can also conduct internal unbiased estimates of the generalization error and improve model accuracy. Meanwhile, the RF includes its own regularization through tree pruning and ensemble learning. Furthermore, we performed model validation based on the validation set through the estimation of the AUC value of time-dependent ROC. Patients were then assigned to high-risk group and low-risk group in line with the median value of risk score computed by different models. Kaplan–Meier analysis and log-rank test were performed between the groups to evaluate the prediction capacity. Moreover, we carried out the decision curve analysis (DCA) based on validation set to compare the net benefit under a range of threshold probabilities of each model.

2.4 Selection of prognosis-related histopathological image features

Two machine learning methods including least absolute shrinkage and selection operator Cox (LASSO-Cox) regression (R package "glmnet") and support vector machines-recursive feature elimination (SVM-RFE) (R package "e1071") were performed independently to identify potential informative image features related to prognostic prediction. LASSO-Cox regression applies L1 regularization, effectively reducing multicollinearity, selecting the most

4 https://cellprofiler.org/

5 http://metascape.org

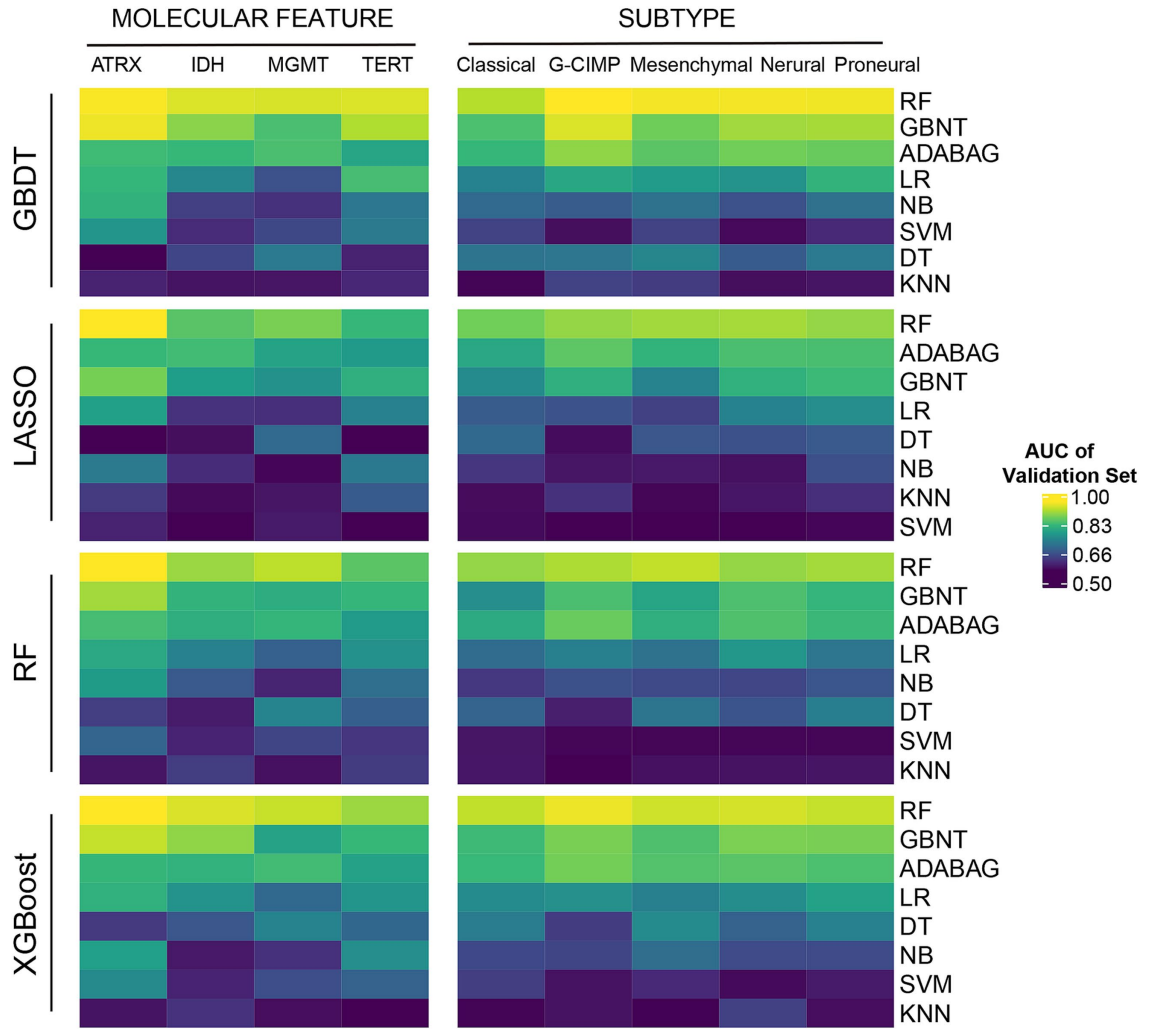


FIGURE 2
The predictive power of HIF in molecular features. Four machine learning algorithms (GBDT, LASSO, RF, and XGBoost) were applied for feature selection. Eight machine learning classifiers (RF, GBDT, Addaboost (ADABAG), LR, DT, SVM, NB, and KNN) were applied for molecular feature classification.

survival-associated features and mitigating overfitting by shrinking less relevant coefficients to zero (39). The SVM model can classify data points by maximizing the distance of the hyperplane with high accuracy, thus identifying predictive models or classifiers. SVM-RFE is a feature selection algorithm according to recursive feature deletion sequences with maximum interval principle. It ranks features based on their contribution to classification performance, iteratively eliminating the least informative ones. The integration of LASSO-Cox and SVM-RFE has been demonstrated to improve the model's generalizability and predictive performance by reducing overfitting and enhancing feature selection reliability (40, 41). Eventually, the features within the intersection of the results by two algorithms were identified as the prognostic histopathological image features (PHIF).

2.5 Gene co-expression network analysis

We performed weighted gene co-expression network analysis (WGCNA) based on training set to investigate the association of the prognostic histopathological image features and corresponding gene expression, aiming to further understand the upstream biological mechanisms. WGCNA (42) has been applied to identify modules of genes with highly correlated expression by analyzing the connections between corresponding genes and converting the expression profile into the weighted network. Co-expressed gene networks may facilitate the identification of underlying biological processes, candidate biomarkers and certain clinical traits. Additionally, we applied Metascape for enrichment analysis to estimate the interlinkage between key modules.

3 Results

3.1 Prediction performance of HIF on somatic mutations and molecular subtypes

In total we included 439 GBM patients with the matched information of histopathological images and other omics from TCGA

portal. To minimize overfitting caused by high-dimensional image features, we initially employed XGBoost, GBDT, LASSO, and RF for feature selection and extracted 550 histopathological image features (HIFs) out of the segmented tumor tissue images. Subsequently, to evaluate the clinical practicability of the 550 HIFs, we employed eight algorithms (RF, GBDT, AdaBoost, LR, DT, SVM, NB, and KNN) as classifiers in predicting four common somatic mutations (ATRX, IDH, MGMT, and TERT) and five RNA-based molecular subtypes (classical, mesenchymal, neural, proneural, and G-CIMP). We systematically compared the predictive performances of all classifiers across multiple molecular features, and RF consistently achieved the highest predictive accuracy among the eight classifiers, independent of the feature selection method used. The AUC values for RF models showed superior classification ability across all tested molecular characteristics as shown in Figure 2 and Supplementary material 1. Therefore, we selected RF as a robustly performed algorithm for subsequent prognostic model construction. Additionally, the HIF models validated by GBDT and AdaBoost (ADABAG) also achieved a relatively accurate classification effect under different feature screening methods, which indicates the clinical practicability of HIFs in distinguishing the somatic mutations and molecular subtypes of GBM.

3.2 Prognostic value evaluation of histopathological image features

To assess the correlation between histopathological image features (HIFs) and the prognosis of GBM patients, we conducted survival analyses based on individual HIFs. We first assigned the patients into two groups in line with the median value of each HIF (higher than median vs. lower than median) for survival analyses. Afterwards, we carried out univariate Cox analyses based on all HIFs to identify protective prognostic imaging factors, and the top 20 features significantly correlated with the overall survival (OS) was demonstrated in Figure 3A. The four most significant HIFs, with the smallest *p*-value included one Zernike shape feature (Median_Cells_AreaShape_Zernike_5_5) and three cell texture features (Mean_Cells_Texture_Contrast_3_45, Mean_Cells_Texture_DifferenceEntropy_3_45 and StDev_Cells_Texture_SumAverage_3_0). In particular, Zernike features are a series of 30 shape features based on Zernike polynomials, ranging from order 0 to order 9, which have been frequently extracted for representing the shape parameters in cell nucleus. Cell texture features quantify the correlations between nearby pixels in the regions of interest, which suggests that the global modes of cell nuclei and cytoplasm are all related to clinical survival outcomes. The Kaplan–Meier survival curves of four image features indicated significant differences between groups with high-value and low-value features, demonstrating the feasibility of HIFs in predicting the survival of GBM patients (Figure 3B).

Additionally, according to the expression level of the four predictive features mentioned above, we evaluated the sub-images of high-expressed and low-expressed prognostic features. We utilized TCGA internal validation and TMA external validation cohorts to assess the robustness of the predictive models and reduce the potential overfitting to the specific characteristics of the initial dataset. These validation steps serve as important safeguards against overfitting and bias, which enhances the reliability of our models across diverse datasets. The patients were identified as high-risk and low-risk groups based on the median value of risk scores, and the representative histopathological sub-images showed visible differences in TCGA and TMA external validation cohorts (Figure 3C). The image processing involving cell recognition and segmentation was conducted by CellProfiler, and different cell types were also outlined.

3.3 Integrated prognostic model of histopathological image features and genomics

To develop a more accurate predictive model for overall survival (OS) in GBM patients, we estimated the prognostic value of genetic profiles and further incorporated the HIFs with genomics data. Patients were randomly assigned into training ($n = 136$) and validation ($n = 135$) sets. To enhance the stability of the measurement, we estimated the mutation status of genes in training set and included the 100 most common somatic mutations in the prognostic model to reduce the dimensionality of the genomics data. The top 15 genes with the most frequent alterations are presented in Figure 4A. Based on the HIFs and 100 mutations we constructed prognosis-relate models in the training set. We applied time-dependent ROC in the validation set since it is more appropriate to represent time-to-event outcomes in the prognostic models compared to the classical ROC curve analysis approach (43). As illustrated in Figures 4C–E, the AUCs for histopathological image features (HIF) model exceeded those of genomics (G) model in 1-year (0.715 vs. 0.634), 3-year (0.813 vs. 0.723) and 5-year (0.829 vs. 0.692) respectively. Moreover, the integrated model of HIF and genomics (HIF + G) reached a better predictive capacity in 3-year and 5-year (AUC = 0.826 and 0.834) than the former two single-omics models. According to the median value of risk score acquired from each model, the patients were then divided into high-risk and low-risk groups. The HIF model and integrative model (HIF + G) showed more accurate prognostic performance (HR = 3.86, 95%CI: 2.67–5.30, $p < 0.001$, Figure 4) as depicted in Kaplan–Meier curves (Figure 4B).

To further validate the predictive power of the prognostic model, we implemented an external verification using the TMA-GBM cohort. Patients in the external validation set were also divided into high-risk and low-risk groups according to the median risk score. The Kaplan–Meier survival curve revealed a significant difference in survival probability between the groups ($p = 0.039$, Figure 4F). The 1-year, 3-year and 5-year AUCs of time-dependent ROC were 0.716, 0.712, and 0.703, respectively (Figure 4G). The results thus verified the prognostic capacity of the HIFs in GBM patients.

3.4 Integrated prognostic model of HIF and transcriptomics

Transcriptomics can serve as an approach for a comprehensive understanding of the interconnection between the genome, proteome, and cellular phenotype by analyzing the RNA transcripts that reflect the underlying genotype. Based on the training set, we involved 100 whole expressed mRNA genes to decrease the dimensionality and further build the transcriptomics predictive model of OS. The patients

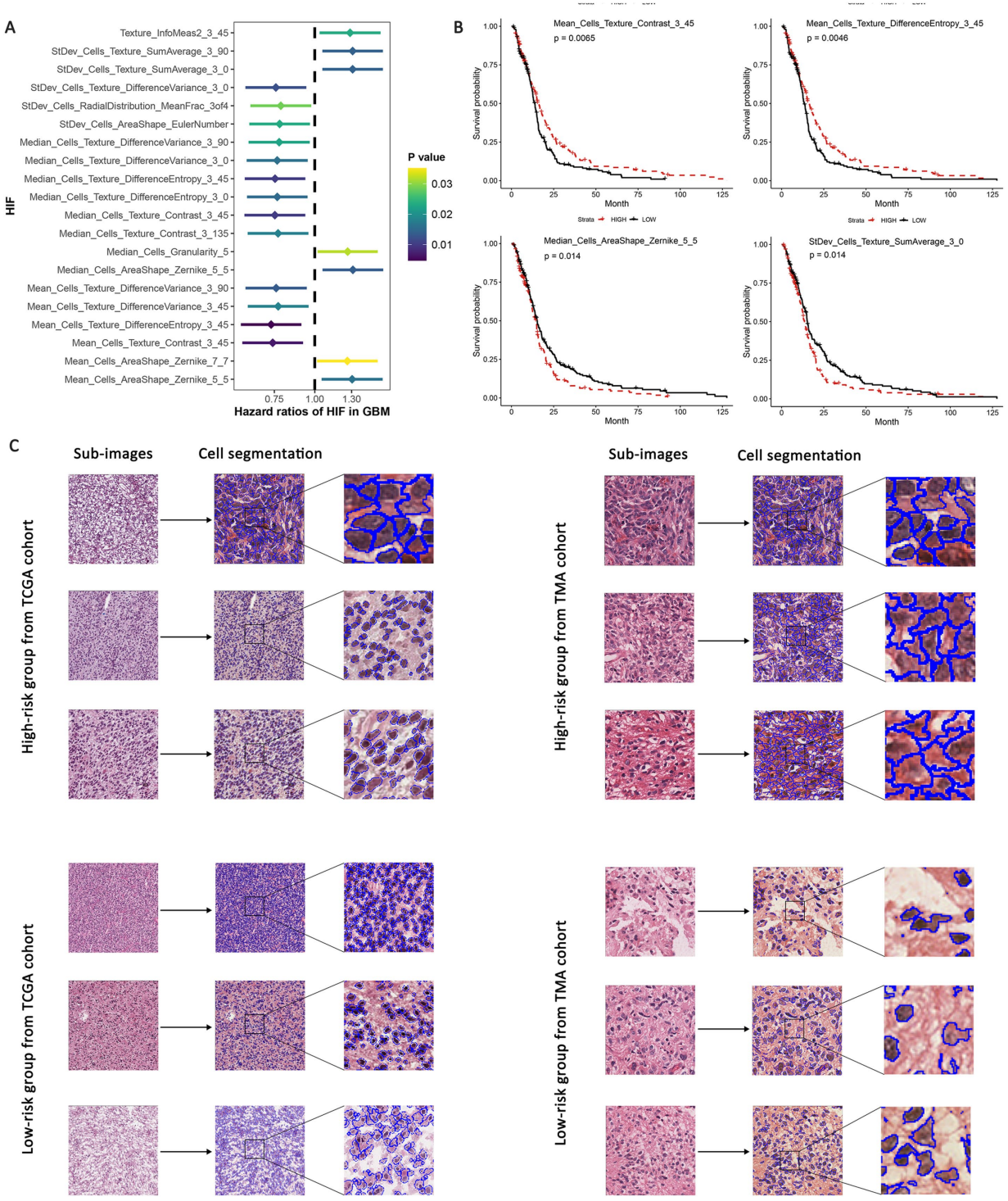


FIGURE 3
Univariate survival analyses based on HIF. GBM patients were assigned into high-risk and low-risk group according to the median value of each feature. **(A)** Hazard ratio of survival difference between two groups in univariate Cox regression. **(B)** Kaplan–Meier curves for groups with high-value and low-value "Median_Cells_AreaShape_Zernike_5_5," "Mean_Cells_Texture_Contrast_3_45," "Mean_Cells_Texture_DifferenceEntropy_3_45" and "StDev_Cells_Texture_SumAverage_3_0." **(C)** Representative sub-images of high-risk and low-risk groups in both TCGA and TMA validation cohorts.

were categorized into short-term group (deceased, 12 months ≥ OS ≥1 month) and long-term group (OS ≥60 months) according to the clinical survival status (4, 5, 7, 8). In addition, we applied Metascape for pathways enrichment in the short-term survival group based on the mRNA sequencing data (Figure 5A). Regulation of insulin-like growth factor (IGF) transport and uptake by insulin-like growth factor binding proteins (IGFBPs) has been proven to modulate essential cellular processes and be implicated in certain disorders including malignant, metabolic and immune diseases (44, 45). Previous studies have reported the potential effect of IGF in biological processes associated

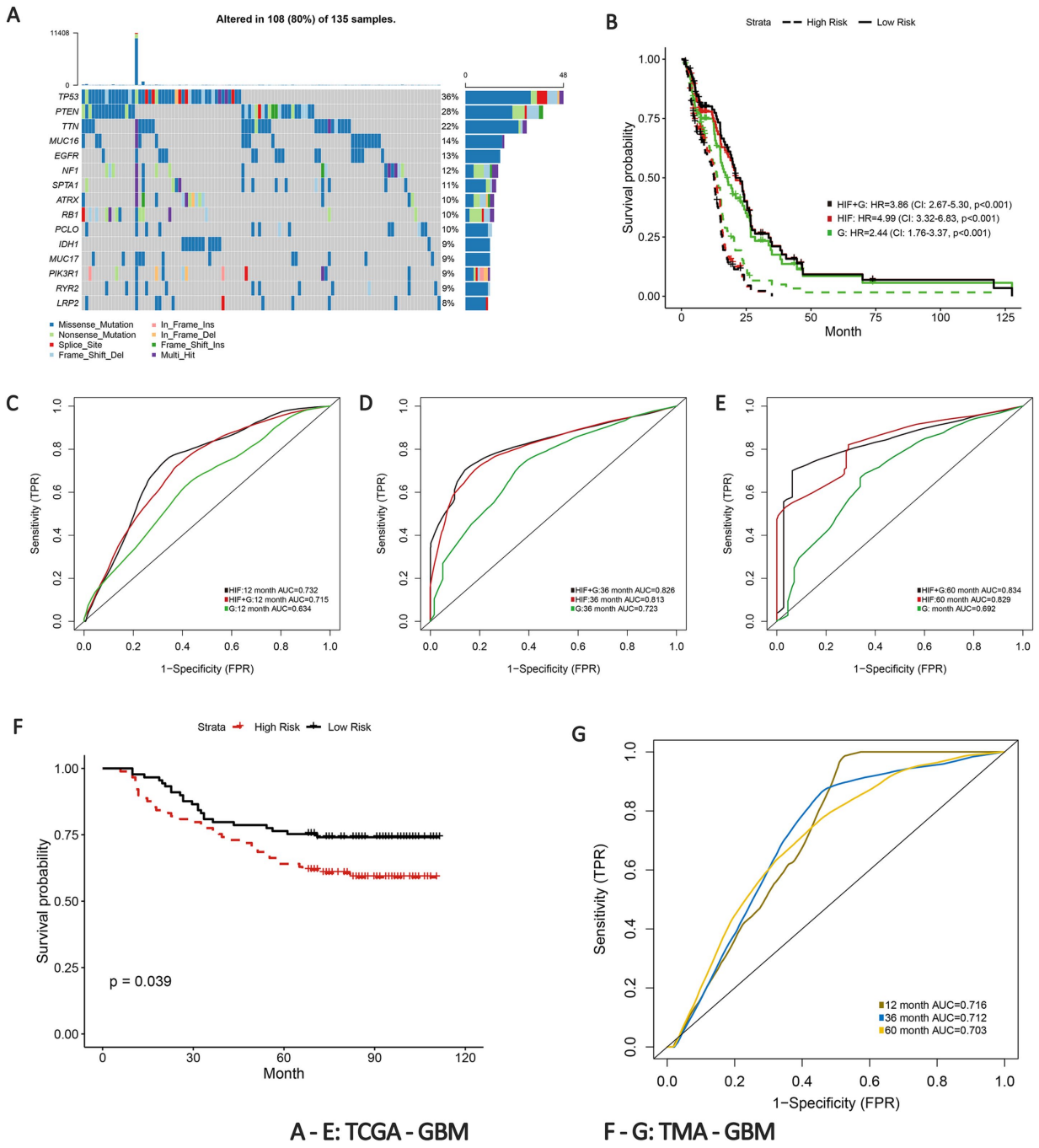


FIGURE 4
Prognostic models integrating HIF and genomics. **(A)** The waterfall plot of the top 15 most common somatic mutations in training set. **(B)** Kaplan–Meier curves of histopathological image features model (HIF), genomics model (G) and integrative histopathology + genomics model (HIF + G) in the validation set. **(C–E)** The **(C)** 1-year, **(D)** 3-year, and **(E)** 5-year area under the time-dependent receiver operating curve (AUC) of the three prognostic models in the validation set. **(F)** Kaplan–Meier curves of high-risk group and low-risk group in the TMA external validation cohort. **(G)** Time-dependent ROC of 1-year, 3-year, and 5-year OS in the TMA external validation cohort.

with tumor growth and invasion inhibition in GBM (46), which may suggest a new effective target for anti-cancer treatment strategies.

As demonstrated in the validation set, the transcriptomics model (RNA) displayed a good predictive performance for OS (1-year AUC = 0.751, 3-year AUC = 0.795 and 5-year AUC = 0.809), which were about equal to the HIF model (1-year AUC = 0.722, 3-year AUC = 0.815 and 5-year AUC = 0.835). Furthermore, we incorporated the transcriptomics and image features as the integrated model (HIF + RNA), which achieved the highest accuracy with the 1-year, 3-year and 5-year AUC increased to 0.769, 0.831 and 0.848 (Figures 5C–E). Additionally, Kaplan–Meier survival analyses also revealed significant differences in survival outcomes between the two groups, with the integrative HIF + RNA model presenting the most notable prognostic value (HR = 7.15, 95%CI: 4.51–10.41, $p < 0.001$, Figure 5B).

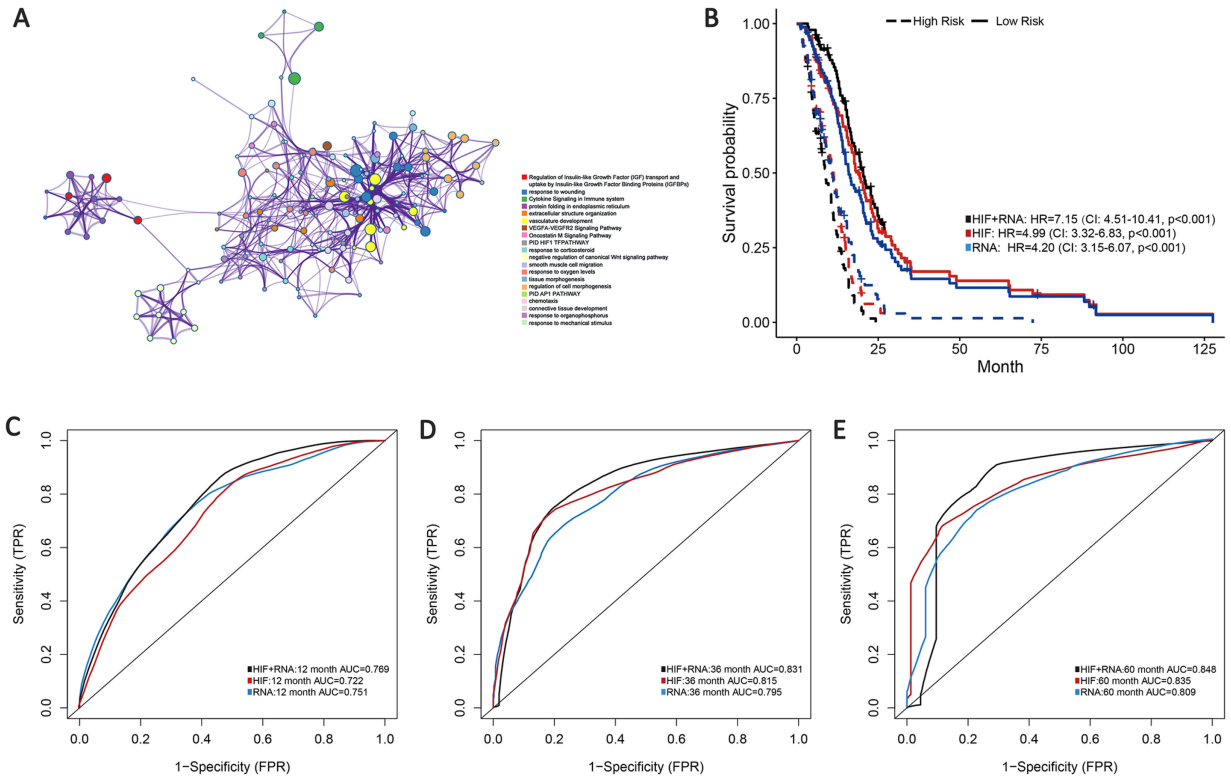


FIGURE 5
Prognostic models integrating HIF and transcriptomics (RNA). **(A)** Metascape enrichment network visualization cluster of genes and associated biological pathways based on training set. Each circled node represents a term and each color represents its cluster identification, showing the intra-cluster and inter-cluster similarities of enriched terms. **(B)** Kaplan–Meier curves of prognostic models (HIF, RNA, and HIF + RNA) in the validation set. **(C–E)**. The **(C)** 1-year, **(D)** 3-year and **(E)** 5-year AUCs of the three prognostic models in the validation set.

3.5 Integrated prognostic model of HIF and proteomics

To improve the prognostic prediction of GBM we also incorporated proteomics profile from TCPA portal for further analysis through the reverse phase protein array (RPPA), a high-throughput proteomics method that can assess protein expression and activation states in abundant samples using small amounts of material. In total we involved 179 eligible protein profiles in the proteomics model based on the validation set. The integration of image features and proteomics features (HIF + P) achieved the highest AUCs in 1-year, 3-year and 5-year compared with the proteomics model (0.752 vs. 0.743, 0.835 vs. 0.813, 0.854 vs. 0.818) or the HIF model alone (Figures 6A–C). As shown in the survival analyses, patients in the high-risk group were significantly related to poor OS, and the integrated model (HIF + P) attained the best performance in prognosis prediction among the three models (HR = 6.35, 95%CI: 4.05–9.20, $p < 0.001$, Figure 6D).

3.6 Integrated multi-omics features for survival prediction

According to the previous analyses, the histopathological image features have presented certain effectiveness in prognostic prediction for GBM patients, and histopathology + omics models have also indicated enhancement in predictive performance and accuracy than the single-omics models. Therefore, we expect to explore the prognostic capacity of a multi-omics predictive model incorporating all the omics features (HIF, genomics, transcriptomics, and proteomics). Based on the validation set, the multi-omics model achieved a 1-year AUC of 0.820, 3-year AUC of 0.926 and 5-year AUC of 0.878, representing an improvement over the HIF + genomics, HIF + transcriptomics and HIF + proteomics models (Figure 7A). Kaplan–Meier survival analysis illustrated a significant difference in survival between high-risk and low-risk groups (HR = 13.14, 95% CI: 7.95–25.95, $p < 0.001$, Figure 7B). Furthermore, the multi-omics model demonstrated superior net benefit in survival prediction compared to the other models (Figure 7C).

In order to identify the histopathological image features with higher prognostic value for OS, LASSO-Cox regression and SVM-RFE were performed independently. These combined approaches help mitigate the risk of overfitting and ensure the robustness of selected features across different selection frameworks. Previous studies (39–41) have demonstrated that the combination of LASSO and SVM-RFE enhances the reliability of prognostic feature identification in cancer research. A total of five imaging features involved in prognosis were selected via LASSO-Cox regression model, and SVM-RFE selected 12 imaging features with the most significant predictive ability. Ultimately, three overlapped features were identified as prognostic histopathological image features (PHIF), including StDev_Cells_AreaShape_FormFactor, StDev_Cells_AreaShape_Orientation and Mean_Cells_Texture_InfoMeas1_MaskedHematoxylin_3_90

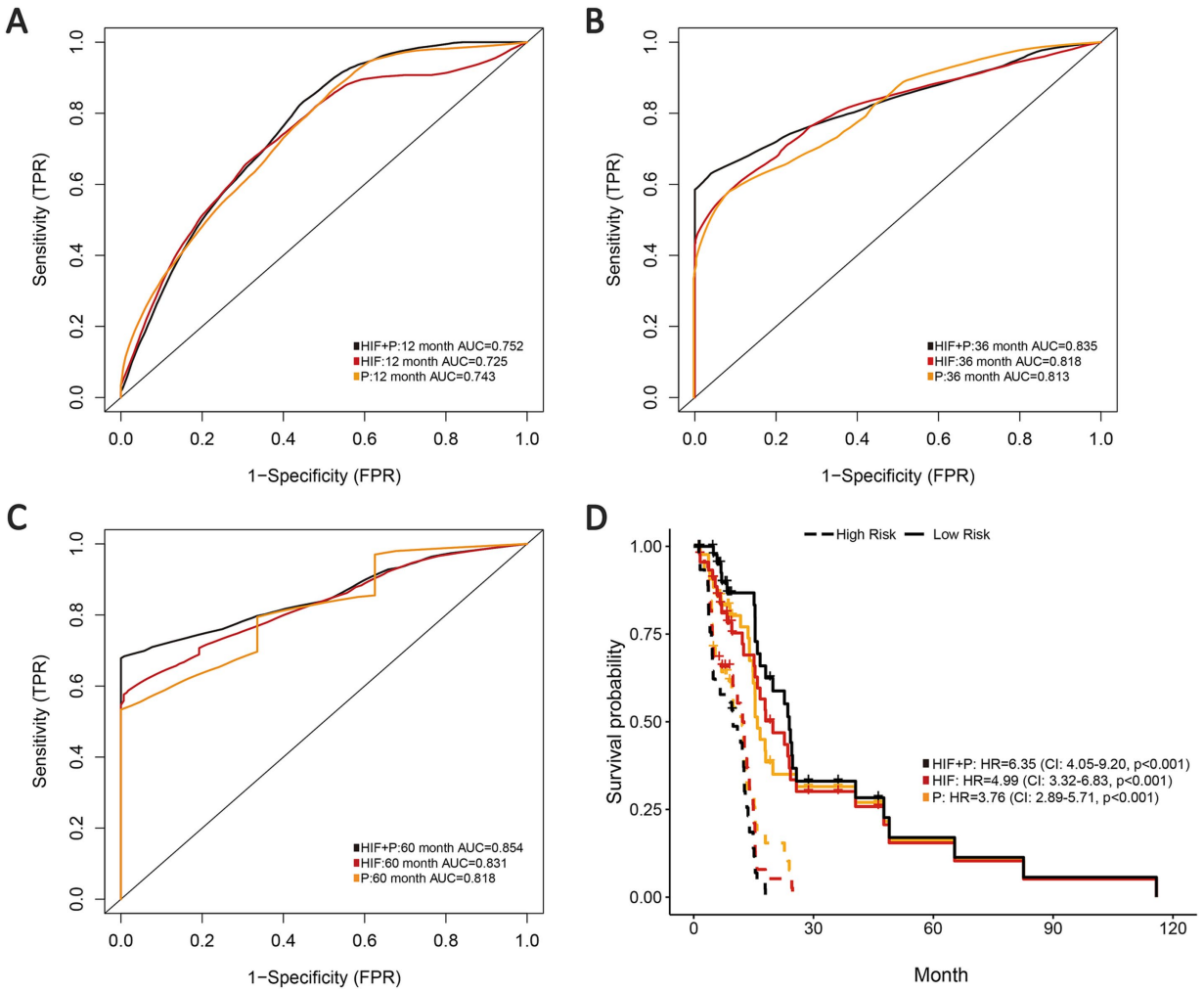


FIGURE 6
Prognostic models integrating HIF with proteomics (P). **(A–C)** The **(A)** 1-year, **(B)** 3-year and **(C)** 5-year AUCs of the three prognostic models (HIF, P and HIF + P) in the validation set. **(D)** Kaplan–Meier curves of the three prognostic models in the validation set.

(Figures 8A,B). Representative sub-images and detailed information of patients with high expressed and low expressed PHIF were displayed in Figure 8C and Supplementary material 4.

To explore the upstream genetic mechanisms, we employed WGCNA to construct a gene co-expression network in the training set and identify the gene clusters significantly correlated with the PHIF in GBM samples. Module-trait correlation analysis showed that the red module (219 genes) and turquoise module (868 genes) were significantly associated with the three prognostic image features of GBM among the six identified gene co-expression modules (Figure 9A). Therefore, we defined the red and turquoise module as the key modules of significant prognostic relevance for subsequent research.

Subsequently, we performed an enrichment analysis to explain the biological interpretations of the gene expression profile in the two modules. Genes in the red module were significantly related to several biological processes and pathways such as defense response to other organism, myeloid leukocyte activation, leukocyte cell–cell adhesion, activation of immune response and response to bacterium (Figure 9B). The results indicated that these genes may be involved in immune function, a crucial aspect of tumor immunology, which plays an important role in tumor initiation and progression. The genes in the turquoise module were primarily enriched in categories related to cell morphogenesis involved in differentiation, regulation of neuron differentiation and nervous system development, synapse organization and signaling (Figure 9C). These findings implied that turquoise module genes may have potential association with central nervous system pathways and cerebral function, which may correspond to tumorigenesis and progression in GBM.

4 Discussion

In this study, we extracted quantitative image features from histopathological images of GBM patients, and subsequently constructed machine learning classifiers based on the HIFs to discriminate the common molecular features of GBM. A predictive model incorporating HIFs was established in the training set, with its prognostic validity subsequently verified in both internal and external validation cohorts. The results demonstrated the prognostic

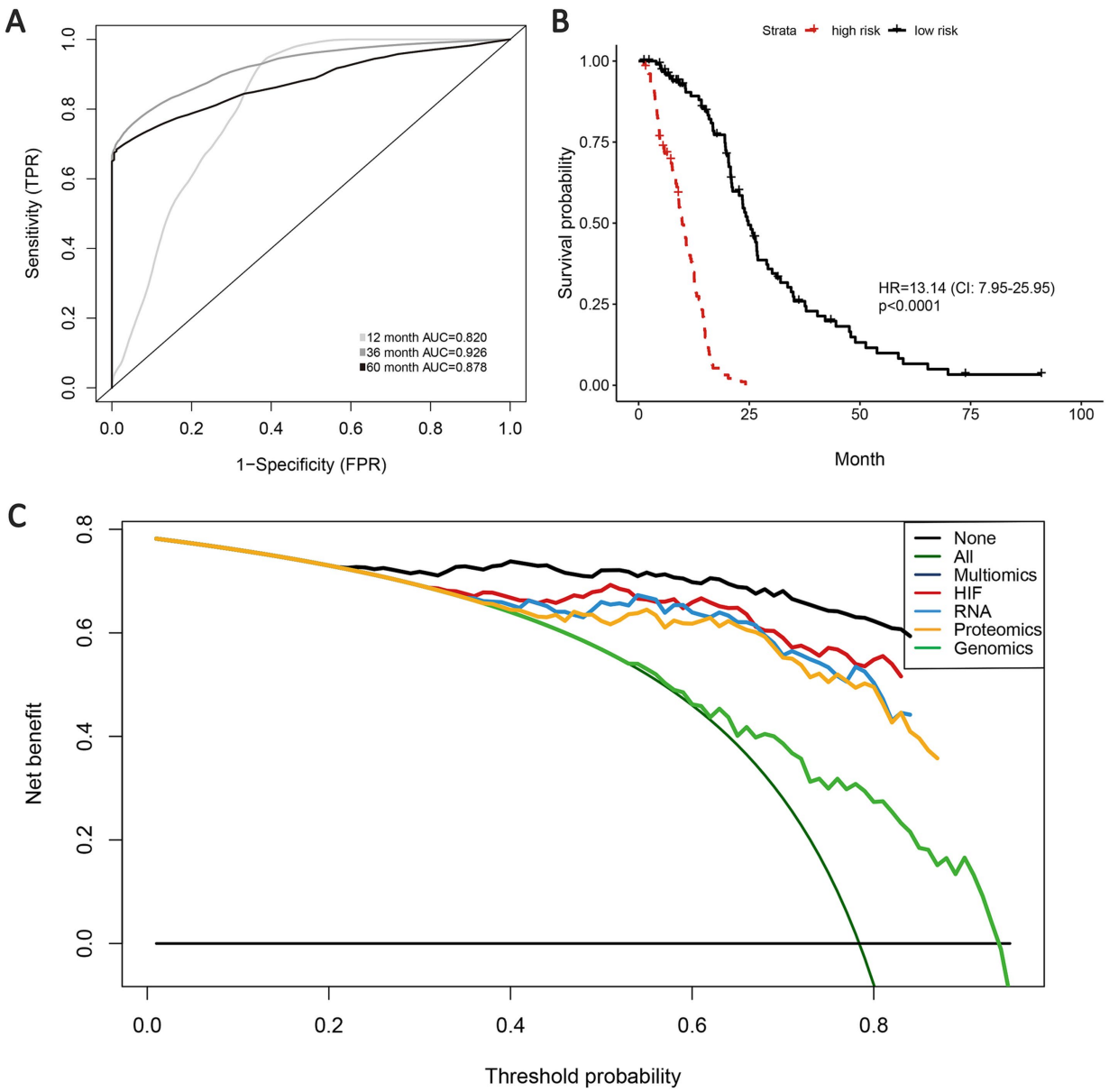


FIGURE 7
Prognostic models of survival integrating HIF and multiple omics features. **(A)** AUCs of multi-omics model in the validation set. Kaplan–Meier curve of multi-omics model (integrating HIF, radiomics, genomics, transcriptomics, proteomics) in the validation set. **(B)** Decision curves analysis for different models in the validation set. **(C)** The gray oblique line represented the net benefit of intervention for all patients, while the horizontal line represented the net benefit of no intervention. The multi-omics model achieved higher net benefit than single-omics models across the major range of threshold probability.

robustness of the predictive model. To enhance the predictive performance, comprehensive prognostic models were built by integrating HIFs with multi-omics data. Based on machine learning approaches, we selected prognostic histopathological image features (PHIF) and identified gene modules most strongly correlated with PHIF through bioinformatics techniques. Notably, the predictive power of OS in patients was significantly enhanced in multi-omics models compared with the single-omics models, suggesting that this approach may be promising for risk stratification and individualized treatments for GBM patients.

Based on histopathological image features, we performed the prediction of the common somatic mutations (ATRX, IDH, and TERT) and methylation (MGMT) in GBM through combinations of eight independent machine learning algorithms. IDH mutations, which occur in approximately 12% of GBM cases, are a well-established prognostic marker associated with prolonged OS (47). The mutation can induce downstream effects on cellular metabolism and epigenetic regulation (48). Previous studies have reported the predictive value of MRI radiomics models for identifying IDH1 mutations in GBM (49, 50), as well as the characterization of core signaling pathways in IDH wild-type tumors (51). The prediction ability of histopathological image features in IDH mutation has not been widely explored, while it may represent an important avenue for further research in

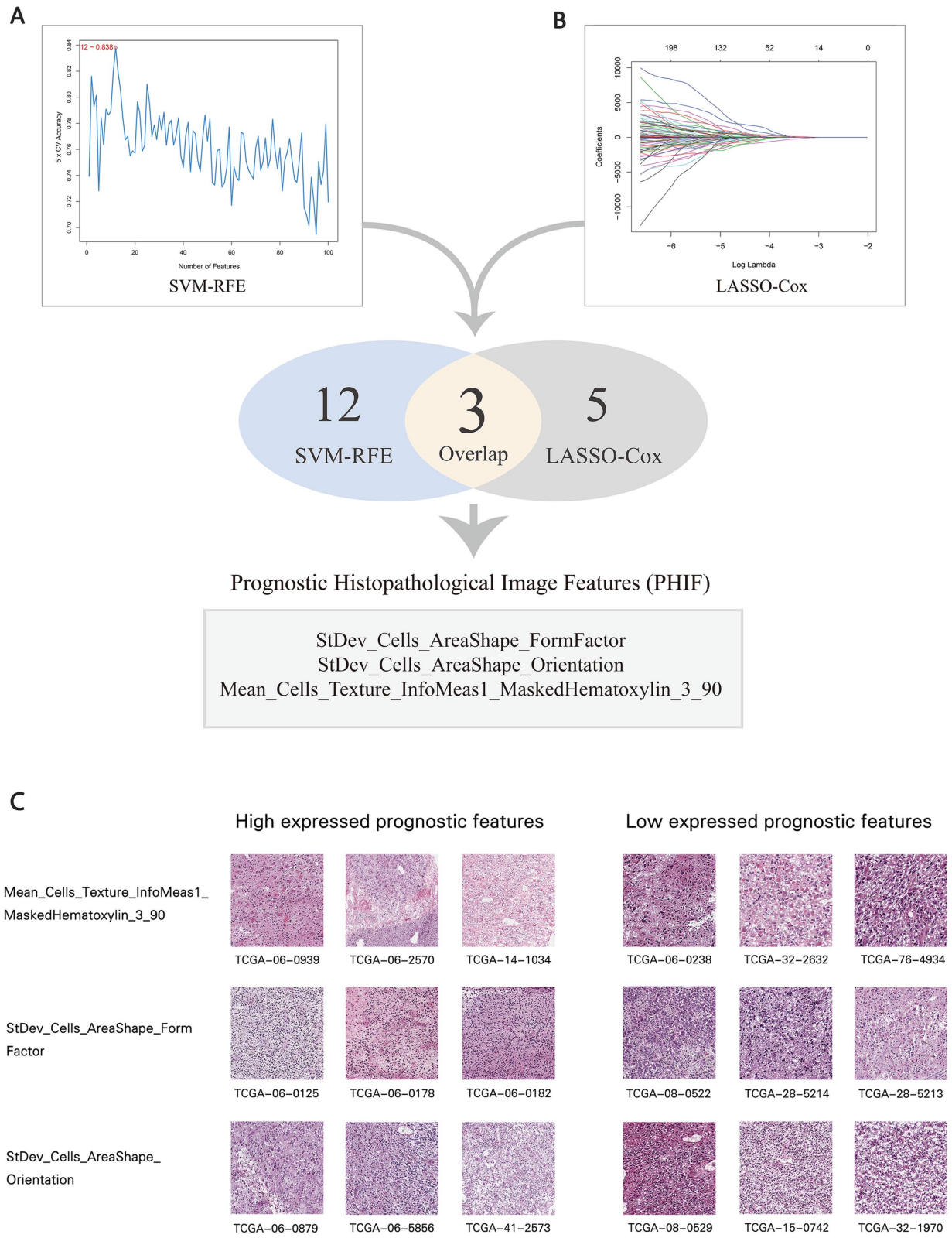


FIGURE 8
Selection of prognostic histopathological image features (PHIF). **(A)** Twelve image features were selected by SVM-RFE. **(B)** Five image features were selected by LASSO-COX regression model. Three image features within the overlap were defined as PHIF. Three image features within the overlap were defined as PHIF. **(C)** Representative sub-images of patients with high expressed and low expressed PHIF. The groups were defined by the median value of each image feature.

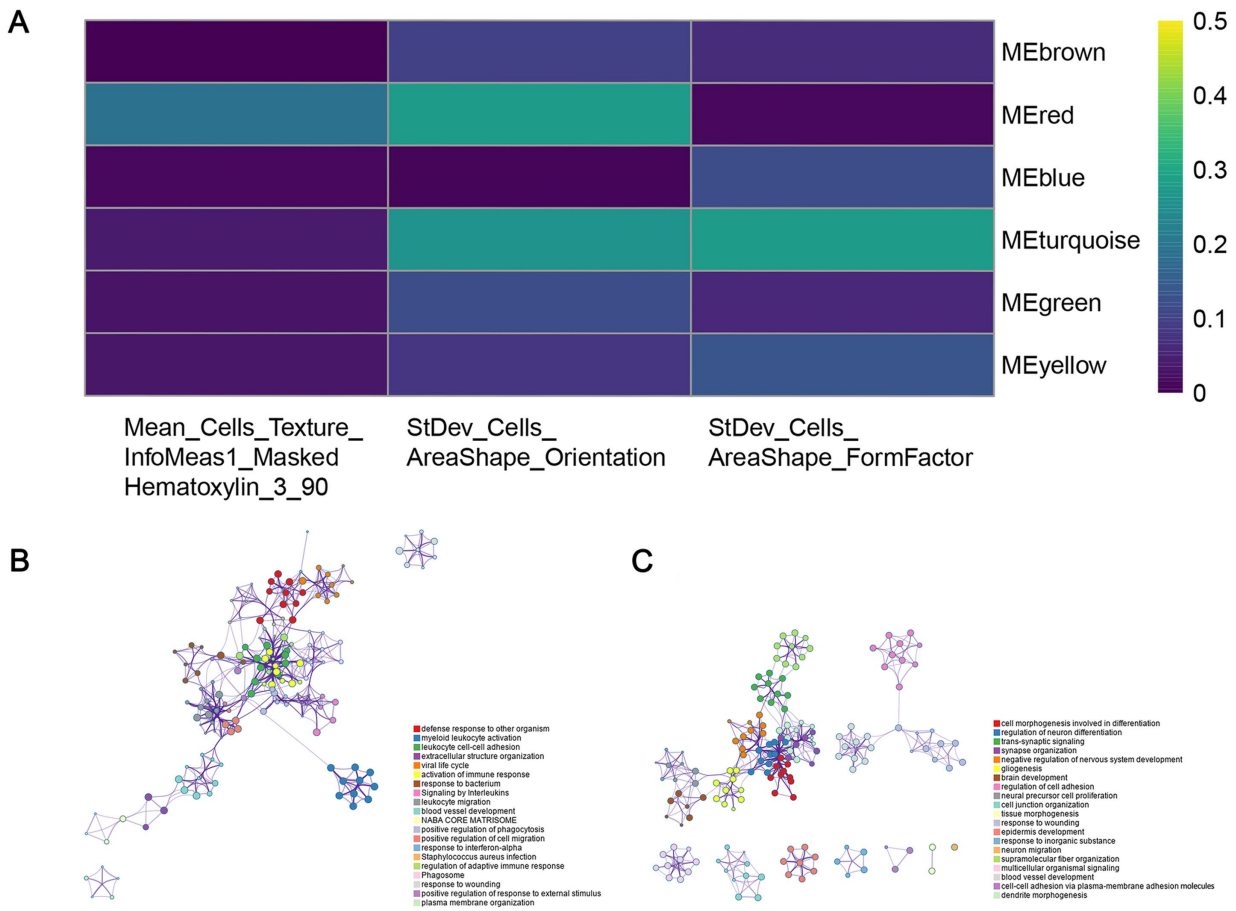


FIGURE 9
Identification of co-expressed gene modules. **(A)** Heatmap of the relationship between gene modules and prognostic histopathological image features (PHIF) through WGCNA. The red module and turquoise module showed the most significant correlation. **(B)** Metascape enrichment network of genes in the red module. Each circled node represents a term and each color represents its cluster identification, showing the intra-cluster and inter-cluster similarities of enriched terms. **(C)** Metascape enrichment network of genes in the turquoise module.

prognostic evaluation and targeted therapies for GBM. MGMT methylation status and TERT promoter mutations have also been recognized as powerful diagnostic and prognostic indicators in GBM (2, 52). Meanwhile, we also conducted the prediction of four mRNA-based molecular subtypes (classical, mesenchymal, neural, proneural) and the G-CIMP methylator phenotype. The prognostic significance of G-CIMP+ subsets among glioma types has been investigated in previous studies (53, 54). For instance, 1p/19q codeletion and MGMT promoter methylation may act as therapeutic predictive markers in GBM (55). Our random forest predictive model based on HIFs exhibited certain accuracy and effectiveness in predicting GBM molecular characteristics, which may contribute to improving current clinical examinations and diagnostic practices.

Subsequently, we constructed prognostic models through random forest algorithm based on single-omics and integrated multi-omics data. Image features of histopathology tissue slides can infer morphological changes in tumor cells and microenvironment, which have proven valuable in identifying pathology biomarkers and predicting clinical outcomes through machine learning techniques (56–58). A fair number of computational histopathologic models have also been applied in the prognostic prediction of diseases such as breast (59), lung (60) and colorectal cancers (61). Consistent with previous studies, the image features with significant prognostic power of OS we selected primarily pertained to Zernike and cell texture (i.e., contrast, sum average, and difference entropy). Zernike shape features in nuclei and cytoplasm are extracted frequently to identify long and short term survival (62). In addition, the texture features are frequently used to represent the distribution and variation of pixel intensity, as well as the relationship between pairs with different intensity values in the regions of interest. While many studies have established prognostic modules based on single-omics data source or combination of quantitative histopathological image features and genomics features (21, 53), our study focused on a more comprehensive evaluation of image features to provide additional prognostic efficiency and precision of the prognostic model. By integrating HIFs with genomics, transcriptomics and proteomics data, we developed a multi-omics model incorporating all features, which eventually achieved superior prediction performance compared to other models. Additionally, we further proposed external validation by involving an extra TMA cohort, further supporting the robustness and generalizability of our findings.

An intriguing observation in our study was that the model based solely on HIFs slightly outperformed the combined HIF and genomics (HIF + G) model in terms of predictive performance, as shown in Figure 4B. This unexpected finding prompted further reflection on the interaction between histopathological and genomic data in prognostic modeling. One possible explanation lies in feature redundancy and confounding effects that HIFs inherently capture tumor morphological and microstructural features, which may already correlate with patient prognosis. The addition of genomic features that provide overlapping or weakly correlated prognostic signals may introduce noise rather than improving predictive accuracy. This aligns with established principles in machine learning, where the mere inclusion of additional variables does not necessarily enhance model performance; instead, feature interactions must be carefully managed to avoid confounding effects. Moreover, the non-linearity between histopathological and genomic data may contribute to this outcome. While HIFs reflect macroscopic tumor morphology, genomic alterations influence prognosis through intricate molecular pathways that may not exhibit direct correlations with image-derived features. Traditional machine learning models may struggle to capture these complex interactions effectively, highlighting the need for alternative fusion strategies such as deep learning or graph neural networks to better integrate data from different modalities.

Despite the robust predictive power of HIFs alone, we emphasize the importance of multi-omics integration for comprehensive patient profiling. While the HIF + G model did not significantly outperform the HIF model alone, the incorporation of transcriptomic and proteomic data substantially improved the accuracy of our prognostic models. This suggests that multi-omics integration holds promise for enhancing model generalizability and robustness across diverse patient populations. Further optimization of feature selection and model refinement will be necessary to fully leverage the potential of multi-omics data.

Through SVM-RFE and LASSO-Cox regression machine learning algorithms, we identified three prognostic histopathological image features (PHIF) concerning cell morphology and texture. We also explored the upstream molecular mechanisms of these features by identifying relevant gene co-expression modules via weighted gene co-expression network analysis (WGCNA). Enrichment analysis of the red and turquoise gene modules demonstrated significant prognostic association with molecular pathways mainly involved in immune response, cell morphogenesis involved in differentiation, development and regulation of central nervous system function. For instance, leukocyte cell adhesion plays a crucial role in the progression and resolution of innate immunity (63). Myeloid leucocyte activation reveals exposure to activating factors and has been regarded as one of the major forces in immunosuppression in tumor progression (64). The genes enriched in cell morphogenesis related pathways might suggest the association with tumor angiogenesis and cell adhesion. In addition, regulation of neuron differentiation, trans-synaptic signaling and gliogenesis also suggest a close connection with biological processes in GBM development (65–67). The results may offer an opportunity to comprehend the association of histopathological image features and the upstream mechanisms of the oncogenesis and progression of GBM.

In conclusion, this study demonstrated the potential of histopathological image features in predicting molecular characteristics and classifying molecular subtypes. By integrating histopathological image features with multi-omics data, we developed comprehensive prognostic models and subsequently analyzed the associated upstream biological processes. The integrative multi-omics model has the potential to enhance prediction performance for OS with greater accuracy and robustness, thereby contributing to risk stratification, prognostic evaluation, and personalized treatment strategies for GBM patients.

However, several limitations should be addressed. Firstly, while the prognostic models were validated using an external TMA cohort to assess prediction stability, a larger-scale multi-center dataset is needed to enhance the applicability and reliability of our findings. Secondly, the genomic features of patients with intermediate survival (12–60 months) warrant further investigation, as they may provide additional insights into treatment response and prognostic markers. Additionally, discrepancies and potential biases in multi-omics data could impact the results. Future research should explore alternative data integration strategies to optimize the synergy between histopathology and molecular alterations. We also acknowledge the lack of unified visualization for all survival curves and model comparisons. Although constrained by computational limitations, we recognize the value of such visual summaries and are committed to improving model visualization and interpretability in future work, hoping to provide clearer insights for both clinical and research applications. Lastly, further clinical and experimental research is required to elucidate the molecular mechanisms underlying the relationship between histopathological image features and survival outcomes in GBM patients.

Data availability statement

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

Author contributions

YH: Conceptualization, Methodology, Writing – original draft. LC: Data curation, Methodology, Software, Writing – review & editing. ZZ: Validation, Writing – review & editing. YuL: Writing – review & editing, Supervision. LH: Supervision, Writing – review & editing. YaL: Supervision, Writing – review & editing. PL: Data curation, Writing – review & editing. FS: Data curation, Writing – review & editing. ZL: Conceptualization, Supervision, Writing – review & editing. ZZ: Conceptualization, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by Sichuan Science and Technology Program (2024YFFK0054 and 2024YFFK0063), Project of Chengdu Medical Research (No. 2023009)

and The Key Projects of Science and Technology Department of Sichuan Province (2024YFFK0063).

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.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2025.1510793/full#supplementary-material

References

1. Ye F, Gao Q, Cai MJ. Therapeutic targeting of EGFR in malignant gliomas. *Expert Opin Ther Targets*. (2010) 14:303–16. doi: 10.1517/14728221003598948

2. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol*. (2016) 131:803–20. doi: 10.1007/s00401-016-1545-1

3. Ostrom Q, Cioffi G, Gittleman H, Patil N, Waite K, Kruchko C, et al. CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2012–2016. *Neuro-Oncol*. (2019) 21:v1–v100. doi: 10.1093/neuonc/noz150

4. Ostrom QT, Price M, Neff C, Cioffi G, Waite KA, Kruchko C, et al. CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2015–2019. *Neuro-Oncol*. (2022) 24:v1–v95. doi: 10.1093/neuonc/noac202

5. Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJB, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med*. (2005) 352:987–96. doi: 10.1056/NEJMoa043330

6. Ostrom Q, Bauchet L, Davis FG, Deltour I, Fisher JL, Langer CE, et al. The epidemiology of glioma in adults: a "state of the science" review. *Neuro-Oncol*. (2014) 16:896–913. doi: 10.1093/neuonc/nou087

7. Waite KA, Cioffi G, Kruchko C, Patil N, Brat DJ, Bruner JM, et al. Aligning the Central Brain Tumor Registry of the United States (CBTRUS) histology groupings with current definitions. *Neurooncol Pract*. (2022) 9:317–27. doi: 10.1093/nop/npac025

8. Poon MTC, Sudlow CLM, Figueroa JD, Brennan PM. Longer-term (≥ 2 years) survival in patients with glioblastoma in population-based studies pre- and post-2005: a systematic review and meta-analysis. *Sci Rep*. (2020) 10:11622. doi: 10.1038/s41598-020-68011-4

9. Stupp R, Taillibert S, Kanner A, Read W, Steinberg DM, Lhermitte B, et al. Effect of tumor-treating fields plus maintenance temozolomide vs. maintenance temozolomide alone on survival in patients with glioblastoma: a randomized clinical trial. *JAMA*. (2017) 318:2306–16. doi: 10.1001/jama.2017.18718

10. Kebir S, Hattingen E, Niessen M, Rauschenbach L, Fimmers R, Hummel T, et al. Olfactory function as an independent prognostic factor in glioblastoma. *Neurology*. (2020) 94:e529–37. doi: 10.1212/WNL.0000000000008744

11. Navarria P, Pessina F, Clerici E, Rossini Z, Franceschini D, D'Agostino G, et al. Is IDH status the only factor predicting prognosis in newly diagnosed anaplastic glioma patients? Outcome evaluation and prognostic factor analysis in a single-institution large series. *J Neurosurg*. (2020) 135:64–77. doi: 10.3171/2020.5.JNS201116

12. Piepoli T, Jakupoglu C, Gu W, Lualdi E, Suarez-Merino B, Poliani PL, et al. Expression studies in gliomas and glial cells do not support a tumor suppressor role for LGI1. *Neuro-Oncol*. (2006) 8:96–108. doi: 10.1215/15228517-2005-006

13. Richardson TE, Snuderl M, Serrano J, Karajannis MA, Heguy A, Oliver D, et al. Rapid progression to glioblastoma in a subset of IDH-mutated astrocytomas: a genome-wide analysis. *J Neuro-Oncol*. (2017) 133:183–92. doi: 10.1007/s11060-017-2431-y

14. Huang Q, Li F, Chen Y, Hong F, Wang H, Chen J. Prognostic factors and clinical outcomes in adult primary gliosarcoma patients: a surveillance, epidemiology, and end results (SEER) analysis from 2004 to 2015. *Br J Neurosurg*. (2020) 34:161–7. doi: 10.1080/02688697.2019.1699903

15. Wick W, Gorlia T, Bady P, Platten M, van den Bent MJ, Taphoorn MJB, et al. Phase II study of radiotherapy and temsirolimus versus radiochemotherapy with temozolomide in patients with newly diagnosed glioblastoma without MGMT promoter hypermethylation (EORTC 26082). *Clin Cancer Res*. (2016) 22:4797–806. doi: 10.1158/1078-0432.CCR-15-3153

16. Wick W, Weller M, van den Bent M, Sanson M, Weiler M, von Deimling A, et al. MGMT testing—the challenges for biomarker-based glioma treatment. *Nat Rev Neurol*. (2014) 10:372–85. doi: 10.1038/nrneurol.2014.100

17. Kikuchi Z, Shibahara I, Yamaki T, Yoshioka E, Shofuda T, Ohe R, et al. TERT promoter mutation associated with multifocal phenotype and poor prognosis in patients with wild-type glioblastoma. *Neurooncol Adv*. (2020) 2:vdaa114. doi: 10.1093/noajnl/vdaa114

18. Pekmezci M, Rice T, Molinaro AM, Walsh KM, Decker PA, Hansen H, et al. Adult infiltrating gliomas with WHO 2016 integrated diagnosis: additional prognostic roles of ATRX and TERT. *Acta Neuropathol*. (2017) 133:1001–16. doi: 10.1007/s00401-017-1690-1

19. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*. (2010) 17:98–110. doi: 10.1016/j.ccr.2009.12.020

20. Weller M, Pfister SM, Wick W, Hegi ME, Reifenberger G, Stupp R. Molecular neuro-oncology in clinical practice: a new horizon. *Lancet Oncol*. (2013) 14:e370–9. doi: 10.1016/S1470-2045(13)70168-2

21. Maier A, Syben C, Lasser T, Riess C. A gentle introduction to deep learning in medical image processing. *Z Med Phys*. (2019) 29:86–101. doi: 10.1016/j.zemedi.2018.12.003

22. Das A, Nair M, Peter SD. Computer-aided histopathological image analysis techniques for automated nuclear atypia scoring of breast cancer: a review. *J Digit Imaging*. (2020) 33:1091–121. doi: 10.1007/s10278-019-00295-z

23. Ikehara H, Saito Y, Matsuda T, Uraoka T, Murakami Y. Diagnosis of depth of invasion for early colorectal cancer using magnifying colonoscopy. *J Gastroenterol Hepatol*. (2010) 25:905–12. doi: 10.1111/j.1440-1746.2010.06275.x

24. Yu KH, Zhang C, Berry GJ, Altman RB, Ré C, Rubin DL, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat Commun*. (2016) 7:12474. doi: 10.1038/ncomms12474

25. Cheng J, Zhang J, Han Y, Wang X, Ye X, Meng Y, et al. Integrative analysis of histopathological images and genomic data predicts clear cell renal cell carcinoma prognosis. *Cancer Res*. (2017) 77:e91–e100. doi: 10.1158/0008-5472.CAN-17-0313

26. Yu KH, Wang F, Berry GJ, Ré C, Altman RB, Snyder M, et al. Classifying non-small cell lung cancer types and transcriptomic subtypes using convolutional neural networks. *J Am Med Inform Assoc*. (2020) 27:757–69. doi: 10.1093/jamia/ocz230

27. Huang C, Cintra M, Brennan K, Zhou M, Colevas AD, Fischbein N, et al. Development and validation of radiomic signatures of head and neck squamous cell carcinoma molecular features and subtypes. *EBioMedicine*. (2019) 45:70–80. doi: 10.1016/j.ebiom.2019.06.034

28. Goode A, Gilbert B, Harkes J, Jukic D, Satyanarayanan M. OpenSlide: a vendor-neutral software foundation for digital pathology. *J Pathol Inform*. (2013) 4:27. doi: 10.4103/2153-3539.119005

29. Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol*. (2006) 7:R100. doi: 10.1186/gb-2006-7-10-r100

30. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med*. (1997) 16:385–95. doi: 10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3

31. Breiman L. Random forests. *Mach Learn*. (2001) 45:5–32. doi: 10.1023/A:1010933404324

32. Friedman JH. Greedy function approximation: a gradient boosting machine. *J Ann Stat*. (2001) 29. doi: 10.1214/aos/1013203451

33. Chen T, Guestrin C (2016). XGBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 785–794.

34. Collins M, Schapire RE, Singer Y. Logistic regression, AdaBoost and Bregman distances. *Mach Learn*. (2002) 48:253–85. doi: 10.1023/A:1013912006537

35. Safavian SR, Landgrebe D. A survey of decision tree classifier methodology. *IEEE Trans Syst Man Cybern*. (1991) 21:660–74. doi: 10.1109/21.97458

36. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. (1995) 20:273–97. doi: 10.1007/BF00994018

37. Friedman N, Geiger D, Goldszmidt M. Bayesian network classifiers. *Mach Learn*. (1997) 29:131–63. doi: 10.1023/A:1007465528199

38. Keller J, Gray M, Givens J. A fuzzy K-nearest neighbor algorithm. *IEEE Trans Syst Man Cybern*. (1985) SMC-15:580–5. doi: 10.1109/TSMC.1985.6313426

39. Lin S, Wei C, Wei Y, Fan J. Construction and verification of an endoplasmic reticulum stress-related prognostic model for endometrial cancer based on WGCNA and machine learning algorithms. *Front Oncol*. (2024) 14:1362891. doi: 10.3389/fonc.2024.1362891

40. Shang JR, Zhu J, Bai L, Kulabiek D, Zhai XX, Zheng X, et al. Adipocytes impact on gastric cancer progression: prognostic insights and molecular features. *World J Gastrointest Oncol*. (2024) 16:3011–31. doi: 10.4251/wjgo.v16.i7.3011

41. Xu Y, Wang X, Yuan W, Zhang L, Chen W, Hu K. Identification of BANF1 as a novel prognostic biomarker in gastric cancer and validation via *in-vitro* and *in-vivo* experiments. *Aging*. (2024) 16:1808–28. doi: 10.18632/aging.205461

42. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. (2008) 9:559. doi: 10.1186/1471-2105-9-559

43. Kamarudin A, Cox T, Kolamunnage-Dona R. Time-dependent ROC curve analysis in medical research: current methods and applications. *BMC Med Res Methodol*. (2017) 17:53. doi: 10.1186/s12874-017-0332-6

44. Guo T, Feng Y, Liu Q, Yang X, Jiang T, Chen Y, et al. MicroRNA-320a suppresses in GBM patients and modulates glioma cell functions by targeting IGF-1R. *Tumour Biol*. (2014) 35:11269–75. doi: 10.1007/s13277-014-2283-4

45. Ho K, Chen PH, Shih CM, Lee YT, Cheng CH, Liu AJ, et al. miR-4286 is involved in connections between IGF-1 and TGF-β Signaling for the mesenchymal transition and invasion by glioblastomas. *Cell Mol Neurobiol*. (2020) 42:791–806. doi: 10.1007/s10571-020-00977-1

46. Bach LJ, Duan C. IGF-binding proteins. *J Mol Endocrinol*. (2018) 61:T11–28. doi: 10.1530/JME-17-0254

47. Parsons DW, Jones Ŝ, Zhang X, Lin JCH, Leary RJ, Angenendt P, et al. An integrated genomic analysis of human glioblastoma multiforme. *Science*. (2008) 321:1807–12. doi: 10.1126/science.1164382

48. Waitkus M, Diplas BH, Yan H. Isocitrate dehydrogenase mutations in gliomas. *Neuro-Oncol*. (2016) 18:16–26. doi: 10.1093/neuonc/nov136

49. Chang K, Bai HX, Zhou H, Su C, Bi WL, Agbodza E, et al. Residual convolutional neural network for the determination of IDH status in low- and high-grade gliomas from MR imaging. *Clin Cancer Res*. (2018) 24:1073–81. doi: 10.1158/1078-0432.CCR-17-2236

50. Chang P, Grinband J, Weinberg BD, Bardis M, Khy M, Cadena G, et al. Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas. *AJNR Am J Neuroradiol*. (2018) 39:1201–7. doi: 10.3174/ajnr.A5667

51. Park JE, Kim HS, Park SY, Nam SJ, Chun SM, Jo Y, et al. Prediction of core signaling pathway by using diffusion- and perfusion-based MRI radiomics and next-generation sequencing in isocitrate dehydrogenase wild-type glioblastoma. *Radiology*. (2020) 294:388–97. doi: 10.1148/radiol.2019190913

52. Lee SY. Temozolomide resistance in glioblastoma multiforme. *Genes Dis*. (2016) 3:198–210. doi: 10.1016/j.gendis.2016.04.007

53. Mur P, Rodríguez de Lope Á, Díaz-Crespo FJ, Hernández-Iglesias T, Ribalta T, Fiaño C, et al. Impact on prognosis of the regional distribution of MGMT methylation with respect to the CpG island methylator phenotype and age in glioma patients. *J Neuro-Oncol*. (2015) 122:441–50. doi: 10.1007/s11060-015-1738-9

54. Noushmehr H, Weisenberger DJ, Diefes K, Phillips HS, Pujara K, Berman BP, et al. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell*. (2010) 17:510–22. doi: 10.1016/j.ccr.2010.03.017

55. Ceccarelli M, Barthel FP, Malta TM, Sabedot TS, Salama SR, Murray BA, et al. Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell*. (2016) 164:550–63. doi: 10.1016/j.cell.2015.12.028

56. Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat Med*. (2019) 25:1054–6. doi: 10.1038/s41591-019-0462-y

57. 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

58. Cheng J, Mo X, Wang X, Parwani A, Feng Q, Huang K. Identification of topological features in renal tumor microenvironment associated with patient survival. *Bioinformatics*. (2018) 34:1024–30. doi: 10.1093/bioinformatics/btx723

59. Chen J, Li Y, Xu J, Gong L, Wang LW, Liu WL, et al. Computer-aided prognosis on breast cancer with hematoxylin and eosin histopathology images: a review. *Tumour Biol*. (2017) 39:101042831769455. doi: 10.1177/1010428317694550

60. Luo X, Zang X, Yang L, Huang J, Liang F, Rodriguez-Canales J, et al. Comprehensive computational pathological image analysis predicts lung cancer prognosis. *J Thorac Oncol*. (2017) 12:501–9. doi: 10.1016/j.jtho.2016.10.017

61. Caie PD, Zhou Y, Turnbull AK, Oniscu A, Harrison DJ. Novel histopathologic feature identified through image analysis augments stage II colorectal cancer clinical reporting. *Oncotarget*. (2016) 7:44381–94. doi: 10.18632/oncotarget.10053

62. Li S, Lee M-C, Pun C-M. Complex Zernike moments features for shape-based image retrieval. *IEEE Trans Syst Man Cybern*. (2009) 39:227–37. doi: 10.1109/TSMCA.2008.2007988

63. Woodside DG, Vanderslice P. Chapter 3—inflammation and regulation by integrin cell adhesion antagonists In: JK Actor and KC Smith, editors. Translational inflammation. London: Academic Press (2019). 43–68.

64. Dubinski D, Wölfer J, Hasselblatt M, Schneider-Hohendorf T, Bogdahn U, Stummer W, et al. $CD4^+$ T effector memory cell dysfunction is associated with the accumulation of granulocytic myeloid-derived suppressor cells in glioblastoma patients. *Neuro-Oncol*. (2016) 18:807–18. doi: 10.1093/neuonc/nov280

65. Dong Y, Han Q, Zou Y, Deng Z, Lu X, Wang X, et al. Long-term exposure to imatinib reduced cancer stem cell ability through induction of cell differentiation via activation of MAPK signaling in glioblastoma cells. *Mol Cell Biochem*. (2012) 370:89–102. doi: 10.1007/s11010-012-1401-0

66. Hulleman E, Helin K. Molecular mechanisms in gliomagenesis. *Adv Cancer Res*. (2005) 94:1–27. doi: 10.1016/S0065-230X(05)94001-3

67. Lu W, Xie H, Yuan C, Li JJ, Li ZY, Wu AH. Identification of potential biomarkers and candidate small molecule drugs in glioblastoma. *Cancer Cell Int*. (2020) 20:419. doi: 10.1186/s12935-020-01515-1

Glossary

AdaBoost - Adaptive boosting

AUC - Area under the curve

CI - Confidence interval

DCA - Decision curve analysis

DEG - Differently expressed gene

DT - Decision tree

GBDT - Gradient boosting decision tree

GBM - Glioblastoma

HIF - Histopathological image features

HR - Hazard ratio

KNN - K-nearest neighbor

LASSO - Least absolute shrinkage and selection operator

LR - Logistic regression

NB - Naive Bayesian

OS - Overall survival

PHIF - Prognostic histopathological image features

RF - Random forest

ROC - Receiver operating characteristic

SVM - Support vector machine

TMA - Tissue microarrays

WGCNA - Weighted gene co-expression network analysis

XGBoost - Extreme gradient boosting

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 06 June 2025
DOI 10.3389/fmed.2025.1572500

Check for updates

OPEN ACCESS

EDITED BY
HaiHui Huang,
Shaoguan University, China

REVIEWED BY
Xuejing Sun,
University of Pittsburgh, United States
Abdullah Al Marzan,
Toxicology Society of Bangladesh, Bangladesh

*CORRESPONDENCE
C. George Priya Doss
✉ georgepriyadoss@vit.ac.in

RECEIVED 07 February 2025
ACCEPTED 12 May 2025
PUBLISHED 06 June 2025

CITATION
Loganathan T and Doss CGP (2025) Multi-omics insights into biomarkers of breast cancer associated diabetes: a computational approach.
Front. Med. 12:1572500.
doi: 10.3389/fmed.2025.1572500

COPYRIGHT
© 2025 Loganathan and Doss. 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.

Multi-omics insights into biomarkers of breast cancer associated diabetes: a computational approach

Tamizhini Loganathan and C. George Priya Doss*

Laboratory of Integrative Genomics, Department of Integrative Biology, School of Bio Sciences and Technology, Vellore Institute of Technology (VIT), Vellore, Tamil Nadu, India

Introduction: Breast cancer (BC) and diabetes are multifaceted diseases with interconnected molecular mechanisms that are not yet fully elucidated. These diseases share common risk factors, biological pathways, and treatment outcomes.

Methods: This study utilizes an integrative computational approach to investigate the interplay between BC and diabetes in African American (AA) and European American (EA) cohorts. It employs transcriptomic and exomic analyses to identify shared pathways and potential therapeutic targets.

Results: The pooled cohort of differential expression analysis identified 2,815 genes differentially expressed in BC patients with diabetes compared to those without diabetes, including 1824 upregulated and 990 downregulated genes. We reanalyzed transcriptomic data by stratifying BC patients with and without diabetes into two cohorts, identifying 3,245 DEGs in AA and 3,208 DEGs in EA, with 786 genes commonly altered between both groups. Whole-exome sequencing (WES) of 23 BC patients with diabetes revealed 899 variants across 208 unique genes, predominantly missense mutations. Among these, nine key genes were prioritized, with *TNFRSF1B* (L264P) and *PDPN* (A105G) identified as the most deleterious variants. Functional enrichment analyses highlighted the significant involvement of pathways related to extracellular matrix organization, angiogenesis, immune regulation, and signaling processes critical to cancer progression and metabolic dysfunction. The TNF pathway emerged as a central link connecting chronic inflammation, insulin resistance, and tumor growth. TNF-mediated mechanisms, including NF-κB activation, oxidative stress, and epithelial-to-mesenchymal transition (EMT), were found to drive both diseases, promoting tumorigenesis, immune evasion, and metabolic dysregulation.

Conclusion: This study provides critical molecular insights into the shared mechanisms of BC and diabetes, identifying the TNF pathway as a key therapeutic target to improve outcomes for patients with these interconnected conditions.

KEYWORDS

breast cancer, diabetes, transcriptomics, exome analysis and TNF pathway, bioinformatics

1 Introduction

Breast cancer (BC) is a multifaceted disease characterized by a wide range of genetic, molecular, and phenotypic variations (1). It remains one of the most prevalent malignancies among women worldwide, with significant heterogeneity in its clinical presentation, prognosis, and therapeutic response (2). Concurrently, diabetes, a chronic metabolic disorder characterized by hyperglycemia and insulin resistance, has been increasingly recognized as a comorbidity that influences cancer risk, progression, and treatment outcomes (3, 4). The intersection of BC and diabetes presents a unique and challenging clinical scenario that warrants a deeper understanding of the underlying molecular mechanisms and potential biomarkers (5). Diabetes has been implicated in altering the tumor microenvironment, enhancing chronic inflammation, promoting oxidative stress, and disrupting metabolic pathways, all of which can contribute to cancer initiation and progression (6, 7). The coexistence of diabetes with BC introduces additional layers of complexity, influencing tumor biology, therapeutic efficacy, and patient survival (3). Patients with diabetes are often associated with poor outcomes, including higher recurrence rates and reduced overall survival, potentially due to delayed diagnosis, altered pharmacokinetics of anticancer drugs, and the impact of hyperglycemia on cancer cell metabolism (8).

Advancements in high-throughput technologies, such as transcriptomics and exome sequencing, have significantly enhanced our ability to understand the molecular landscape of diseases (9). Exome sequencing facilitates the identification of somatic mutations, copy number variations, and other genomic alterations that drive cancer development (10). Conversely, transcriptomics provides insights into gene expression patterns, revealing dysregulated pathways and potential therapeutic targets (11, 12). Integrating transcriptomics and exome data has proven to be a powerful approach to uncover genetic and transcriptomic alterations, providing a more comprehensive understanding of the molecular mechanisms driving conditions such as cancer and other diseases. This integration has the potential to identify novel biomarkers and therapeutic targets. Biomarkers are invaluable for stratifying patients, predicting therapeutic responses, and monitoring disease progression (13–15). Few studies have explored the diabetes-associated gene expression profiles in BC, revealing the unique signatures that could be targeted therapeutically or used as diagnostic tools (16–19). Understanding the molecular interplay between BC and diabetes can pave the way for personalized medicine approaches, ensuring more effective and tailored treatments.

In this study, we aim to explore the biomarker landscape in BC with diabetes (African American (AA) and European American (EA) cohorts) through a comprehensive analysis of transcriptomics and exome data. By examining the transcriptomic and genomic profiles specific to this cohort, we seek to identify key molecular players and pathways that underlie the interaction between these two conditions. Our findings could provide insights into the mechanistic basis of BC in diabetic patients, highlight potential therapeutic vulnerabilities, and contribute to the development of precision oncology strategies. The detailed workflow is illustrated in Figure 1.

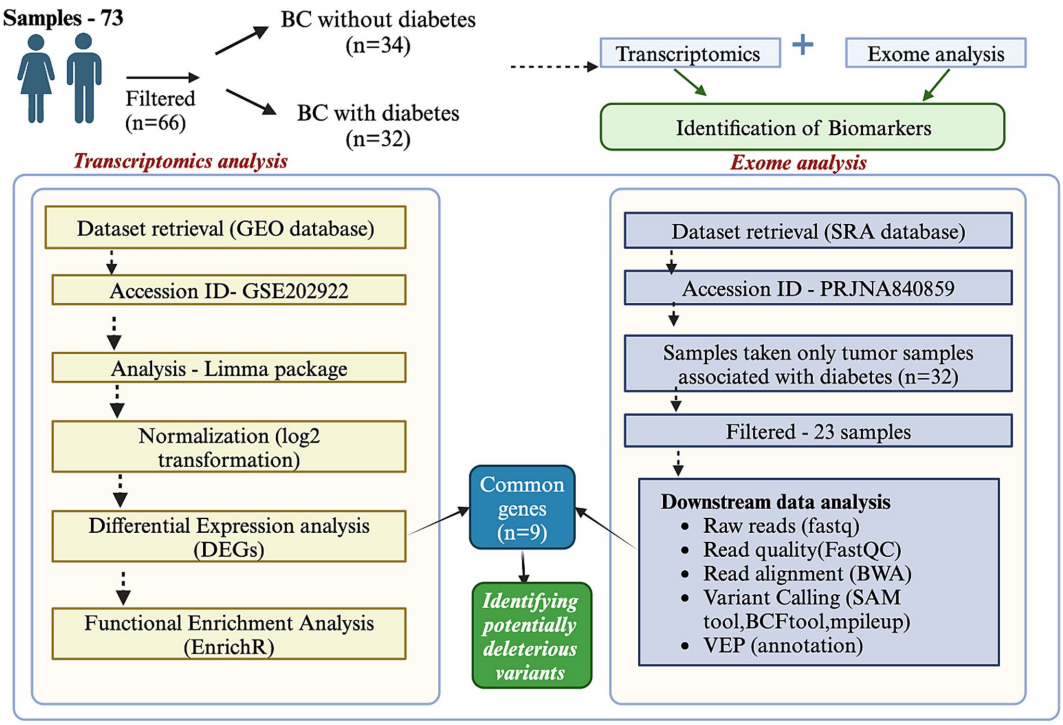


FIGURE 1
Integration of transcriptomics and exome data analysis. The figure illustrates the workflow and outcomes of integrating transcriptomics and exome data analysis. Transcriptomics data provides insights into differential gene expression across conditions, while exome data reveals coding region mutations. The integration identifies overlapping features, including genes with significant expression changes and mutations. This combined approach highlights key biomarkers, potential driver genes, and pathways associated with the biological process of interest.

2 Materials and methods

2.1 Transcriptomics data analysis

All data used in this study were obtained from the NCBI database. We acquired the gene expression profiling dataset produced through high-throughput sequencing (GSE202922) (16) using Illumina HiSeq 3,000 from the publicly available GEO database (20). The dataset has a total of 73 samples, and a further 66 samples have raw counts. A total of 66 samples were included in this study, comprising 32 diabetic and 34 non-diabetic cases. The detailed metadata information, along with transcriptomics data of the 66 samples, were described in Supplementary Table 1 and Supplementary Figure 1A. We also conducted race-specific transcriptomic analyses using datasets from African American (AA) and European American (EA) cohorts. The metadata for these cohorts is provided in Supplementary Table 2. GEO2R is a web-based analysis tool that enables user to compare multiple sample groups within a GEO Series to find deregulated genes under certain experimental conditions (21). Moreover, differentially expressed genes (DEGs) were detected using the limma R package (22), applying a threshold of $|\log2FoldChange| >= 0.5$, adj $p <= 0.05$, and $p <= 0.05$. All statistical analyses and data visualization were carried out using R/Bioconductor packages. Statistical plots such as boxplot and UMAP plot were performed and analyzed.

2.2 Exome data analysis

The study also utilized Whole Exome Sequencing (WES) data with ID: PRJNA840859 comprising 32 individuals with BC-associated diabetes (16). Supplementary Figure 1B provides detailed information on the selected exome data. After verifying the availability of exome data, 23 sample reads were retrieved and analyzed. These samples were subjected to exome sequence analysis. Sequencing was performed on Illumina NovaSeq 6,000 systems, generating paired-end reads. A shell script was employed to download the sequencing reads from the ENA database (23). The exome sequencing pipeline involves a comprehensive workflow for processing, analyzing, and interpreting genetic data to ensure high accuracy and reliability in identifying variants. The process begins with quality control using FastQC (24), which evaluates critical metrics such as read quality scores, GC content, and adapter contamination. This step helps to identify potential issues in the raw FASTQ files, ensuring only high-quality reads proceed to the next stage. Tools like Trimmomatic remove low-quality bases and adapter sequences in the read preprocessing step. Reads with quality scores below a threshold (commonly Q30) are trimmed or discarded, producing a clean dataset suitable for downstream analysis. Next, the high-quality reads are aligned to the human reference genome GRCh38 (25) using the BWA-MEM algorithm (26), a widely used tool for efficient and accurate alignment of short-read sequences. This step generates SAM files containing mapped reads and their corresponding positions on the genome. These SAM files are converted into BAM format using SAMtools, sorted by coordinate order, and indexed to enable efficient querying and visualization in downstream applications. The variant calling step identifies genetic variants such as SNPs and indels (27). BCFtools generate a pileup of aligned reads, and variants are called highly confidently (28). The resulting data is output in the Variant Call Format (VCF), which contains detailed information about each identified variant.

Once variants are called, they undergo filtering and annotation. Each sample VCF was merged using the "VCFmerge tool" and the Galaxy tool. The Ensembl Variant Effect Predictor (VEP) was used to annotate the functional consequences of genes (29). Filtering ensures that only high-confidence variants are retained by removing low-quality or potentially false-positive calls. Tools like the BCFtools filter allow for applying stringent criteria, such as minimum quality scores or read depth thresholds. Annotating the filtered variants with databases such as dbSNP and ClinVar provides functional insights, including potential pathogenicity, population frequency, and relevance to known diseases. The missense variants were retrieved and further used for functional analysis.

2.3 Functional enrichment analysis

Functional analysis of the differentially expressed genes (DEGs) identified from the transcriptomic analysis was conducted using EnrichR (30). Additionally, common genes identified from both transcriptomic and exome analyses were analyzed. Functional enrichment analysis included Gene Ontology categories: Biological Process (GO-BP), Cellular Component (GO-CC), and Molecular Function (GO-MF), as well as pathway analyses using KEGG and Reactome. Protein-coding genes with a p-value < 0.05 were used as the background gene set.

2.4 Identification of potentially deleterious variants

Genes featuring missense variants from a curated in-house list of cancer-associated genes were subsequently examined for functional effects using the PredictSNP web tool (31). This examination utilized six well-known predictive tools, MAPP, PhD-SNP, PolyPhen-1, PolyPhen-2, SIFT, and SNAP, to detect potentially harmful variants (missense).

MAPP demonstrated that the likelihood of disease or cancer risk is closely linked to breaches of physicochemical limitations due to amino acid variations (32). PhD-SNP, based on support vector machines (SVMs), was used to determine whether a given point mutation was a neutral polymorphism or associated with genetic disorders (33). PolyPhen-1 analyzed the impact of missense variants on protein structure and function (34). In contrast, PolyPhen-2 incorporated both sequence- and structure-based features, utilizing a Naïve Bayesian classifier to predict the consequences of amino acid substitutions. Variants identified as "probably damaging" or "possibly damaging" (scores ≥0.5) were categorized as harmful, whereas "benign" variants (scores <0.5) were regarded as acceptable. Scores nearer to 1.0 were more prone to be damaging (35).

SIFT predicted the potential harm of variants using a normalized probability score, where scores <0.05 were deemed harmful and scores ≥0.05 were considered neutral. The SIFT score assessed the effect of amino acid substitutions on protein function (36). SNAP was used to evaluate the functional impact of missense variants (37). Protein stability alterations due to single-point variants were forecasted using

I-Mutant 2.0, which categorized variants into two groups: reduced stability (<0 kcal/mol -decrease) and enhanced stability (>0 kcal/mol – increase) (38).

The evolutionary conservation of amino acid positions for the most deleterious variants were assessed using the ConSurf online tool. Conservation scores ranges from 1 (most variable positions) to 9 (most conserved positions), providing insights into the variants' functional significance (39).

3 Results

3.1 Transcriptomics and functional analysis of pooled cohort

To identify differentially expressed genes (DEGs) between BC patients with and without diabetes, we utilized normalized expression data from the GEO database. The GEO2R tool, based on the limma package, was employed for the analysis. 2,814 DEGs were analyzed across 66 samples, including BC without and BC with diabetes samples. A boxplot is a graphical representation of the distribution of a dataset that shows its central tendency and variability. It provides a concise summary of the data's statistical properties across samples. The boxplot of each group comparison is mentioned in the Figure 2A. A UMAP plot is a dimensionality reduction technique that is particularly useful for visualizing high-dimensional data, such as gene expression values plotted in Figure 2B. The comparison revealed 2,814 DEGs comprising 1824 upregulated and 990 downregulated genes (p-value <= 0.05, adj p-value <= 0.05, |log2 fold change| > =0.5). The DEGs were visualized using a volcano plot (Figure 2C), highlighting significant genes with biological relevance. The detailed results of DEGs are mentioned in Supplementary Table 3. The heatmap illustrates the expression levels of selected genes across 66 samples, with rows representing genes and columns representing samples, as mentioned in Figure 2D.

The functional enrichment analysis of 2,814 DEGs was performed using EnrichR. The background genes are protein-coding genes with p-value<=0.05. The functional terms are GO (Gene ontology) terms and KEGG pathways. Significant enrichment is seen in processes such as extracellular matrix organization, regulation of cell migration, angiogenesis, and circulatory system development. These are key processes in tissue remodeling, cancer metastasis, and vascular development. Highlighted components include collagen-containing extracellular matrix, cell junctions, plasma membrane raft, and sarcolemma. These components are critical for cellular integrity, signaling, and intercellular communication. Functions such as tyrosine kinase activity, platelet-derived growth factor binding, and kinase inhibitor activity dominate. These molecular functions are

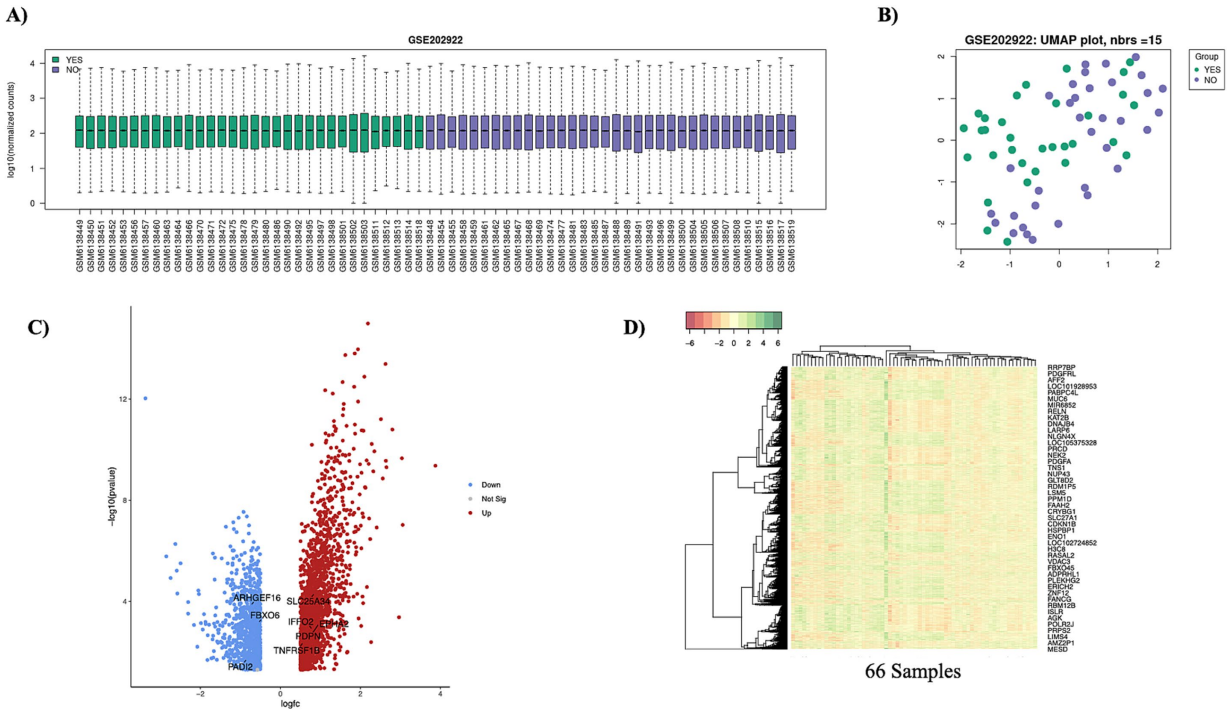


FIGURE 2
Statistical plots of transcriptomics data. **(A)** The boxplot represents the distribution of normalized transcriptomics data across all samples. Each box corresponds to an individual sample, with the central line representing the median expression level. **(B)** The UMAP plot illustrates the clustering of transcriptomics data, with each point representing an individual sample. Samples are color-coded based on their respective groups (BC with diabetes vs. BC without diabetes). This visualization highlights the underlying structure and relationships in the dataset, revealing group-specific patterns. The "YES" label represents the BC with diabetes, and the "NO" label represents the BC without diabetes. The color-coded representation of the "YES" label is green, and the "NO" label is purple. **(C)** The volcano plot shows the relationship between statistical significance for all genes. Significant upregulated and downregulated genes are highlighted in distinct colors with respective thresholds. This visualization identifies key differentially expressed genes. The red denotes the upregulated genes, and the blue indicates the down-regulated genes. **(D)** The heatmap visualizes the expression levels of selected genes across 66 samples. Rows represent genes, and columns represent samples.

often associated with signaling pathways and therapeutic targets in cancer and other diseases. Enriched pathways include systemic lupus erythematosus, cell cycle regulation, ECM-receptor interaction, and PI3K-Akt signaling. These pathways are relevant to immune disorders, cancer progression, and extracellular matrix interactions. The results suggest an association with processes and pathways related to cancer progression, immune regulation, and extracellular matrix dynamics. The length of the bar indicates the top function in the barplot. The detailed functional results are mentioned in Figure 3.

3.2 Transcriptomics and functional analysis of AA and EA cohorts

To identify DEGs between BC patients with and without diabetes, we analyzed normalized expression data from the GEO database separately for African American (AA) and European American (EA) cohorts. The analysis was performed using the GEO2R tool, which is based on the limma package. In the African American (AA) cohort, a total of 3,245 differentially expressed genes (DEGs) were identified from 57 samples, including 1,922 upregulated and 1,323 downregulated genes, based on thresholds of p-value ≤ 0.05, adjusted p-value ≤ 0.05, and $|\log_2$ fold change$| \geq 0.5$. Similarly, in the European American (EA) cohort, 3,208 DEGs were detected across 17 samples, with 1,640 genes upregulated and 1,568 downregulated using the same statistical criteria. Notably, 786 DEGs were found to be shared between the AA and EA cohorts. The statistical plots of boxplot and UMAP were performed and mentioned in Supplementary Figure 2. The detailed information of DEGs of both the cohorts were mentioned in the Supplementary Tables 4, 5.

The functional enrichment analysis of each cohort was performed. The functional enrichment analysis of 3,245 (AA cohort) and 3,208 (EA cohort) DEGs was performed using EnrichR. The background genes are protein-coding genes with p-value<=0.05. Some of the KEGG's significant functions are cell cycle, ECM receptor interactions, PI3K-Akt signaling, and AGE-RAGE signaling pathway in diabetic complications, and these functions were specific to the AA cohort. Some of the important functions in the EA cohort are Oxidative phosphorylation and Diabetic cardiomyopathy. The detailed information on these enrichment analyses is mentioned in Supplementary Figures 3, 4.

The Venn diagram illustrates the overlap in transcriptomic data between the African American (AA) and European American (EA) cohorts, revealing 786 genes common to both groups (Supplementary Figure 5A). Further functional analysis of GO-BP,GO-CC,GO-MF and KEGG pathways were performed on common genes. Some of the important functions are Notch signaling pathway and Hippo signaling pathway. Both these functions were related to BC and diabetes. The detailed functional enrichment analysis were mentioned in the Supplementary Figures 5B–E.

3.3 Exome data analysis

We retrieved WES datasets for BC with diabetes from the NCBI SRA database. The tumor data of BC with diabetes ($n = 23$) were only taken for further analysis. Each sample was processed using a computational pipeline tailored to laboratory protocols. Sequence quality was assessed using the FastQC tool. High-quality data for analysis was ensured by trimming low-quality reads, removing

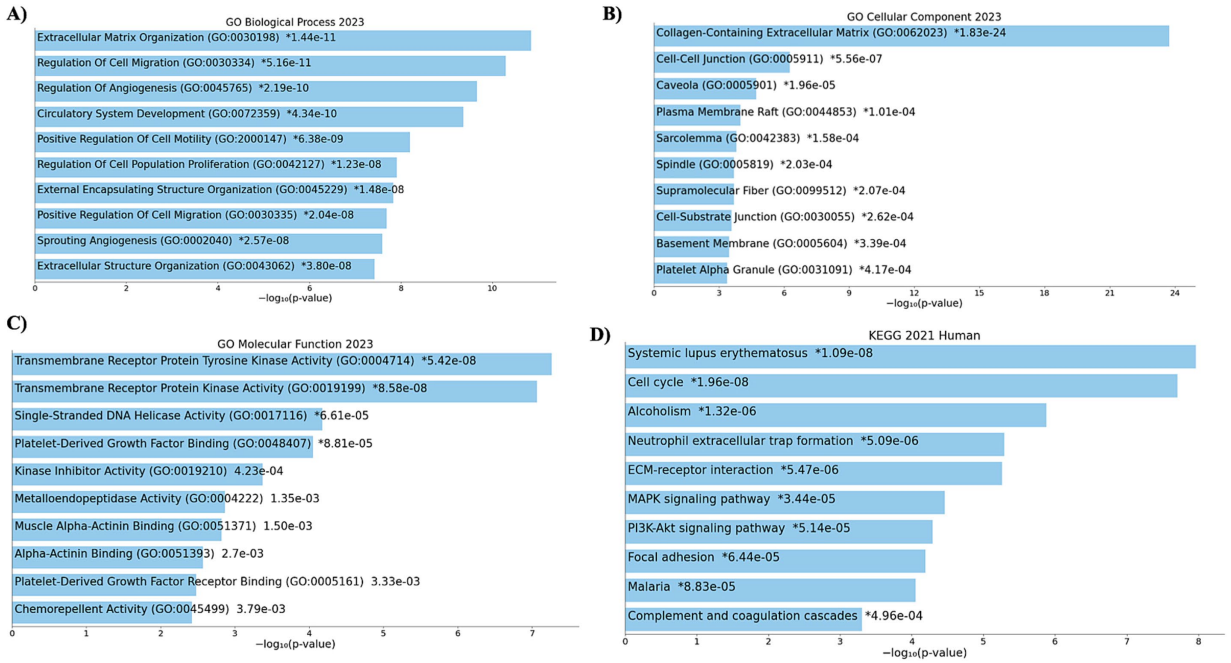


FIGURE 3
Functional enrichment analysis (transcriptomics). The figure showcases the functional enrichment analysis of differentially expressed genes (DEGs) derived from transcriptomics data. **(A)** GO-BP, **(B)** GO-CC, **(C)** GO-MF, and **(D)** KEGG pathways. Bar sizes indicate statistical significance (adjusted p-values) and gene ratios, providing insights into DEGs' molecular and functional context.

adapters, and further validating the sequences' base quality. Following the evaluation of read quality, the final reads were mapped to the human reference genome GRCh38.p13 (hg38) utilizing the BWA aligner with default settings. Every dataset attained a total alignment rate surpassing 85%. SAM tools were employed to process and enhance the sequenced files further in the "SAM" format. The SAM files were first transformed into BAM format by utilizing the "samtools view" command. This transformation enabled later processes, including file sorting, indexing, and arranging mapped reads for further analysis. Prior to indexing, samtools organized the aligned reads and clustered them according to particular genomic areas. In conclusion, base calls from the mapped reads aligned to the reference sequence were compiled using the "samtools mpileup" command.

After processing the output from "mpileup" with BCFtools, SNPs in relation to the reference genome were identified and interpreted as variations. The VCF and its binary counterpart, BCF, were used in the analysis to handle the data. The resulting output for each dataset was provided in a VCF format, containing detailed information about variant positions, types, and quality. Each VCF file was annotated using the Ensembl VEP database (release 113) which provided a thorough analysis of the variants detected in each sample. All identified variants are single-nucleotide (SNVs), accounting for 100% of the dataset. There are two types of variants: non-coding variants and coding variants. The non-coding variants constitute 56.9% of the total, including regions like upstream, downstream, and intronic variants. The coding variants represent 43.1%, further categorized into missense variants 96%, synonymous variants 3%, stop-gained 1%, and no start-lost variants. The results are depicted in Figure 4A. Among the 3,238 observed missense variants in the VCF file, filtered 899 missense variants were chosen for further analysis (no novel variants were detected) (Figure 4B). After removing duplicate genes in the 298 overlapped genes, 208 unique genes were finalized for further analysis. The detailed results of missense variants are mentioned in Supplementary Table 6. The distribution of 208 genes with respective metadata is depicted in Figure 5.

3.4 Identification of shared genes and their respective functional analysis

The Venn diagram illustrates the overlap between transcriptomic and exomic data. A total of 2,804 unique genes were identified only in transcriptomic analysis. One hundred ninety-seven unique genes are found exclusively in the exomic data. Eleven genes are shared between the datasets, representing key potentially important genes across transcriptional and mutational levels. The two genes with no variations were excluded from the analysis. The nine genes comprises six upregulated genes (*SKI, TNFRSF1B, PDPN, SLC25A34, EPHA2,* and *IFFO2*) and three down-regulated genes (*ARHGEF16, FBXO6,* and *PADI2*). The results of overlap genes are mentioned in Figure 6A. We compared the selected genes across two different cohorts AA and EA populations. Four genes such as *SKI, TNFRSF1B, SLC25A34,* and *EPHA2,* were present in both cohorts.

The functional analysis of nine genes was performed and analyzed, including the functional categories such as GO-BP, GO-CC, GO-MF, and the Reactome pathway. The GO-BP enriched terms include

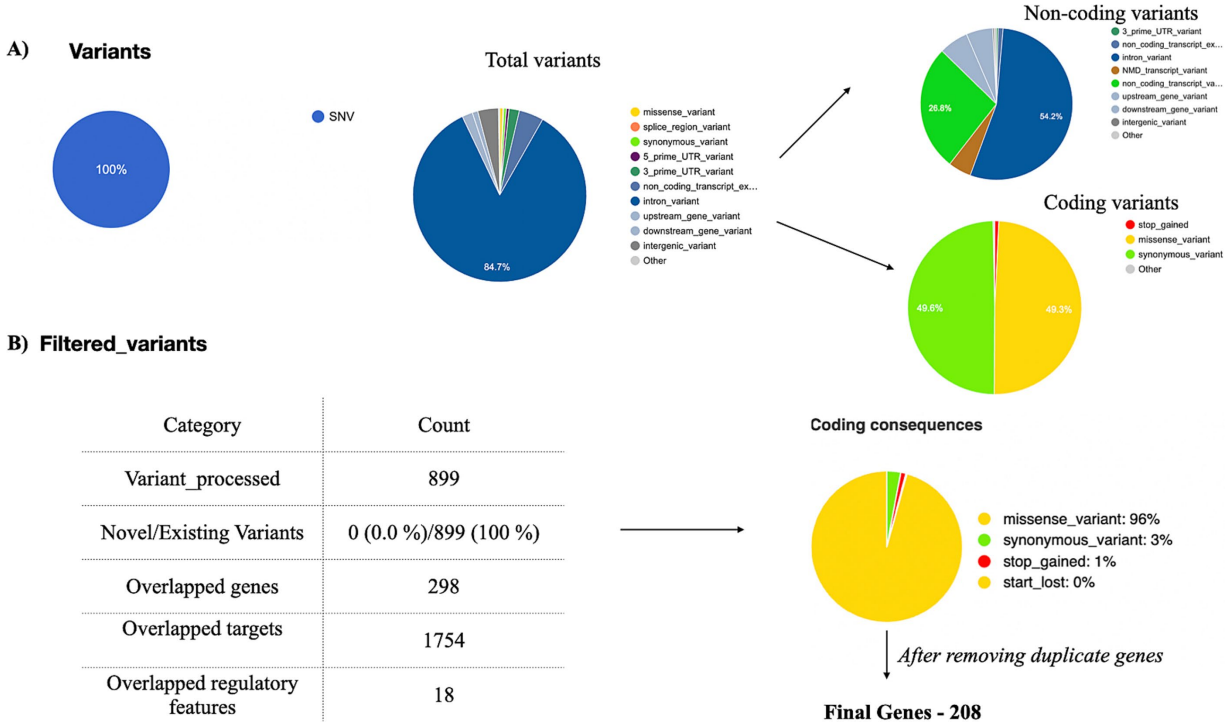


FIGURE 4
Exome data analysis. Exome data analysis was conducted on tumor samples (n = 23) from breast cancer patients with diabetes. Variants were annotated using the VEP tool. **(A)** A pie chart was generated to classify coding and non-coding variants. **(B)** Among the filtered coding variants, 96% were identified as missense variants. A total of 208 missense variants, derived from all 23 samples, were selected for further analysis.

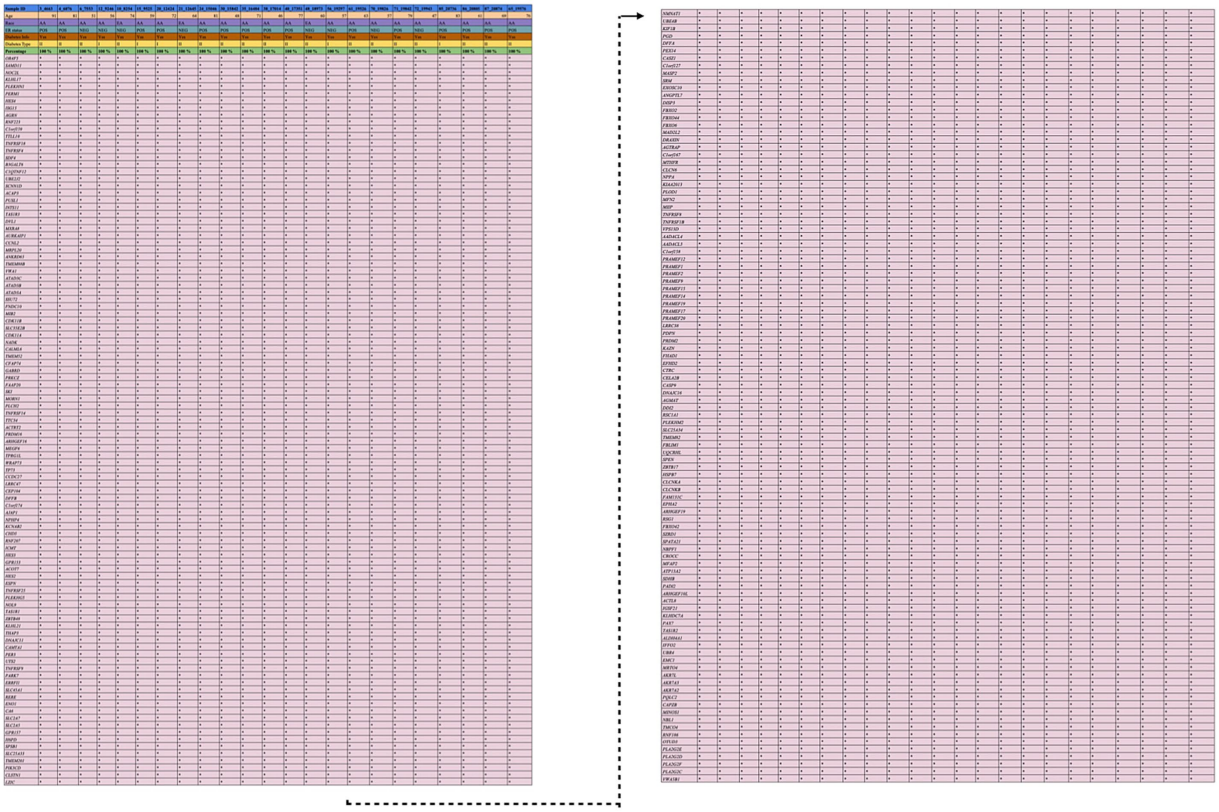


FIGURE 5
Distribution of variants. This figure illustrates the distribution of genetic variants observed in 23 samples, stratified by metadata variables such as age, race, ER status, diabetes type, and diabetes info. Each sample is represented as a distinct bar or point, categorized by metadata groups.

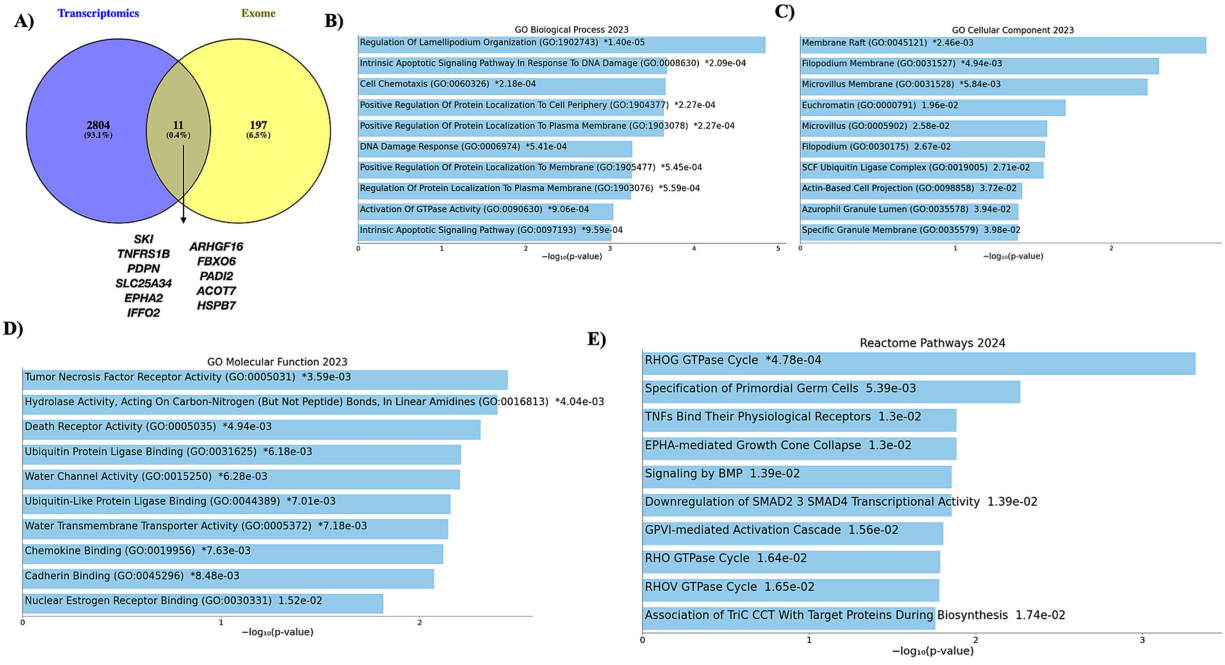


FIGURE 6
Overlapping genes and functional analysis. This figure shows the overlapping genes identified through integrative analysis of exome sequencing and transcriptomics data. The shared genes represent a subset with potential biological functions. **(A)** Eleven genes overlapped the exome and transcriptomics data. The functional analysis of common genes was performed. **(B)** GO-BP, **(C)** GO-CC, and **(D)** GO-MF. **(E)** Reactome pathways.

protein localization processes and synaptic pathways, reflecting cellular organization and signaling roles. The GO-CC, the enriched components, such as synaptic and endosomal compartments, highlight cellular compartmentalization for signaling and transport. The GO-MF of enriched functions includes ubiquitin-protein ligase binding, TNF activity, chemokine activity, and cadherin binding, which are crucial for protein regulation and cellular interactions. The Reactome pathways enriched are RHOG GTPase cycle, TNFs Bind their physiological receptors. The other pathways include EPHA-mediated growth cone collapse and signaling pathways. The functional results are mentioned in the Figures 6B–E. These variations were taken for further analysis.

3.5 Identification of potential variants

The analysis involved nine genes, focusing on identifying the most deleterious variants using a comprehensive suite of online prediction tools. These tools included PredictSNP, MAPP, PhD-SNP, PolyPhen-1, PolyPhen-2, SIFT, SNAP, and PANTHER, each offering distinct methodologies for assessing variant pathogenicity. The prediction results provided detailed insights into the potential impact of these variants on protein function and structure. A summary of the findings, highlighting the pathogenicity scores from each tool for the identified variants, is presented in Table 1. This table serves as a consolidated resource, showcasing the comparative outcomes from all tools, thus facilitating an in-depth evaluation of the most deleterious genetic changes. Among these nine genes, the TNFRSF1B (L264P) and PDPN (A105G) were the top 2 variants predicted by the above tools.

Variants classified as neutral were excluded from stability analysis using I-Mutant 2.0. analysis. Among these nine gene variants, the results revealed distinct patterns of stability changes. Three gene variants exhibited an increase in protein stability upon mutation. This indicates that these mutations potentially enhance the structural integrity or thermodynamic stability of the proteins, which could impact their functional roles positively or negatively, depending on the biological context. The remaining six gene variants showed a decrease in protein stability upon mutation. A reduction in stability suggests that these mutations may disrupt the protein's structural conformation, potentially leading to misfolding, aggregation, or loss of function (38). Such destabilizing mutations could contribute to disease pathogenesis or altered protein activity. The results are mentioned in Table 2. Among these mutations, the TNFRSF1B variant (L264P) is the most deleterious variant confirmed by all computational tools. ConSurf is a tool that analyses the evolutionary conservation of amino acid positions in protein sequences. The variant (TNFRSF1B-L264P) is categorized as a highly conserved position with a significant score (score range of 9), it may suggest a deleterious impact. The ConSurf results are mentioned in Figure 7.

4 Discussion

Integrating transcriptomics and exomic analyses combines the strengths of both methods to achieve a comprehensive understanding of genomic and transcriptomic changes in biological systems (40, 41). This integration represents a powerful approach to elucidating the molecular mechanisms underlying complex diseases, facilitating the identification of robust biomarkers and therapeutic targets (42). Several studies have successfully integrated transcriptomics and exomic data to provide deeper insights into biological mechanisms, disease pathogenesis, and therapeutic strategies (43–47). The etiology of BC associated with diabetes remains poorly understood. We aimed to identify differentially expressed genes (DEGs) between BC patients without diabetes and those with diabetes. Our analysis included 66 samples, comparing BC without diabetes to BC with diabetes, and identified 2,815 DEGs, comprising 1,824 upregulated and 990 downregulated genes with statistical significance. This integrative analysis provides insights into the gene expression changes associated with diabetes in BC patients, with visualizations effectively summarizing statistical properties, significant DEGs, and their expression patterns.

Functional enrichment analysis of the 2,814 DEGs was performed using EnrichR, with protein-coding genes as the background. The study focused on GO terms and KEGG pathways, revealing significant enrichment in processes and pathways related to cancer progression, immune regulation, and extracellular matrix (ECM) dynamics. Notable enrichment was observed in processes such as extracellular matrix organization, regulation of cell migration, angiogenesis, and circulatory system development. These processes are crucial for tissue remodeling, cancer metastasis, and vascular development. Key components highlighted include collagen-containing extracellular matrix, cell junctions, plasma membrane rafts, and sarcolemma, which are essential for cellular integrity, signaling, and intercellular communication. Functions like tyrosine kinase activity, platelet-derived growth factor binding, and kinase inhibitor activity were dominant, indicating relevance to signaling pathways and therapeutic targets. The enrichment analysis underscores the involvement of key processes, components, and pathways in cancer progression, immune system regulation, and extracellular matrix interactions, offering potential insights into disease mechanisms and therapeutic targets. These processes are fundamental biological functions and pathways in BC and diabetes, as reported in several studies (48–54).

To identify DEGs between BC patients with and without diabetes, normalized expression data from the GEO database were analyzed separately for AA and EA cohorts. In the AA cohort, 3,245 DEGs were identified from 57 samples, including 1,922 upregulated and 1,323 downregulated genes, while in the EA cohort, 3,208 DEGs were detected across 17 samples, with 1,640 upregulated and 1,568 downregulated genes. A total of 786 DEGs were found to be common between the two cohorts. Key KEGG pathways identified in the AA cohort included cell cycle, ECM-receptor interaction, PI3K-Akt signaling, and AGE-RAGE signaling in diabetic complications. In contrast, significant pathways in the EA cohort included oxidative phosphorylation and diabetic cardiomyopathy. A Venn diagram illustrating the overlap between AA and EA transcriptomic profiles revealed 786 shared genes. Some of the key KEGG pathways, such as Notch signaling and Hippo signaling, both of which are relevant to breast cancer and diabetes (55–58).

The study analyzed WES data from 23 BC patients with diabetes, sourced from the NCBI SRA database, using a customized computational pipeline. Annotation via the Ensembl VEP database classified these variants into non-coding (56.9%) and coding (43.1%). Among coding variants, 96% were missense, 3% synonymous, and 1% stop-gained, with no start-lost variants detected. A total of 899 variants were analyzed, with no novel variants identified. After

TABLE 1 Prediction of deleterious variants of common genes by different tools.

Protein	Uniport ID	Amino_acid_change	Existing_variation	PredictSNP prediction	MAPP prediction	PhD-SNP prediction	PolyPhen-1 prediction	PolyPhen-2 prediction	SIFT prediction	SNAP prediction	PANTHER prediction
SKI	P12755	A62G	rs28384811	Neutral	Neutral	Neutral	Neutral	Deleterious	Deleterious	Neutral	Unknown
SKI	P12755	E491D	rs1266460001	Neutral	Neutral	Neutral	Neutral	Deleterious	Neutral	Neutral	Neutral
TNFRSF1B	P20333	M196R	rs1061622	Neutral	Deleterious	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
TNFRSF1B	P20333	E232K	rs5746026	Neutral	Deleterious	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
TNFRSF1B*	P20333	L264P	rs2229700	Deleterious	Deleterious	Deleterious	Deleterious	Neutral	Deleterious	Deleterious	Deleterious
PDPN	Q86YL7	M43V	rs141726617	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
PDPN*	Q86YL7	A105G	rs2486188	Neutral	Deleterious	Neutral	Deleterious	Deleterious	Neutral	Neutral	Neutral
PDPN	Q86YL7	A147G	rs2486188	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
SLC25A34	Q6PIV7	I215M	rs62621224	Neutral	Deleterious	Deleterious	Neutral	Neutral	Neutral	Neutral	Neutral
EPHA2	P29317	V747I	rs145592908	Neutral	Neutral	Deleterious	Neutral	Deleterious	Neutral	Neutral	Unknown
EPHA2	P29317	M631T	rs34021505	Neutral	Neutral	Neutral	Deleterious	Neutral	Neutral	Neutral	Neutral
EPHA2	P29317	V541M	rs61731097	Neutral	Neutral	Neutral	Neutral	Neutral	Deleterious	Neutral	Neutral
EPHA2	P29317	G391R	rs34192549	Neutral	Na	Neutral	Deleterious	Neutral	Deleterious	Neutral	Neutral
EPHA2	P29317	D232G	rs114498261	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral
IFFO2	Q5TF58	V352I	rs6675316	Neutral	Neutral	Neutral	Neutral	Deleterious	Neutral	Neutral	Neutral
ARHGEF16	Q5VV41	V137M	rs3806164	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Deleterious
ARHGEF16	Q5VV41	H370Y	rs2185639	Neutral	Neutral	Deleterious	Neutral	Neutral	Neutral	Neutral	Deleterious
FBXO6	Q9NRD1	R60Q	rs3125818	Neutral	Neutral	Neutral	Neutral	Deleterious	Neutral	Neutral	Neutral
FBXO6	Q9NRD1	V72M	rs766167101	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Neutral	Unknown
FBXO6	Q9NRD1	V290I	rs140436527	Neutral	Na	Neutral	Neutral	Neutral	Neutral	Neutral	Unknown
PADI2	Q9Y2J8	Y275H	NA	Neutral	Neutral	Neutral	Deleterious	Deleterious	Neutral	Neutral	Neutral
PADI2	Q9Y2J8	D259N	rs150731573	Neutral	Neutral	Deleterious	Neutral	Neutral	Deleterious	Neutral	Neutral

* Represents the most deleterious variants from all the tools.

10.3389/fmed.2025.1572500

TABLE 2 Prediction of protein stability using I-Mutant 2.0.

Protein	Uniport ID	Amino_acid_ change	Existing_ variation	Stability	RI	DDG_value (kcal/mol)
SKI	P12755	A62G	rs28384811	Decrease	5	−0.28
SKI	P12755	E491D	rs1266460001	Increase	6	0.04
TNFRSF1B	P20333	M196R	rs1061622	Decrease	7	−1.07
TNFRSF1B	P20333	L264P	rs2229700	Decrease	4	−2.29
PDPN	Q86YL7	A105G	rs2486188	Decrease	7	−1.68
SLC25A34	Q6PIV7	I215M	rs62621224	Decrease	7	−1.91
EPHA2	P29317	V747I	rs145592908	Decrease	7	−0.57
EPHA2	P29317	M631T	rs34021505	Increase	1	0.1
EPHA2	P29317	D232G	rs114498261	Decrease	3	−0.9
IFFO2	Q5TF58	V352I	rs6675316	Decrease	8	−1.04
ARHGEF16	Q5VV41	V137M	rs3806164	Decrease	7	−2.43
ARHGEF16	Q5VV41	H370Y	rs2185639	Increase	4	2.02
FBXO6	Q9NRD1	R60Q	rs3125818	Decrease	8	−0.94
PADI2	Q9Y2J8	Y275H	NA	Decrease	6	−0.62
PADI2	Q9Y2J8	D259N	rs150731573	Decrease	0	−0.56

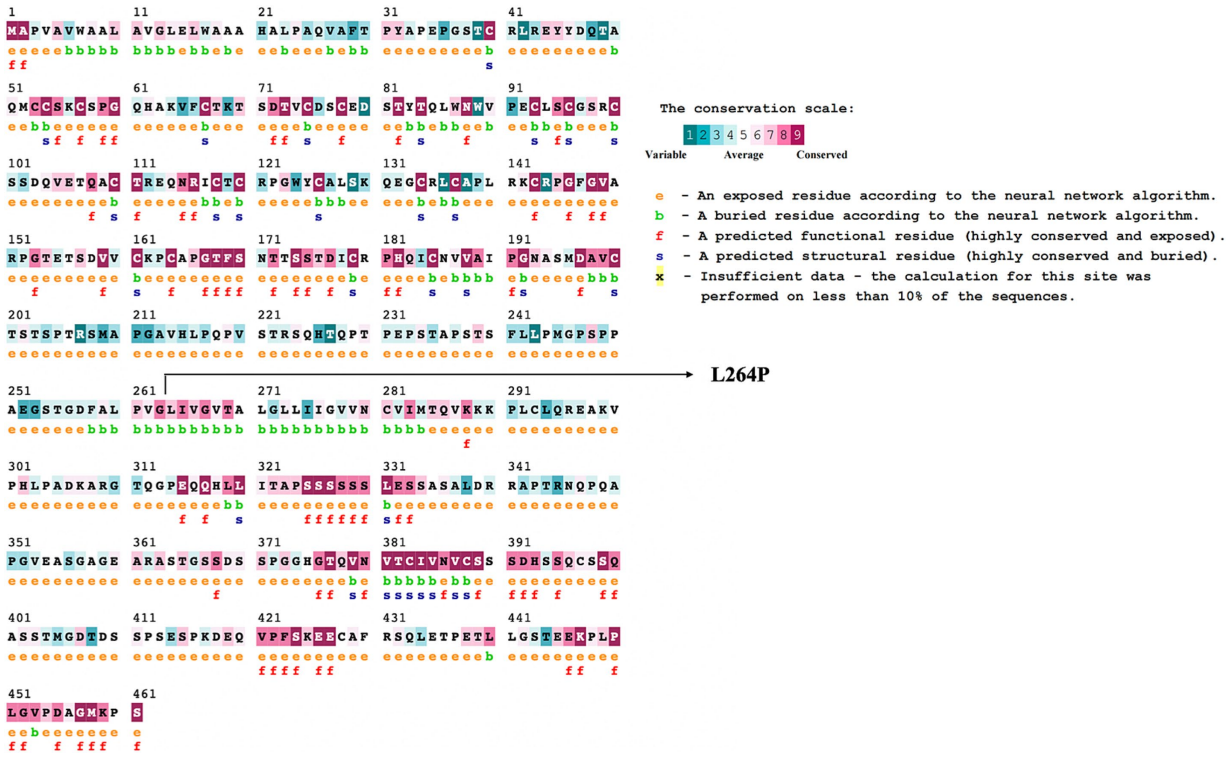


FIGURE 7
ConSurf analysis of potential deleterious variant (TNFRSF1B-L264P). This figure presents the results of a ConSurf analysis, highlighting the evolutionary conservation of amino acid residues in the TNFRSF1B (L264P) protein. Residues are color-coded based on their conservation scores, ranging from highly conserved (dark shades) to variable (light shades).

removing duplicates, these variants spanned 298 genes, which were reduced to 208 unique genes. A Venn diagram illustrated the overlap between transcriptomic and exomic datasets, identifying 2,804 genes unique to transcriptomics, 197 genes exclusive to exomics, and 11 common genes (Figure 6A). Two genes without mutations were excluded, leaving nine key genes: *SKI* (59, 60), *TNFRSF1B* (61, 62), *PDPN* (62, 63), *SLC25A34* (64), *EPHA2* (65, 66), *IFFO2* (67, 68), *ARHGEF16* (69, 70), *FBXO6* (71, 72), and *PADI2* (73, 74) for further

analysis. Among these, six were upregulated, and three were downregulated. These genes play significant roles in both diabetes and BC. We analyzed gene expression across two cohorts—AA and EA populations and mapped these four genes (*SKI*, *TNFRSF1B*, *SLC25A34*, and *EPHA2*) that were consistently present in both groups.

Functional analysis of these nine genes revealed enriched terms across GO categories and Reactome pathways. GO-BP terms included processes like protein localization and synaptic pathways. GO-CC analysis highlighted synaptic and endosomal compartments, indicating roles in cellular organization and signaling. GO-MF terms included ubiquitin-protein ligase binding, TNF activity, chemokine activity, and cadherin binding, essential for protein regulation and interactions. These biological functions were enriched in BC and diabetes in other studies (75–79). Reactome pathways featured RHOG GTPase cycle, TNF-receptor binding, EPHA-mediated growth cone collapse, and other signaling pathways. Among these, the TNF pathway is significant in connecting BC and diabetes (16, 80, 81). The analysis focused on identifying the most deleterious variants using a comprehensive suite of online prediction tools. Among the nine genes analyzed, TNFRSF1B (L264P) and PDPN (A105G) were identified as the top two variants predicted to be most deleterious. These mutations remain poorly characterized and have not been extensively studied. TNFRSF1B (also known as TNFR2), a receptor for the pro-inflammatory cytokine TNF-α, is primarily expressed in immune cells, endothelial cells, and certain tumor cells, playing a pivotal role in immune regulation, inflammation, and cell survival. As chronic inflammation is a common feature of both BC and diabetes, TNFRSF1B may represent a molecular link between these diseases. It contributes to shared inflammatory pathways by promoting a pro-inflammatory microenvironment, and the presence of missense mutations in TNFRSF1B among BC patients with diabetes may exacerbate both tumor progression and metabolic dysfunction. Given its involvement in both cancer and metabolic disease, TNFRSF1B holds potential as a biomarker for identifying at-risk BC patients with diabetes and guiding personalized treatment strategies. Moreover, targeting TNFRSF1B signaling such as through TNF-α inhibitors could offer therapeutic benefits by mitigating inflammation and tumor development. Understanding genetic variations in TNFRSF1B may also inform precision medicine approaches that address the dual challenges of cancer and metabolic dysregulation (5, 16, 82–84).

TNF pathway plays a crucial role in linking chronic inflammation, metabolic dysfunction, and cancer progression, providing an everyday mechanistic basis for its involvement in diabetes and BC. TNF, produced by adipocytes and macrophages in adipose tissue, is elevated in obesity and diabetes (85, 86). It inhibits insulin signaling by phosphorylating insulin receptor substrate-1 (*IRS1*), disrupting pathways essential for glucose uptake. TNF-induced NF-κB activation and oxidative stress exacerbate inflammation, worsening insulin resistance (87). TNF-mediated inflammation also contributes to beta-cell dysfunction, reducing insulin secretion. Prolonged TNF signaling increases circulating free fatty acids, further impairing metabolic homeostasis (88). In BC, chronic TNF secretion by cancer-associated macrophages and stromal cells creates a pro-inflammatory environment that supports tumor growth (89). NF-κB activation in cancer cells increases the expression of anti-apoptotic genes, helping tumor cells evade programmed cell death (90). TNF drives epithelial-to-mesenchymal transition (EMT), enhancing cancer cell motility and invasion, and promotes angiogenesis via VEGF induction, facilitating tumor vascularization and growth (91). The cross-talk between diabetes and BC with shared mechanisms. Obesity and hyperglycemia heighten TNF levels, creating a pro-inflammatory milieu (92). TNF exacerbates oxidative stress, which damages DNA and increases cancer risk (93). TNF-mediated immune suppression allows cancer cells to escape immune surveillance. Insulin resistance and hyperinsulinemia, driven by TNF, activate pathways like PI3K/AKT, promoting cancer cell proliferation (94). Elevated TNF levels in diabetic patients may accelerate BC progression through increased inflammation and angiogenesis (95, 96). These mechanisms are illustrated in a simplified manner in Figure 8.

Our analysis identifies the TNF pathway as a crucial mediator in the interplay between BC and diabetes. While pathways such as PI3K-AKT, JAK–STAT, and mTOR are also implicated, our differential expression analysis reveals a significant enrichment of TNF receptor activity among genes common to both conditions. This indicates that TNF signaling plays a pivotal role in inflammation, apoptosis, and immune regulation, potentially driving the interactions between these diseases. Although the PI3K-AKT and MAPK pathways contribute broadly, TNF signaling stands out as a central hub, highlighting its potential as a therapeutic target (5, 50, 97). Further studies are needed to refine these insights. Targeting the TNF gene or its variants could have substantial therapeutic implications, especially for research on comorbidities. Anti-TNF therapies could reduce inflammation, benefiting patients with both metabolic disorders and cancer. Combining TNF inhibitors with treatments specific to metabolic or cancer conditions may offer synergistic benefits, particularly for patients with both diabetes and BC. The TNF pathway exemplifies how chronic inflammation is a common factor in complex diseases like diabetes and BC, emphasizing the importance of addressing systemic inflammation in therapeutic strategies.

5 Conclusion

This study provides a comprehensive examination of the biomarker landscape in BC associated with diabetes through integrative transcriptomics and exome analysis. Utilizing computational approaches, we identified key differentially expressed genes, mutations, and genes with potential deleterious variants that may elucidate the interplay between these conditions. Our findings highlight potential biomarkers and therapeutic targets that could enhance stratification, diagnosis, and treatment for patients with comorbid BC and diabetes. Future studies validating these biomarkers in experimental and clinical settings could significantly advance our understanding and management of this complex disease intersection.

5.1 Limitation of the study

We acknowledge the limitation of our Whole Exome Sequencing (WES) analysis due to the relatively small sample size ($n = 23$). This constraint primarily arises from our focus on integrating transcriptomic and exomic data specifically for BC patients with diabetes, ensuring a well-defined cohort for robust multi-omics analysis. Additionally, the stringent patient selection

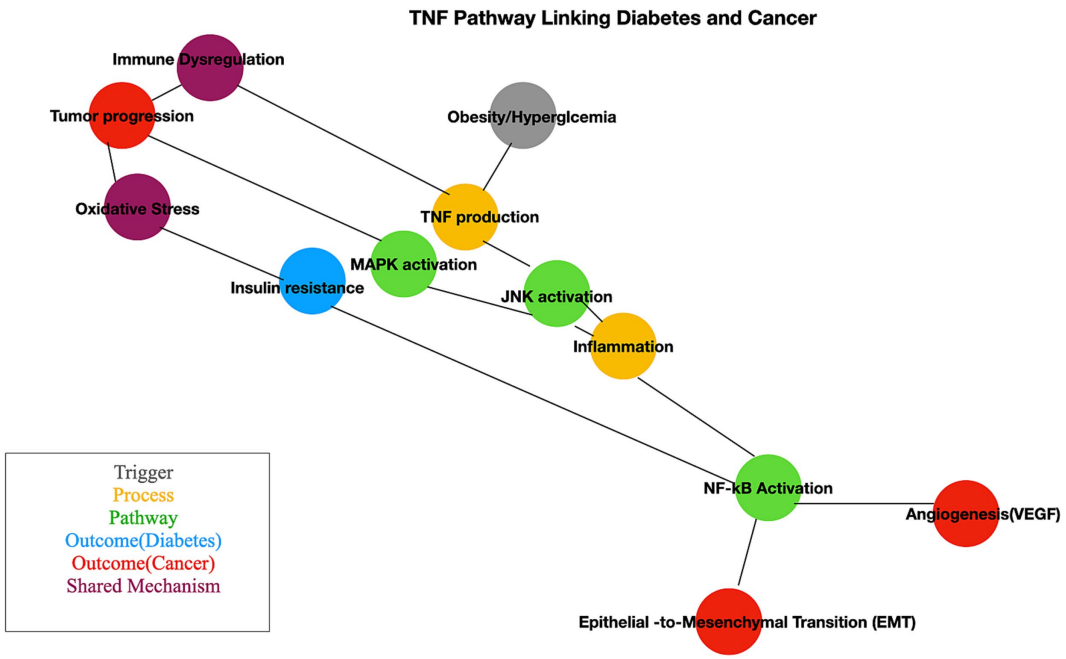


FIGURE 8
Pathway mechanism linked with diabetes and cancer. This schematic illustrates the interconnected molecular mechanisms linking diabetes and cancer. Key pathways include insulin signaling, chronic inflammation, oxidative stress, and altered metabolism. The figure highlights how hyperinsulinemia and insulin resistance influence cancer cell proliferation and survival through pathways like PI3K/AKT/mTOR and MAPK.

criteria and data availability restricted our analysis to tumor samples alone, as paired normal controls were not available within the dataset.

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

TL: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. CD: Conceptualization, Formal analysis, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing.

Funding

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

Acknowledgments

The authors would like to take this opportunity to thank the management of Vellore Institute of Technology (VIT), Vellore, India, for providing the necessary facilities and encouragement to carry out this work.

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.

Generative AI statement

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

Publisher's note

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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2025.1572500/full#supplementary-material

SUPPLEMENTARY FIGURE 1
(A) Selection of samples from transcriptomics data. **(B)** Selection of samples from exomic data.

SUPPLEMENTARY FIGURE 2
Statistical plots of transcriptomics data. The boxplot represents the distribution of normalized transcriptomics data across all samples. Each box corresponds to an individual sample, with the central line representing the median expression level. **(A)** Boxplot of AA cohort; **(B)** Boxplot of EA cohort. The UMAP plot illustrates the clustering of transcriptomics data, with each point representing an individual sample. Samples are color-coded based on their respective groups (BC with diabetes vs BC without diabetes). **(C)** UMAP of AA cohort; **(D)** UMAP of EA cohort.

SUPPLEMENTARY FIGURE 3
The figure presents the functional enrichment analysis of DEGs identified from transcriptomic data of the AA cohort. Panels include **(A)** GO-BP, **(B)** GO-CC, **(C)** GO-MF, and **(D)** KEGG pathways. Bar lengths represent both statistical significance (adjusted *p*-values) and gene ratios, offering insights into the molecular roles and functional relevance of the DEGs.

SUPPLEMENTARY FIGURE 4
The figure presents the functional enrichment analysis of DEGs identified from transcriptomic data of the AA cohort. Panels include **(A)** GO-BP, **(B)** GO-CC, **(C)** GO-MF, and **(D)** KEGG pathways. Bar lengths represent both statistical significance (adjusted *p*-values) and gene ratios, offering insights into the molecular roles and functional relevance of the DEGs.

SUPPLEMENTARY FIGURE 5
This figure displays the overlapping genes identified through a transcriptomics analysis of the AA and EA cohort. These shared genes represent a subset with potential biological significance and functional relevance. **(A)** 786 overlapping genes were present in both cohorts. The functional analysis of common genes was performed. **(B)** GO-BP, **(C)** GO-CC, and **(D)** GO-MF. **(E)** KEGG pathways.

SUPPLEMENTARY TABLE 1
Metadata information of transcriptomics data.

SUPPLEMENTARY TABLE 2
Detailed metadata information of AA and EA cohorts.

SUPPLEMENTARY TABLE 3
Detailed information of differentially expressed genes (pooled cohort) in comparison of BC with and without diabetes.

SUPPLEMENTARY TABLE 4
Detailed information of differentially expressed genes of the AA cohort in comparison of BC with and without diabetes.

SUPPLEMENTARY TABLE 5
Detailed information of differentially expressed genes of the EA cohort in comparison of BC with and without diabetes.

SUPPLEMENTARY TABLE 6
Detailed information of missense data.

References

1. Rossi C, Cicalini I, Cufaro MC, Consalvo A, Upadhyaya P, Sala G, et al. Breast cancer in the era of integrating "omics" approaches. *Oncogenesis*. (2022) 11:17–3. doi: 10.1038/s41389-022-00393-8

2. Guo L, Kong D, Liu J, Zhan L, Luo L, Zheng W, et al. Breast cancer heterogeneity and its implication in personalized precision therapy. *Exp Hematol Oncol*. (2023) 12:3. doi: 10.1186/s40164-022-00363-1

3. Shahid RK, Ahmed S, Le D, Yadav S. Diabetes and Cancer: risk, challenges, management and outcomes. *Cancers*. (2021) 13:5735. doi: 10.3390/cancers13225735

4. Giovannucci E, Harlan DM, Archer MC, Bergenstal RM, Gapstur SM, Habel LA, et al. Diabetes and Cancer: A consensus report. *Diabetes Care*. (2010) 33:1674–85. doi: 10.2337/dc10-0666

5. Durrani IA, Bhatti A, John P. Integrated bioinformatics analyses identifying potential biomarkers for type 2 diabetes mellitus and breast cancer: in SIK1-ness and health. *PLoS One*. (2023) 18:e0289839–9. doi: 10.1371/journal.pone.0289839

6. Zhang Y-Y, Li Y-J, Xue C-D, Li S, Gao Z-N, Qin K-R. Effects of T2DM on cancer progression: pivotal precipitating factors and underlying mechanisms. *Front Endocrinol*. (2024) 15:1396022. doi: 10.3389/fendo.2024.1396022

7. Rojas A, Lindner C, Schneider I, Gonzàlez I, Araya H, Morales E, et al. Diabetes mellitus contribution to the remodeling of the tumor microenvironment in gastric cancer. *World J Gastrointest Oncol*. (2021) 13:1997–2012. doi: 10.4251/wjgo.v13.i12.1997

8. Samuel SM, Varghese E, Varghese S, Büsselberg D. Challenges and perspectives in the treatment of diabetes associated breast cancer. *Cancer Treat Rev*. (2018) 70:98–111. doi: 10.1016/j.ctrv.2018.08.004

9. Churko JM, Mantalas GL, Snyder MP, Wu JC. Overview of high throughput sequencing technologies to elucidate molecular pathways in cardiovascular diseases. *Circ Res*. (2013) 112:1613–23. doi: 10.1161/circresaha.113.300939

10. Warr A, Robert C, Hume D, Archibald A, Deeb N, Watson M. Exome sequencing: current and future perspectives. *G3 (Bethesda)*. (2015) 5:1543–50. doi: 10.1534/g3.115.018564

11. Verbist B, Klambauer G, Vervoort L, Talloen W, Shkedy Z, Thas O, et al. Using transcriptomics to guide lead optimization in drug discovery projects: lessons learned from the QSTAR project. *Drug Discov Today*. (2015) 20:505–13. doi: 10.1016/j.drudis.2014.12.014

12. Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T. Transcriptomics technologies. *PLoS Comput Biol*. (2017) 13:e1005457. doi: 10.1371/journal.pcbi.1005457

13. Goel A, Ward DG, Noyvert B, Yu M, Gordon NS, Abbotts B, et al. Combined exome and transcriptome sequencing of non-muscle-invasive bladder cancer: associations between genomic changes, expression subtypes, and clinical outcomes. *Genome Med*. (2022) 14:59. doi: 10.1186/s13073-022-01056-4

14. Codina-Solà M, Rodríguez-Santiago B, Homs A, Santoyo J, Rigau M, Aznar-Laín G, et al. Integrated analysis of whole-exome sequencing and transcriptome profiling in males with autism spectrum disorders. *Mol Autism*. (2015) 6:21. doi: 10.1186/s13229-015-0017-0

15. Nikas JB, Mitanis NT, Nikas EG. Whole exome and transcriptome RNA-sequencing model for the diagnosis of prostate Cancer. *ACS Omega*. (2019) 5:481–6. doi: 10.1021/acsomega.9b02995

16. Panigrahi G, Candia J, Dorsey TH, Tang W, Ohara Y, Byun JS, et al. Diabetes-associated breast cancer is molecularly distinct and shows a DNA damage repair deficiency. *JCI Insight*. (2023) 8:e170105. doi: 10.1172/jci.insight.170105

17. Zhao X, Jia C, Ji S, Lu K, Yang P, Wang Y. Development and validation of a metabolic syndrome and its components to predict the efficacy of neoadjuvant chemotherapy in breast cancer: an observational, single-center, cohort study. *Medicine*. (2025) 104:e41221–1. doi: 10.1097/md.0000000000041221

18. Cleveland RJ, North KE, Stevens J, Teitelbaum SL, Neugut AI, Gammon MD. The association of diabetes with breast cancer incidence and mortality in the Long Island breast Cancer study project. *Cancer Causes Control*. (2012) 23:1193–203. doi: 10.1007/s10552-012-9989-7

19. Lao C, Gurney J, Stanley J, Krebs J, Meredith I, Campbell I, et al. Association of diabetes and breast cancer characteristics at diagnosis. *Cancer Causes Control*. (2022) 34:103–11. doi: 10.1007/s10552-022-01654-y

20. 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

21. Davis S, Meltzer PS. GEOquery: a bridge between the gene expression omnibus (GEO) and BioConductor. *Bioinformatics*. (2007) 23:1846–7. doi: 10.1093/bioinformatics/btm254

22. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*. (2004) 3:Article3, 1–25. doi: 10.2202/1544-6115.1027

23. Yuan D, Ahamed A, Burgin J, Cummins C, Devraj R, Gueye K, et al. The European nucleotide archive in 2023. *Nucleic Acids Res*. (2023) 52:D92–7. doi: 10.1093/nar/gkad1067

24. Andrews S. Babraham bioinformatics – FastQC a quality control tool for high throughput sequence data. Babrahamacuk (2010). Available online at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. (Accessed February 7, 2025).

25. Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, et al. The complete sequence of a human genome. *Science*. (2022) 376:44–53. doi: 10.1126/science.abj6987

26. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*. (2009) 25:1754–60. doi: 10.1093/bioinformatics/btp324

27. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. (2009) 25:2078–9. doi: 10.1093/bioinformatics/btp352

28. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *GigaScience*. (2021) 10:giab008. doi: 10.1093/gigascience/giab008

29. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl variant effect predictor. *Genome Biol*. (2016) 17:122. doi: 10.1186/s13059-016-0974-4

30. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles G, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. (2013) 14:128. doi: 10.1186/1471-2105-14-128

31. Bendl J, Stourac J, Salanda O, Pavelka A, Wieben ED, Zendulka J, et al. PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS Comput Biol*. (2014) 10:e1003440. doi: 10.1371/journal.pcbi.1003440

32. Stone EA. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome Res*. (2005) 15:978–86. doi: 10.1101/gr.3804205

33. Capriotti E, Calabrese R, Casadio R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics*. (2006) 22:2729–34. doi: 10.1093/bioinformatics/btl423

34. Ramensky V. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res*. (2002) 30:3894–900. doi: 10.1093/nar/gkf493

35. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. (2010) 7:248–9. doi: 10.1038/nmeth0410-248

36. Ng PC. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res*. (2003) 31:3812–4. doi: 10.1093/nar/gkg509

37. Bromberg Y, Rost B. SNAP: predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Res*. (2007) 35:3823–35. doi: 10.1093/nar/gkm238

38. Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res*. (2005) 33:W306–10. doi: 10.1093/nar/gki375

39. Ben Chorin A, Masrati G, Kessel A, Narunsky A, Sprinzak J, Lahav S, et al. ConSurf-DB: an accessible repository for the evolutionary conservation patterns of the majority of PDB proteins. *Protein Sci*. (2019) 29:258–67. doi: 10.1002/pro.3779

40. Cotto KC, Feng Y, Ramu A, Richters MM, Freshour S, Skidmore ZL, et al. Integrated analysis of genomic and transcriptomic data for the discovery of splice-associated variants in cancer. *Nat Commun*. (2023) 14:1589. doi: 10.1038/s41467-023-37266-6

41. Du P, Fan R, Zhang N, Wu C, Zhang Y. Advances in integrated multi-omics analysis for drug-target identification. *Biomol Ther*. (2024) 14:692–2. doi: 10.3390/biom14060692

42. Jiang L, Zhang S, Jiang C, Chen H, Huang J, Yang J, et al. Integrative biomarker discovery and immune profiling for ulcerative colitis: a multi-methodological approach. *Sci Rep*. (2024) 14:24290. doi: 10.1038/s41598-024-75797-0

43. Hicks EM, Seah C, Cote A, Marchese S, Brennand KJ, Nestler EJ, et al. Integrating genetics and transcriptomics to study major depressive disorder: a conceptual framework, bioinformatic approaches, and recent findings. *Transl Psychiatry*. (2023) 13:129–14. doi: 10.1038/s41398-023-02412-7

44. San Lucas FA, Allenson K, Bernard V, Castillo J, Kim DU, Ellis K, et al. Minimally invasive genomic and transcriptomic profiling of visceral cancers by next-generation sequencing of circulating exosomes. *Ann Oncol*. (2016) 27:635–41. doi: 10.1093/annonc/mdv604

45. Hu Y, Zhang X, Wang O, Cui M, Li X, Wang M, et al. Integrated whole-exome and transcriptome sequencing of sporadic parathyroid adenoma. *Front Endocrinol*. (2021) 12:631680. doi: 10.3389/fendo.2021.631680

46. Seo JS, Lee JW, Kim A, Shin JY, Jung YJ, Lee SB, et al. Whole exome and transcriptome analyses integrated with microenvironmental immune signatures of lung squamous cell carcinoma. *Cancer Immunol Res*. (2018) 6:848–59. doi: 10.1158/2326-6066.cir-17-0453

47. Liu J, McCleland M, Stawiski EW, Gnad F, Mayba O, Haverty PM, et al. Integrated exome and transcriptome sequencing reveals ZAK isoform usage in gastric cancer. *Nat Commun*. (2014) 5:3830. doi: 10.1038/ncomms4830

48. Wang W, Hapach LA, Griggs L, Smart K, Wu Y, Taufalele PV, et al. Diabetic hyperglycemia promotes primary tumor progression through glycation-induced tumor extracellular matrix stiffening. *Sci Adv*. (2022) 8:eabo1673. doi: 10.1126/sciadv.abo1673

49. Okumura M, Yamamoto M, Sakuma H, Kojima T, Maruyama T, Jamali M, et al. Leptin and high glucose stimulate cell proliferation in MCF-7 human breast cancer cells: reciprocal involvement of PKC-α and PPAR expression. *Biochim Biophys Acta*. (2002) 1592:107–16. doi: 10.1016/s0167-4889(02)00276-8

50. Durrani IA, Bhatti A, John P. The prognostic outcome of "type 2 diabetes mellitus and breast cancer" association pivots on hypoxia-hyperglycemia axis. *Cancer Cell Int*. (2021) 21:351. doi: 10.1186/s12935-021-02040-5

51. Chou P, Choi HH, Huang Y, Fuentes-Mattei E, Velazquez-Torres G, Zhang F, et al. Impact of diabetes on promoting the growth of breast cancer. *Cancer Commun*. (2021) 41:414–31. doi: 10.1002/cac2.12147

52. Lero M, Shaw LM. Diversity of insulin and IGF signaling in breast cancer: implications for therapy. *Mol Cell Endocrinol*. (2021) 527:111213–3. doi: 10.1016/j.mce.2021.111213

53. Fountas A, Diamantopoulos L-N, Tsatsoulis A. Tyrosine kinase inhibitors and diabetes: A novel treatment paradigm? *Trends Endocrinol Metab*. (2015) 26:643–56. doi: 10.1016/j.tem.2015.09.003

54. Butti R, Das S, Gunasekaran VP, Yadav AS, Kumar D, Kundu GC. Receptor tyrosine kinases (RTKs) in breast cancer: signaling, therapeutic implications and challenges. *Mol Cancer*. (2018) 17:34. doi: 10.1186/s12943-018-0797-x

55. Edwards A, Brennan K. Notch Signalling in breast development and Cancer. *Front Cell Dev Biol*. (2021) 9:692173. doi: 10.3389/fcell.2021.692173

56. Bonegio R, Susztak K. Notch signaling in diabetic nephropathy. *Exp Cell Res*. (2012) 318:986–92. doi: 10.1016/j.yexcr.2012.02.036

57. Ardestani A, Maedler K. The hippo signaling pathway in pancreatic β-cells: functions and regulations. *Endocr Rev*. (2017) 39:21–35. doi: 10.1210/er.2017-00167

58. Yousefi H, Delavar MR, Piroozian F, Baghi M, Nguyen K, Cheng T, et al. Hippo signaling pathway: A comprehensive gene expression profile analysis in breast cancer. *Biomed Pharmacother*. (2022) 151:113144. doi: 10.1016/j.biopha.2022.113144

59. Liao H-Y, Da C-M, Wu Z-L, Zhang H-H. Ski: double roles in cancers. *Clin Biochem*. (2020) 87:1–12. doi: 10.1016/j.clinbiochem.2020.10.015

60. Diaz M, Martel N, Fitzsimmons RL, Eriksson NA, Cowin GJ, Thomas GP, et al. Ski overexpression in skeletal muscle modulates genetic programs that control susceptibility to diet-induced obesity and insulin signaling. *Obesity (Silver Spring)*. (2012) 20:2157–67. doi: 10.1038/oby.2012.101

61. Tabassum R, Chavali S, Mahajan A, Ghosh S, Madhu SV, Tandon N, et al. Association analysis of TNFRSF1B polymorphisms with type 2 diabetes and its related traits in North India. *Genomic Med*. (2008) 2:93–100. doi: 10.1007/s11568-009-9031-7

62. Xu F, Zhou G, Han S, Yuan W, Chen S, Fu Z, et al. Association of TNF-α, TNFRSF1A and TNFRSF1B gene polymorphisms with the risk of sporadic breast Cancer in northeast Chinese Han women. *PLoS One*. (2014) 9:e101138–8. doi: 10.1371/journal.pone.0101138

63. Zhu X, Xu M, Zhao X, Shen F, Ruan C, Zhao Y. The detection of plasma soluble Podoplanin of patients with breast Cancer and its clinical signification. *Cancer Manag Res*. (2020) 12:13207–14. doi: 10.2147/cmar.s281785

64. Huan F, Jiang X. Serum Podoplanin levels as a potential biomarker for diabetic nephropathy progression: A cross-sectional study. *Diabetes Metab Syndr Obes*. (2024) 17:4701–10. doi: 10.2147/DMSO.S500608

65. Roy N, Alencastro F, Roseman BA, Wilson S, Delgado ER, May MC, et al. Dysregulation of lipid and glucose homeostasis in hepatocyte-specific SLC25A34 knockout mice. *Am J Pathol*. (2022) 192:1259–81. doi: 10.1016/j.ajpath.2022.06.002

66. Zhao P, Jiang D, Huang Y, Chen C. EphA2: A promising therapeutic target in breast cancer. *J Genet Genomics*. (2021) 48:261–7. doi: 10.1016/j.jgg.2021.02.011

67. Konstantinova I, Nikolova G, OharaImaizumi M, Meda P, Kučera T, Zarbalis K, et al. EphAEphrinAMediated β cell communication regulates insulin secretion from pancreatic islets. *Cell*. (2007) 129:359–70. doi: 10.1016/j.cell.2007.02.044

68. Mahapatra KD, Pasquali L, Søndergaard JN, Lapins J, Nemeth IB, Baltás E, et al. A comprehensive analysis of coding and non-coding transcriptomic changes in cutaneous squamous cell carcinoma. *Sci Rep*. (2020) 10:3637. doi: 10.1038/s41598-020-59660-6

69. Lee J, Park B, Kim G, Kim K, Pak J, Kim K, et al. Arhgef16, a novel Elmo1 binding partner, promotes clearance of apoptotic cells via RhoG-dependent Rac1 activation. *Biochim Biophys Acta*. (2014) 1843:2438–47. doi: 10.1016/j.bbamcr.2014.07.006

70. Davegårdh C, García-Calzón S, Bacos K, Ling C. DNA methylation in the pathogenesis of type 2 diabetes in humans. *Mol Metab*. (2018) 14:12–25. doi: 10.1016/j.molmet.2018.01.022

71. Liu Y, Pan B, Qu W, Cao Y, Li J, Zhao H. Systematic analysis of the expression and prognosis relevance of FBXO family reveals the significance of FBXO1 in human breast cancer. *Cancer Cell Int*. (2021) 21:130. doi: 10.1186/s12935-021-01833-y

72. Liu B, Lu H, Li D, Xiong X, Gao L, Wu Z, et al. Aberrant expression of FBXO2 disrupts glucose homeostasis through ubiquitin-mediated degradation of insulin receptor in obese mice. *Diabetes*. (2016) 66:689–98. doi: 10.2337/db16-1104

73. Stolp J, Chen Y-G, Cox SL, Henck V, Zhang W, Tsaih S-W, et al. Subcongenic analyses reveal complex interactions between distal chromosome 4 genes controlling Diabetogenic B cells and CD4 T cells in nonobese diabetic mice. *J Immunol*. (2012) 189:1406–17. doi: 10.4049/jimmunol.1200120

74. McElwee JL, Mohanan S, Griffith OL, Breuer H, Anguish LJ, Cherrington BD, et al. Identification of PADI2 as a potential breast cancer biomarker and therapeutic target. *BMC Cancer*. (2012) 12:500. doi: 10.1186/1471-2407-12-500

75. Yang X-D, Xiang D-X, Yang Y-Y. Role of E3 ubiquitin ligases in insulin resistance. *Diabetes Obes Metab*. (2016) 18:747–54. doi: 10.1111/dom.12677

76. Wang Y, Dai J, Zeng Y, Guo J, Lan J. E3 ubiquitin ligases in breast Cancer metastasis: A systematic review of pathogenic functions and clinical implications. *Front Oncol*. (2021) 11:752604. doi: 10.3389/fonc.2021.752604

77. Pan X, Kaminga AC, Wen SW, Liu A. Chemokines in prediabetes and type 2 diabetes: A Meta-analysis. *Front Immunol*. (2021) 12:622438. doi: 10.3389/fimmu.2021.622438

78. Ozga AJ, Chow MT, Luster AD. Chemokines and the immune response to cancer. *Immunity*. (2021) 54:859–74. doi: 10.1016/j.immuni.2021.01.012

79. Andrews JL, Kim AC, Hens JR. The role and function of cadherins in the mammary gland. *Breast Cancer Res*. (2012) 14:203. doi: 10.1186/bcr3065

80. Okita T, Kita S, Fukuda S, Fukuoka K, Kawada-Horitani E, Iioka M, et al. Soluble T-cadherin promotes pancreatic β-cell proliferation by upregulating notch signaling. *iScience*. (2022) 25:105404. doi: 10.1016/j.isci.2022.105404

81. Akash MSH, Rehman K, Liaqat A. Tumor necrosis factor-alpha: role in development of insulin resistance and pathogenesis of type 2 diabetes mellitus. *J Cell Biochem*. (2017) 119:105–10. doi: 10.1002/jcb.26174

82. Sergi D, Melloni M, Passaro A, Neri LM. Influence of type 2 diabetes and adipose tissue dysfunction on breast Cancer and potential benefits from nutraceuticals inducible in microalgae. *Nutrients*. (2024) 16:3243–3. doi: 10.3390/nu16193243

83. Abdolvand M, Shahini Shams Abadi M, Soltani A, Banisharif F, Ghatrehsamani M. Chronic treatment with TNF-α, alone and in combination with Takinib, SB203580 and metformin induce cell death in breast cancer. *Heliyon*. (2023) 9:e21060. doi: 10.1016/j.heliyon.2023.e21060

84. Swaroop J, Naidu J, Rajarajeswari D. Association of TNF-α with insulin resistance in type 2 diabetes mellitus. *Indian J Med Res*. (2012) 135:127–30. doi: 10.4103/0971-5916.93435

85. Liu W, Lu X, Shi P, Yang G, Zhou Z, Li W, et al. TNF-α increases breast cancer stem-like cells through up-regulating TAZ expression via the non-canonical NF-κB pathway. *Sci Rep*. (2020) 10:1804. doi: 10.1038/s41598-020-58642-y

86. Sethi JK, Hotamisligil GS. Metabolic messengers: tumour necrosis factor. *Nat Metab*. (2021) 3:1302–12. doi: 10.1038/s42255-021-00470-z

87. Arcidiacono B, Iiritano S, Nocera A, Possidente K, Nevolo MT, Ventura V, et al. Insulin resistance and Cancer risk: an overview of the Pathogenetic mechanisms. *Exp Diabetes Res*. (2012) 2012:1–12. doi: 10.1155/2012/789174

88. Petersen MC, Shulman GI. Mechanisms of insulin action and insulin resistance. *Physiol Rev*. (2018) 98:2133–223. doi: 10.1152/physrev.00063.2017

89. Baker RG, Hayden MS, Ghosh S. NF-κB, inflammation, and metabolic disease. *Cell Metab*. (2011) 13:11–22. doi: 10.1016/j.cmet.2010.12.008

90. Wang X, Lin Y. Tumor necrosis factor and cancer, buddies or foes? *Acta Pharmacol Sin*. (2008) 29:1275–88. doi: 10.1111/j.1745-7254.2008.00889.x

91. Xia Y, Shen S, Verma IM. NF- B, an active player in human cancers. *Cancer Immunol Res*. (2014) 2:823–30. doi: 10.1158/2326-6066.cir-14-0112

92. Ribatti D, Tamma R, Annese T. Epithelial-mesenchymal transition in cancer: A historical overview. *Transl Oncol*. (2020) 13:100773. doi: 10.1016/j.tranon.2020.100773

93. Rohm TV, Meier DT, Olefsky JM, Donath MY. Inflammation in obesity, diabetes, and related disorders. *Immunity*. (2022) 55:31–55. doi: 10.1016/j.immuni.2021.12.013

94. Arfin S, Jha NK, Jha SK, Kesari KK, Ruokolainen J, Roychoudhury S, et al. Oxidative stress in Cancer cell metabolism. *Antioxidants*. (2021) 10:642. doi: 10.3390/antiox10050642

95. Wu Y, Zhou BP. TNF-α/NF-κB/snail pathway in cancer cell migration and invasion. *Br J Cancer*. (2010) 102:639–44. doi: 10.1038/sj.bjc.6605530

96. Li W, Zhang X, Sang H, Zhou Y, Shang C, Wang Y, et al. Effects of hyperglycemia on the progression of tumor diseases. *J Exp Clin Cancer Res*. (2019) 38:327. doi: 10.1186/s13046-019-1309-6

97. Zhang H-P, Jiang R-Y, Zhu J-Y, Sun K-N, Huang Y, Zhou H-H, et al. PI3K/AKT/mTOR signaling pathway: an important driver and therapeutic target in triple-negative breast cancer. *Breast Cancer*. (2024) 31:539–51. doi: 10.1007/s12282-024-01567-5

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 24 July 2025
DOI 10.3389/fmed.2025.1596788

Check for updates

OPEN ACCESS

EDITED BY
HaiHui Huang,
Shaoguan University, China

REVIEWED BY
Zheng Yuan,
China Academy of Chinese Medical Sciences, China
Raffaella Fiamma Cabini,
University of Ferrara, Ferrara, Italy

*CORRESPONDENCE
Jie Peng
✉ sank44@sina.com
Bing Lu
✉ lbgymaaaa@163.com

†These authors have contributed equally to this work

RECEIVED 20 March 2025
ACCEPTED 14 July 2025
PUBLISHED 24 July 2025

CITATION
Li Y, Zhang M, Hu Y, Zou D, Du B, Mo Y, He T, Zhao M, Li B, Xia J, Huang Z, Lu F, Lu B and Peng J (2025) A novel radiomics model combining GTVp, GTVnd, and clinical data for chemoradiotherapy response prediction in patients with advanced NSCLC.
Front. Med. 12:1596788.
doi: 10.3389/fmed.2025.1596788

COPYRIGHT
© 2025 Li, Zhang, Hu, Zou, Du, Mo, He, Zhao, Li, Xia, Huang, Lu, Lu and Peng. 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 novel radiomics model combining GTVp, GTVnd, and clinical data for chemoradiotherapy response prediction in patients with advanced NSCLC

Ya Li[1,2,3,4†], Min Zhang[1,2,3,4†], Yong Hu[5], Dan Zou[1], Bo Du[5], Youlong Mo[5], Tianchu He[6], Mingdan Zhao[7], Benlan Li[1,2,3,4], Ji Xia[1,2,3,4], Zhongjun Huang[1,2,3,4], Fangyang Lu[1], Bing Lu[2,3,4]* and Jie Peng[1,2,3,4]*

[1]Department of Oncology, The Second Affiliated Hospital of Guizhou Medical University, Kaili, China, [2]Department of Oncology, Affiliated Hospital of Guizhou Medical University, Guiyang, China, [3]Department of Oncology, Affiliated Cancer Hospital of Guizhou Medical University, Guiyang, China, [4]Division of Oncology, School of Clinical Medicine, Guizhou Medical University, Guiyang, China, [5]Department of Oncology, Guiyang Pulmonary Hospital, Guiyang, China, [6]Department of Oncology, Qiandongnan Prefecture People's Hospital, Kaili, China, [7]Department of Oncology, Qiannan Prefecture Hospital of Traditional Chinese Medicine, Duyun, China

Background: Numerous radiomic models have been developed to predict treatment outcomes in patients with NSCLC receiving chemotherapy and radiation therapy. However, computed tomography (CT) radiomic models that integrate the Gross Tumour Volume of the primary lesion (GTVp), the Gross Tumour Volume of nodal disease (GTVnd), and clinical information are relatively scarce and may offer greater predictive accuracy than models focusing on GTVp alone. This study aimed to evaluate the efficacy of a CT radiomic model combining GTVp, GTVnd, and clinical data for predicting treatment response in unresectable stage III–IV NSCLC patients undergoing concurrent chemoradiotherapy.

Methods: A total of 101 patients with unresectable stage III–IV NSCLC were included. GTVp was delineated using lung windows, and GTVnd was delineated using mediastinal windows. Radiological features were extracted using Python 3.6, then subjected to F-test and Lasso regression for feature selection. Logistic regression was performed on the selected radiological features. Clinical information was analysed with univariate and multivariate logistic regression to identify significant clinical variables. Five models were developed and evaluated, incorporating GTVp, GTVnd, and clinical data.

Results: The GTVp-based radiomics model achieved an area under the curve (AUC) of 0.855 in the training cohort and 0.775 in the validation cohort. The multimodal composite model (integrating GTVp, GTVnd, and clinical parameters) significantly outperformed the GTVp-only model, with a training AUC of 0.862 and validation AUC of 0.863, demonstrating superior predictive performance for concurrent chemoradiotherapy response in this patient population.

KEYWORDS
NSCLC, radiomics, chemoradiotherapy, GTVp, GTVnd

1 Introduction

Lung cancer has a high incidence and mortality rate, with an estimated five-year survival of only around 23% (1). It is classified into non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC) based on pathological features, with NSCLC accounting for approximately 85% of cases (2). For patients with inoperable stage III–IV NSCLC, concurrent chemoradiotherapy (CCRT) is a vital treatment approach (3). However, treatment sensitivity varies among individuals (4, 5), affecting prognosis. Notably, the response to cancer therapy is closely linked to prognosis. Notably, patients who respond more favourably to therapy often experience longer progression-free and overall survival then those with poorer responses (6–8).

Imaging remains the primary method for tumour evaluation in clinical practice (9), and radiomics has emerged as a non-invasive, effective tool for prognostic prediction (10–14). Several radiological models have been developed to predict treatment response and outcomes in patients with NSCLC undergoing CCRT (15–17). Approximately 60% of patients with NSCLC present with advanced or locally advanced disease at diagnosis (18), often because of late detection of non-specific symptoms (19), which can lead to mediastinal lymph node metastasis. In such cases, radiation oncologists typically delineate the Gross Tumour Volume of the primary lesion (GTVp) and nodal disease (GTVnd) for chest radiation therapy. However, when extracting CT radiomic features, many researchers focus solely on GTVp while overlooking GTVnd (20, 21). This omission is notable because pre- and post-treatment changes in GTVnd are equally critical for tumour staging (22). Moreover, prior research has shown that combining mediastinal window CT images with lung window CT images can improve both the malignancy of a nodule and its potential indolence (23, 24). Thus, incorporating GTVnd CT images may be crucial for assessing CCRT efficacy.

Despite the demand for multimodal biomarkers in NSCLC management, no prior study has simultaneously integrated CT radiomics features of GTVp (lung window) and GTVnd (mediastinal window) with clinical parameters to predict CCRT response. Therefore, this study aims to develop and validate a composite model, specifically evaluating its performance in predicting short-term CCRT efficacy among patients with unresectable stage III-IV NSCLC.

2 Methods

The study received approval from the Ethics Committee of the Second Affiliated Hospital of Guizhou Medical University (SAHGMU; approval number 2020-LS-03) and was conducted in strict accordance with the Declaration of Helsinki. Informed consent was obtained from all participants.

Figure 1 presents the study flowchart. The inclusion criteria were: (1) pathologically confirmed NSCLC; (2) no surgical indications; (3) no prior therapies (including neoadjuvant chemotherapy, interventional therapy, immunotherapy, or targeted therapy) before CCRT; (4) stage III or IV disease with confirmed mediastinal lymph node metastasis (N2/N3) based on the 8th edition UICC Tumor-Node-Metastasis staging system; (5) availability of standard contrast-enhanced chest CT images obtained within 1 month before and 3 months after treatment completion; and (6) receipt of conventional fractionated radiotherapy (target dose: 60–66 Gy/30–33\u00B0F, intensity-modulated radiotherapy) combined with chemotherapy. For squamous cell carcinoma, weekly paclitaxel plus cisplatin was used, whereas for non-squamous cell carcinoma, pemetrexed was administered every 3 weeks alongside cisplatin (25). The exclusion criteria were: (1) concomitant malignancies, (2) incomplete or poor-quality CT images, and (3) insufficient follow-up data.

This multicentre retrospective study enrolled patients from two distinct cohorts: (1) 77 patients treated at SAHGMU; (2) 24 patients from three regional hospitals (Guiyang Pulmonary Hospital, Qiandongnan People's Hospital, Qiannan Traditional Chinese Medicine Hospital). All cases were recruited consecutively between January 2019 and July 2023. Treatment outcomes were categorized as complete response (CR), partial response (PR), stable disease (SD), or progressive disease (PD) according to RECIST 1.1 (26). Patients with CR or PR were classified into the treatment-sensitive group, while those with SD and PD were classified as treatment-insensitive.

Chest contrast-enhanced CT images were preprocessed using MATLAB 2014b[1] with: (1) Spatial normalization: Rigid registration to the INHALE chest CT atlas via ANTs (v2.3.3) using mutual information; (2) Isotropic resampling: Resampling normalized images to 1 mm isotropic voxels using B-spline interpolation. Following the guidelines of ICRU 83 (27), a radiation oncologist with 10 years of experience in lung cancer treatment delineated the GTVp and GTVnd without access to patient information. ITK-SNAP (version 3.8.0; http://www.itksnap.org) was used to manually label slices layer-by-layer (28). GTVp was delineated in the lung window (WW 1600 HU, WL − 600 HU), and GTVnd in the mediastinal window (WW 250 HU, WL 50 HU). The criteria for defining GTVnd included: (1) short-axis diameter ≥1 cm, (2) presence of ≥3 clustered lymph nodes within a single station, (3) pathological confirmation of metastasis in mediastinal lymph nodes (in select patients), or (4) PET-CT $SUV_{max} > 2.5$ in the region (in select patients). After completing the annotations were completed, the region of interest (ROI) was designated. For each patient, 1,834 radiological features were extracted from the ROIs. These features were standardized using the Z-score and then screened by an F-test in ANOVA, where F is defined as the ratio of between-group variance to within-group variance. To avoid overfitting, LASSO regression with 10-fold cross-validation (via glmnet in R) was performed on each training subset to select the λ minimizing mean square error. Only features selected in ≥80% of folds were retained for the final model. Finally, logistic regression was used to construct the radiological models.

Clinical data—including sex, ethnicity, age, smoking history, pathological type, tumour stage, and haematological markers measured 1 week before treatment (such as carcinoembryonic antigen,

1 https://ww2.mathworks.cn/

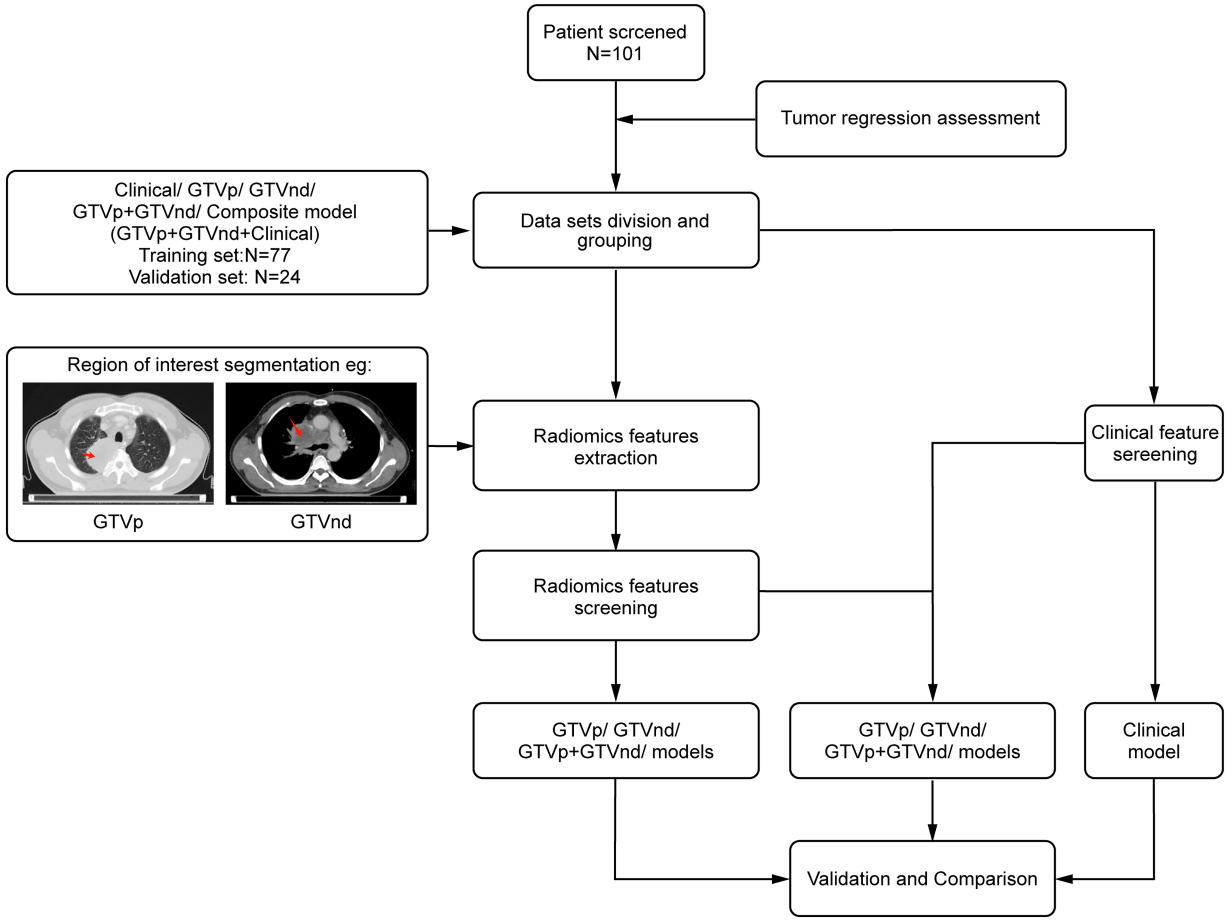


FIGURE 1
Experimental flowchart.

neuron-specific enolase [NSE], white blood cell count, haemoglobin, and platelet levels)—were collected and initially analysed via univariate regression. Factors with $p < 0.05$ underwent multivariate regression, and variables remaining significant ($p < 0.05$) were incorporated into a clinical prediction model built through logistic regression. PyRadiomics was used for radiomic feature extraction (v3.0.1; https://github.com/radiomics/pyradiomics) (29). Statistical modeling was conducted in R (v3.5.1; https://www.r-project.org/). SPSS (v26.0, IBM Corp., Armonk, NY, USA) handled descriptive statistics.

Combination models were constructed using logistic regression with selected radiological and clinical features. Model performance was evaluated through Receiver Operating Characteristic (ROC) curves, Area Under the Curve (AUC), accuracy, precision, recall, and Decision Curve Analysis (DCA). Statistical significance was defined as $p < 0.05$ for all hypothesis tests.

3 Results

A total of 101 participants met the inclusion criteria. Patients were recruited from the SAHGMU ($n = 77$), Guiyang Pulmonary Hospital ($n = 13$), Qiandongnan Prefecture People's Hospital ($n = 5$), Qiannan Prefecture Traditional Chinese Medicine Hospital ($n = 6$). Table 1 shows the clinical information. Guizhou—an ethnically diverse province in southwest China—is home to all four treatment centres included in this study. The principal ethnic groups were Han (39.60%), Miao (29.70%), and Dong (25.74%). The training cohort and external validation cohort exhibited comparable treatment efficacy rates ($p > 0.05$). Table 2 presents the relationship between clinical features and CCRT treatment sensitivity. After screening, only haemoglobin was significantly correlated with CCRT treatment sensitivity. However, as shown in Table 3, the haemoglobin-based clinical model underperformed among the models, with an AUC of 60.65% in the training set and 65.00% in the validation set.

Following the F-test and Lasso regression feature selection, six radiomic features were selected for GTVp (lung window) and four for GTVnd (mediastinal window). Figure 2 and Table 4 illustrate the distribution of these selected features. The predictive performance of the radiological models is shown in Figure 3 and Table 3. In the training set, the composite model—incorporating GTVp, GTVnd, and clinical features—achieved the highest AUC (0.862). The second-ranked model was the GTVp-only model (AUC: 0.855), followed by the GTVp + GTVnd combination (AUC: 0.853). The GTVnd-only model yielded the lowest performance (AUC: 0.734). In the external validation set, the composite model again demonstrated the highest accuracy (AUC: 0.863). The GTVp + GTVnd combination ranked

TABLE 1 Baseline characteristics of patients.

Variables	Categories	Total (n = 101)	Training (n = 77)	External validation (n = 24)	P
Sex, n (%)	Female	17 (16.83)	15 (19.48)	2 (8.33)	0.583
	Male	84 (83.17)	62 (80.52)	22 (91.67)	
Age, n (%)	≤50	17 (16.83)	15 (19.48)	2 (8.33)	0.336
	>50	84 (83.17)	62 (80.52)	22 (91.67)	
Ethnicity, n (%)	Miao	30 (29.70)	27 (35.06)	3 (12.50)	0.005
	Dong	26 (25.74)	22 (28.57)	4 (16.67)	
	Han	40 (39.60)	23 (29.87)	17 (70.83)	
	Others	5 (4.95)	5 (6.49)	0 (0.00)	
Efficacy, n (%)	CR/PR	28 (27.72)	24 (31.17)	4 (16.67)	0.166
	SD/PD	73 (72.28)	53 (68.83)	20 (83.33)	
Histology, n (%)	LUSC	65 (64.36)	46 (59.74)	19 (79.17)	0.238
	LUAD	31 (30.69)	26 (33.77)	5 (20.83)	
	Other	5 (4.95)	5 (6.49)	0 (0.00)	
TNM, n (%)	III	67 (66.34)	53 (68.83)	14 (58.33)	0.342
	IV	34 (33.66)	24 (31.17)	10 (41.67)	
CEA, n (%)	Normal	59 (58.42)	45 (58.44)	14 (58.33)	0.993
	Elevated	42 (41.58)	32 (41.56)	10 (41.67)	
NSE, n (%)	Normal	72 (71.29)	55 (71.43)	17 (70.83)	0.955
	Elevated	29 (28.71)	22 (28.57)	7 (29.17)	
WBC, n (%)	Reduced	3 (2.97)	3 (3.90)	0 (0.00)	0.548
	Normal	86 (85.15)	66 (85.71)	20 (83.33)	
	Elevated	12 (11.88)	8 (10.39)	4 (16.67)	
Hb, n (%)	Reduced	71 (70.30)	53 (68.83)	18 (75.00)	0.564
	Normal	30 (29.70)	24 (31.17)	6 (25.00)	
PLT, n (%)	Reduced	5 (4.95)	3 (3.90)	2 (8.33)	0.567
	Normal	89 (88.12)	69 (89.61)	20 (83.33)	
	Elevated	7 (6.93)	5 (6.49)	2 (8.33)	

LUSC, Lung squamous cell carcinoma; LUAD, Lung adenocarcinoma; CEA, carcinoembryonic antigen; NSE, neuron specific enolase; WBC, White blood cell; Hb, Hemoglobin; PLT, Platelet; Alb, Albumin.

second (AUC: 0.800), the GTVp-only model placed third (AUC: 0.775), and the GTVnd-only model performed poorest (AUC: 0.375).

The DeLong test on the external validation set ROC data (Table 5) showed no statistically significant difference between the composite model and the conventional GTVp model (p = 0.14). Considering the limited sample size of the validation cohort (n = 24), we conducted clinical decision curve analysis to evaluate real-world utility. As shown in Figure 4, the composite model provided a superior net benefit across threshold probabilities compared to both the conventional clinical model and the GTVp model.

4 Discussion

In this study, our radiomic models outperformed the clinical factor model in predicting treatment outcomes. At present, the most commonly used guideline for tumour evaluation is RECIST 1.1; however, metabolic changes in tumour cells induced by chemotherapy and radiation therapy may become apparent earlier than morphological changes (30, 31). While radiation and chemotherapeutic agents effectively inhibit tumour cell proliferation, their structural impact can manifest slowly, making it difficult to detect short-term treatment effects through conventional imaging. Unlike RECIST 1.1, radiomics extracts pre-treatment data from the tumour, thus enabling an earlier assessment of treatment sensitivity before therapy is complete.

Among the 101 patients analysed, decreased haemoglobin emerged as the only clinical feature associated with CCRT sensitivity. Haemoglobin is critical for oxygen transport to tissues. When haemoglobin levels are low, increased anoxia in tumour cells leads to reduced sensitivity to radiotherapy and chemotherapy, ultimately weakening the therapeutic effect (32). In our patient population, over 70% presented with low haemoglobin levels prior to treatment. This could be explained by several factors. First, dietary habits among middle-aged and elderly individuals in Guizhou, who tend to eat more vegetables than meat, can result in insufficient iron intake and anaemia. Second, compromised immunity in cancer patients elevates

TABLE 2 Clinical model: clinical features related to CCRT sensitivity.

Variables	Univariate analysis		Multivariate analysis	
	OR (95%CI) *P*		OR (95%CI) *P*	
Sex				
Female	1.00 (Reference)			
Male	0.77 (0.23 ~ 2.60)	0.673		
Age (years)				
≤50	1.00 (Reference)			
>50	0.13 (0.02 ~ 1.05)	0.055		
Ethnicity				
Miao	1.00 (Reference)			
Dong	2.10 (0.61 ~ 7.23)	0.239		
Han	1.17 (0.42 ~ 3.22)	0.766		
Others	2.00 (0.20 ~ 20.33)	0.558		
Histology				
LUSC	1.00 (Reference)			
LUAD	0.59 (0.23 ~ 1.50)	0.270		
Other	1.31 (0.14 ~ 12.55)	0.817		
TNM				
III	1.00 (Reference)			
IV	3.03 (1.04 ~ 8.88)	0.043		
CEA				
Normal	1.00 (Reference)			
Elevated	1.40 (0.57 ~ 3.46)	0.459		
NSE				
Normal	1.00 (Reference)			
Elevated	0.51 (0.20 ~ 1.28)	0.149		
WBC				
Reduced	1.00 (Reference)			
Normal	0.00 (0.00 ~ Inf)	0.991		
Elevated	0.00 (0.00 ~ Inf)	0.991		
Hb				
Normal	1.00 (Reference)		1.00 (Reference)	
Reduced	2.85 (1.14 ~ 7.16)	0.025	2.85 (1.14 ~ 7.16)	0.025
PLT				
Reduced	1.00 (Reference)			
Normal	4.57 (0.72 ~ 29.14)	0.108		
Elevated	2.00 (0.19 ~ 20.61)	0.560		

OR: Odds Ratio, CI: Confidence Interval; LUSC, Lung squamous cell carcinoma; LUAD, Lung adenocarcinoma; CEA, carcinoembryonic antigen; NSE, neuron specific enolase; WBC, White blood cell; Hb, Hemoglobin; PLT, Platelet.

their risk of secondary infections, which may lead to the excessive destruction of red blood cells. Third, acute and chronic bleeding (e.g., haemoptysis) often associated with lung cancer can further exacerbate anaemia in these patients.

Although the validation set showed that the GTVnd radiomics model alone had a relatively poor predictive performance (AUC: 0.375) compared to the GTVp model (AUC: 0.775), these findings indicate that, in standard CT-based radiomics models for stage III - IV lung cancer, primary tumour features may be more influential than those of metastatic mediastinal lymph nodes. In our study, the radiological features of metastatic mediastinal lymph node lesions sensitive to CCRT all originated from "wavelets" a phenomenon that warrants further inquiry. Moreover, the absence of comprehensive PET/CT scans or mediastinal lymph node biopsies in some patients may have limited the precision of GTVnd delineation, as radiation oncologists relied solely on conventional imaging criteria (e.g., short

TABLE 3 Performance of the models.

Model	Accuracy	Precision	Recall	F1-score	AUC
Clinical					
Training set	68.83%	68.83%	100.00%	81.54%	60.65%
Validation set	83.33%	83.33%	100.00%	90.91%	65.00%
GTVp					
Training set	83.12%	85.71%	90.57%	88.07%	85.53%
Validation set	79.17%	82.61%	95.00%	88.37%	77.50%
GTVnd					
Training set	79.22%	80.33%	92.45%	85.96%	73.43%
Validation set	83.33%	86.36%	85.00%	90.48%	37.50%
GTVp + GTVnd					
Training set	83.12%	84.48%	92.45%	88.29%	85.30%
Validation set	83.33%	83.33%	100.00%	90.91%	80.00%
Composite model (GTVp + GTVnd + clinical)					
Training set	83.12%	84.48%	92.45%	88.29%	86.16%
Validation set	83.33%	83.33%	100.00%	90.91%	86.25%

Data in parentheses are 95% CIs. AUC, area under the curve; GTVp, gross tumor volume of the primary lesion; GTVnd, gross tumor volume of nodal disease.

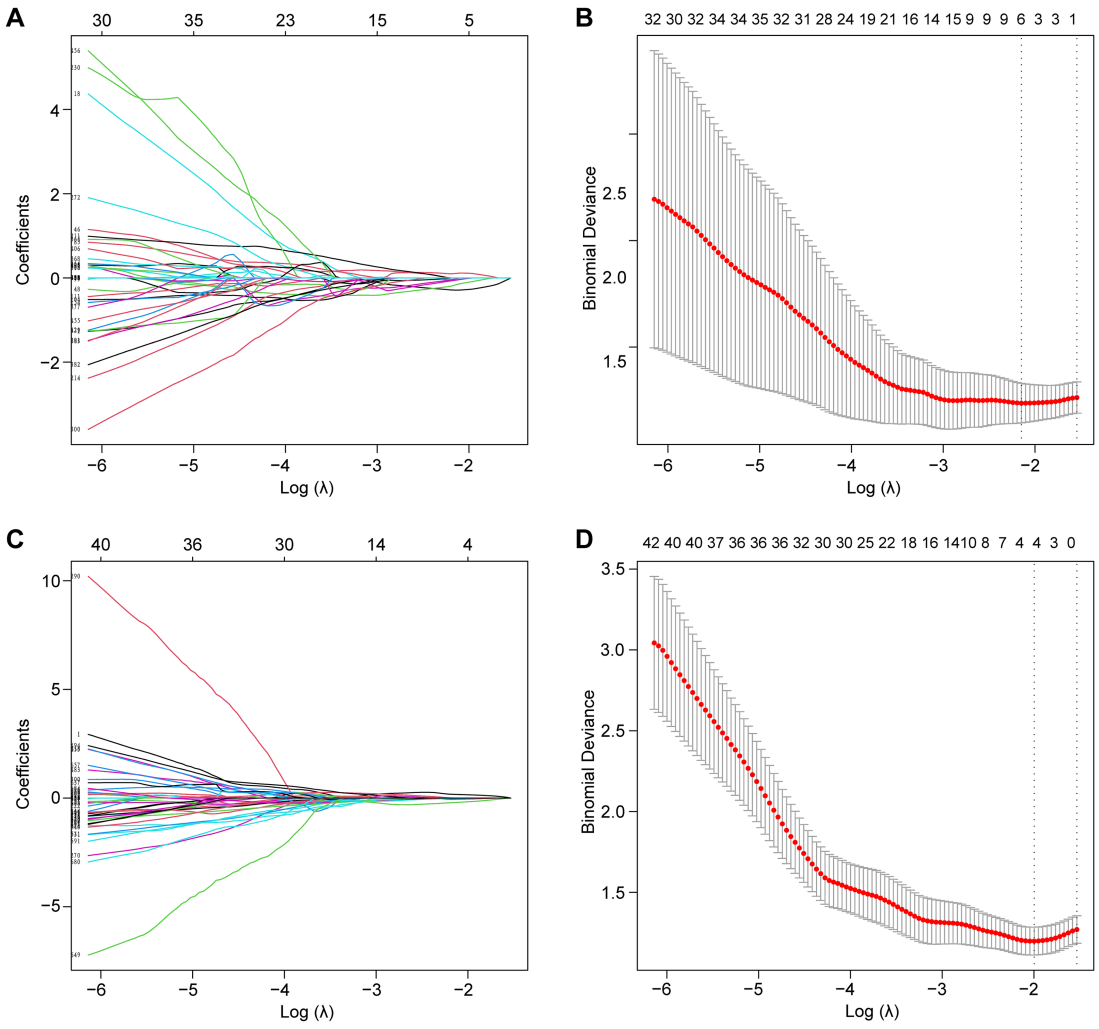


FIGURE 2
Lasso regression was employed to screen radiological features. **(A)** LASSO coefficient curve for the GTVp group. **(B)** Cross-validation curve for the GTVp group. **(C)** LASSO coefficient curve for the GTVnd group. **(D)** Cross-validation curve for the GTVnd group.

diameter ≥1 cm or at least three clustered lymph nodes in one region), potentially resulting in a reduced diagnostic rate for positive mediastinal lymph nodes.

We also noted that integrating clinical features with radiological data led to superior predictive performance compared to radiological models alone. The radiomics model combining GTVp and GTVnd (AUC: 0.800) outperformed the individual GTVp and GTVnd models. We compared our model not only to our own previous models but also to similar studies, such as: 1. A 2022 study that used a radiomics nomogram based solely on CT-derived GTVp and clinical features to predict chemoradiotherapy efficacy in locally advanced non-small cell lung cancer, with a training set C-index of 0.796 and a validation set C-index of 0.756 (17); 2. A 2023 study developed a radiomics model based on CT-derived GTVp to predict concurrent chemoradiotherapy in patients with locally advanced non-small cell lung cancer. The study reported that the AUC for the GTV reduction (Criteria A) model was 0.767, while the AUC for the RECIST 1.1 standard (Criteria B) model was 0.771 (16). In contrast, our composite model (GTVp + GTVnd + clinical characteristics) achieved higher AUCs in both the training set (0.862) and the validation set (0.863). Further analysis revealed that the GTVnd features added critical information: (1) "wavelet. LHL_firstorder_10Percentile" quantifies low-intensity pixels in regions with vertical textural detail; (2) "wavelet. LHL_glcm_Contrast" captures roughness/heterogeneity of vertical textures and sensitivity to directional structures; (3) "wavelet. HLH_glszm_SizeZoneNonUniformityNormalized" indicates lesion size heterogeneity; (4) "wavelet. LLL_firstorder_InterquartileRange" stably quantifies slow-varying grayscale distribution in anatomical structures. The inclusion of these GTVnd radiomic features enhanced the model's efficacy.

In conclusion, our composite model (AUC = 0.863) demonstrated notably better performance than the conventional GTVp model (AUC = 0.775), indicating that including GTVnd radiological features can significantly improve the predictive capacity of CT-based models for CCRT outcomes. Decision curve analysis further confirmed that the composite model provided higher accuracy than the GTVp model alone,

TABLE 4 Selected radiological features.

GTVp	GTVnd
lbp.3D.k_glszm_GrayLevelNonUniformityNormalized	wavelet. LHL_firstorder_10Percentile
lbp.3D.k_glrlm_RunLengthNonUniformityNormalized	wavelet. LHL_glcm_Contrast
original_shape_Sphericity	wavelet. HLH_glszm_SizeZoneNonUniformityNormalized
square_glcm_Imc2	wavelet. LLL_firstorder_InterquartileRange
squareroot_glcm_Correlation	
exponential_glrlm_RunLengthNonUniformity	

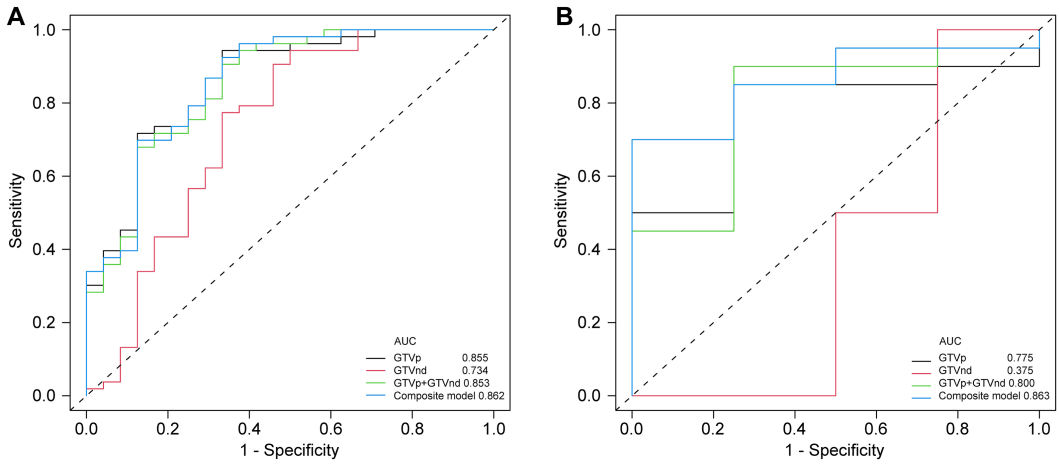


FIGURE 3
Comparison of ROC curves for different models. **(A)** ROC curves of different radiomic models in the training cohort. **(B)** ROC curves of different radiomic models in the validation cohorts.

TABLE 5 DeLong test for AUC values of the validation set.

Model	Clinical	GTVp	GTVnd	GTVp+GTVnd	Composite model
Clinical	1	0.59	0.04	0.55	0.26
GTVp	0.59	1	0.23	0.57	0.14
GTVnd	0.04	0.23	1	0.23	0.09
GTVp + GTVnd	0.55	0.57	0.23	1	0.40
Composite model	0.26	0.14	0.09	0.40	1

GTVp, gross tumor volume of the primary lesion; GTVnd, gross tumor volume of nodal disease.

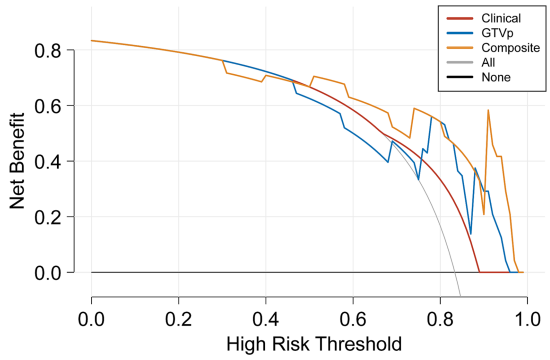


FIGURE 4
Decision curve analysis of the models.

highlighting the importance of incorporating additional radiomic features and clinical data in treatment response predictions. This study is the first to show that CT-based radiomic models integrating GTVnd, GTVp, and clinical information can meaningfully enhance CCRT response prediction in unresectable stage III–IV NSCLC. By extracting a broader range of radiomic features, the composite model offers a more comprehensive assessment of the tumour's biological characteristics, potentially facilitating more individualized cancer treatment strategies. Overall, our findings emphasize the importance of including GTVnd in CT imaging analyses, reinforcing the need for a holistic approach to tumour evaluation.

Despite these promising results, our study has several limitations. First, the use of various CT scanners across four different institutions may have introduced variability in imaging parameters. To reduce this effect, all CT scans were normalized and reconstructed into 1-mm slices. Second, not all patients underwent PET/CT or mediastinal lymph node biopsies, potentially impacting the precision of GTVnd delineation. Previous research indicates that PET/CT is more accurate than conventional CT for detecting malignant lymph nodes (33, 34). Consequently, future research should incorporate PET/CT or biopsy before CCRT to better define GTVnd and improve model accuracy. Third, a single radiation oncologist performed all ROI delineations, restricting our ability to assess inter-observer consistency in radiomic feature extraction. Fourth, due to a relatively small sample size, larger studies are necessary to validate these findings.

5 Conclusion

This study demonstrates that a CT-based model integrating GTVp, GTVnd, and clinical data surpasses the conventional GTVp radiological model in predicting CCRT efficacy for patients with unresectable stage III–IV NSCLC. Such an approach may allow for earlier adjustments to treatment regimens for patients expected to have less favourable outcomes.

Data availability statement

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

Ethics statement

The studies involving humans were approved by the Second Affiliated Hospital of Guizhou Medical University Ethics Committee. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

YL: Conceptualization, Data curation, Investigation, Writing – original draft, Writing – review & editing. MZhang: Conceptualization, Data curation, Investigation, Writing – original draft, Writing – review & editing. YH: Data curation, Writing – original draft, Investigation. DZ: Data curation, Investigation, Writing – original draft. BD: Data curation, Investigation, Writing – original draft. YM: Data curation, Investigation, Writing – original draft. TH: Data curation, Investigation, Writing – original draft. MZhao: Data curation, Investigation, Writing – original draft. BeL: Investigation, Writing – original draft. JX: Investigation, Writing – original draft. ZH: Investigation, Writing – original draft. FL: Investigation, Writing – original draft, Resources, Supervision. BiL: Investigation, Supervision, Writing – original draft, Conceptualization, Data curation, Methodology. JP: Conceptualization, Data curation, Supervision, Writing – original draft, Funding acquisition, Project administration, Resources, Software, Validation, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the Qian Dong Nan Science and Technology Program (No. qdnkhJz [2023] 14), Scientific Research Project of Guizhou Provincial Health and Wellness Commission (No. gzwkj2024-099 and gzwkj2025-608), Public Hospital High-Quality Development Research Public Welfare Project Fund (No. GL-A014), and Spark Program (No. XHJH-0048).

Acknowledgments

We thank all members for their invaluable contributions.

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.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

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

References

1. Rebecca LS, Kimberly DM, Nikita Sandeep W, Ahmedin J. Cancer statistics, 2023. *CA Cancer J Clin*. (2023) 73:17–48. doi: 10.3322/caac.21763

2. Molina JR, Yang PG, Cassivi SD, Schild SE, Adjei AA. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clin Proc*. (2008) 83:584–94. doi: 10.1016/S0025-6196(11)60735-0

3. Maconachie R, Mercer T, Navani N, McVeigh G. Lung cancer: diagnosis and management: summary of updated NICE guidance. *BMJ-Brit Med J*. (2019) 364:l1049. doi: 10.1136/bmj.l1049

4. Zhivotovsky B, Joseph B, Orrenius S. Tumor Radiosensitivity and apoptosis. *Exp Cell Res*. (1999) 248:10–7. doi: 10.1006/excr.1999.4452

5. Zheng WY, Xin Y, Siyang W, Xiaonan W, Qiutao W, Wenhao C, et al. Instability mechanism of Osimertinib in plasma and a solving strategy in the pharmacokinetics study. *Front Pharmacol*. (2022) 13:928983. doi: 10.3389/fphar.2022.928983

6. Park C, Chu HH, Kim JH, Kim SY, Alrashidi I, Gwon DI, et al. Clinical significance of the initial and best responses after chemoembolization in the treatment of intermediate-stage hepatocellular carcinoma with preserved liver function. *J Vasc Interv Radiol*. (2020) 31:1998. doi: 10.1016/j.jvir.2020.04.017

7. Pointer KB, Katipally RR, Bestvina CM, Juloori A, Partouche J, Patel JD, et al. Evaluation of initial metastatic tumor location and radiation response to determine outcomes in patients who received combination stereotactic body radiotherapy and immunotherapy for NSCLC. *Int J Radiat Oncol Biol Phys*. (2021) 111:e449. doi: 10.1016/j.ijrobp.2021.07.1266

8. Tsurugai Y, Kozuka T, Ishizuka N, Oguchi M. Relationship between the consolidation to maximum tumor diameter ratio and outcomes following stereotactic body radiotherapy for stage I non-small-cell lung cancer. *Lung Cancer*. (2016) 92:47–52. doi: 10.1016/j.lungcan.2015.12.003

9. Kurland BF, Gerstner ER, Mountz JM, Schwartz LH, Ryan CW, Graham MM, et al. Promise and pitfalls of quantitative imaging in oncology clinical trials. *Magn Reson Imaging*. (2012) 30:1301–12. doi: 10.1016/j.mri.2012.06.009

10. Peng J, Kang S, Ning Z, Deng H, Shen J, Xu Y, et al. Residual convolutional neural network for predicting response of transarterial chemoembolization in hepatocellular carcinoma from CT imaging. *Eur Radiol*. (2020) 30:413–24. doi: 10.1007/s00330-019-06318-1

11. Jiang X, Zhao H, Saldanha OL, Nebelung S, Kuhl C, Amygdalos I, et al. An MRI deep learning model predicts outcome in rectal Cancer. *Radiology*. (2023) 307:e222223. doi: 10.1148/radiol.222223

12. Peng J, Huang J, Huang G, Zhang J. Predicting the initial treatment response to Transarterial chemoembolization in intermediate-stage hepatocellular carcinoma by the integration of Radiomics and deep learning. *Front Oncol*. (2021) 11:730282. doi: 10.3389/fonc.2021.730282

13. Peng J, Lu F, Huang J, Zhang J, Gong W, Hu Y, et al. Development and validation of a pyradiomics signature to predict initial treatment response and prognosis during transarterial chemoembolization in hepatocellular carcinoma. *Front Oncol*. (2022) 12:853254. doi: 10.3389/fonc.2022.853254

14. Bera K, Braman N, Gupta A, Velcheti V, Madabhushi A. Predicting cancer outcomes with radiomics and artificial intelligence in radiology. *Nat Rev Clin Oncol*. (2022) 19:132–46. doi: 10.1038/s41571-021-00560-7

15. Chen W, Wang L, Hou Y, Li L, Chang L, Li Y, et al. Combined Radiomics-clinical model to predict radiotherapy response in inoperable stage III and IV non-small-cell lung Cancer. *Technol Cancer Res Treat*. (2022) 21:15330338221142400. doi: 10.1177/15330338221142400

16. Zhou C, Hou L, Tang X, Liu C, Meng Y, Jia H, et al. CT-based radiomics nomogram may predict who can benefit from adaptive radiotherapy in patients with local advanced-NSCLC patients. *Radiother Oncol*. (2023) 183:109637. doi: 10.1016/j.radonc.2023.109637

17. Chen X, Tong X, Qiu Q, Sun F, Yin Y, Gong G, et al. Radiomics nomogram for predicting Locoregional failure in locally advanced non-small cell lung Cancer treated with definitive Chemoradiotherapy. *Acad Radiol*. (2022) 29:S53–61. doi: 10.1016/j.acra.2020.11.018

18. Meza R, Meernik C, Jeon J, Cote ML. Lung cancer incidence trends by gender, race and histology in the United States, 1973-2010. *PLoS One*. (2015) 10:e0121323. doi: 10.1371/journal.pone.0121323

19. Wilk AM, Kozłowska E, Borys D, D'Amico A, Fujarewicz K, Gorczewska I, et al. Radiomic signature accurately predicts the risk of metastatic dissemination in late-stage non-small cell lung cancer. *Transl Lung Cancer Res*. (2023) 12:1372–83. doi: 10.21037/tlcr-23-60

20. 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) 48:7946–58. doi: 10.1002/mp.15302

21. Gong J, Bao X, Wang T, Liu J, Peng W, Shi J, et al. A short-term follow-up CT based radiomics approach to predict response to immunotherapy in advanced non-small-cell lung cancer. *Onco Targets Ther*. (2022) 11:2028962. doi: 10.1080/2162402X.2022.2028962

22. Asamura H, Chansky K, Crowley J, Goldstraw P, Rusch VW, Vansteenkiste JF, et al. The International Association for the Study of Lung Cancer lung Cancer staging project: proposals for the revision of the N descriptors in the forthcoming 8th edition of the TNM classification for lung Cancer. *J Thoracic Oncol*. (2015) 10:1675–84. doi: 10.1097/JTO.0000000000000678

23. Nasir M, Farid MS, Suhail Z, Khan MH. Optimal thresholding for multi-window computed tomography (CT) to predict lung cancer. *Appl Sci*. (2023) 13:256. doi: 10.3390/app13127256

24. Lu H, Mu W, Balagurunathan Y, Qi J, Abdalah MA, Garcia AL, et al. Multi-window CT based Radiomic signatures in differentiating indolent versus aggressive lung cancers in the National Lung Screening Trial: a retrospective study. *Cancer Imaging*. (2019) 19:45. doi: 10.1186/s40644-019-0232-6

25. Ettinger DS, Wood DE, Aisner DL, Akerley W, Bauman JR, Bharat A, et al. NCCN guidelines® insights: non-small cell lung Cancer, version 2.2023. *J Natl Comprehens Cancer Network*. (2023) 21:340–50. doi: 10.6004/jnccn.2023.0020

26. 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). *Europ J Cancer*. (2009) 45:228–47. doi: 10.1016/j.ejca.2008.10.026

27. Grégoire V, Mackie TR. State of the art on dose prescription, reporting and recording in intensity-modulated radiation therapy (ICRU report no. 83). *Cancer Radiother*. (2011) 15:555–9. doi: 10.1016/j.canrad.2011.04.003

28. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *NeuroImage*. (2006) 31:1116–28. doi: 10.1016/j.neuroimage.2006.01.015

29. 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:e104–7. doi: 10.1158/0008-5472.CAN-17-0339

30. Yang Y, Tian W, Su L, Li P, Gong X, Shi L, et al. Tumor-infiltrating cytotoxic T cells and tumor-associated macrophages correlate with the outcomes of neoadjuvant Chemoradiotherapy for locally advanced rectal Cancer. *Front Oncol*. (2021) 11:743540. doi: 10.3389/fonc.2021.743540

31. Wang HM, Wu MH, Chang PH, Lin HC, Liao CD, Wu SM, et al. The change in circulating tumor cells before and during concurrent chemoradiotherapy is associated with survival in patients with locally advanced head and neck cancer. *Head Neck*. (2019) 41:2676–87. doi: 10.1002/hed.25744

32. Topkan E, Selek U, Ozdemir Y, Yildirim BA, Guler OC, Mertsoylu H, et al. Chemoradiotherapy-induced hemoglobin nadir values and survival in patients with stage III non-small cell lung cancer. *Lung Cancer*. (2018) 121:30–6. doi: 10.1016/j.lungcan.2018.04.016

33. Ventura E, Islam T, Gee MS, Mahmood U, Braschi M, Harisinghani MG. Detection of nodal metastatic disease in patients with non-small cell lung cancer: comparison of positron emission tomography (PET), contrast-enhanced computed tomography (CT), and combined PET-CT. *Clin Imaging*. (2010) 34:20–8. doi: 10.1016/j.clinimag.2009.03.012

34. Al-Sarraf N, Gately K, Lucey J, Wilson L, McGovern E, Young V. Lymph node staging by means of positron emission tomography is less accurate in non-small cell lung cancer patients with enlarged lymph nodes: analysis of 1,145 lymph nodes. *Lung Cancer*. (2008) 60:62–8. doi: 10.1016/j.lungcan.2007.08.036

frontiers | Frontiers in Medicine

TYPE Review
PUBLISHED 01 September 2025
DOI 10.3389/fmed.2025.1629621

Check for updates

OPEN ACCESS

EDITED BY
HaiHui Huang,
Shaoguan University, China

REVIEWED BY
Dola Sundeep,
Indian Institute of Information Technology Design and Manufacturing, India
Qiang Zhang,
University of Texas MD Anderson Cancer Center, United States

*CORRESPONDENCE
Bhuvic Patel
✉ bhuvic.patel@wustl.edu

RECEIVED 16 May 2025
ACCEPTED 13 August 2025
PUBLISHED 01 September 2025

CITATION
Pugazenthi S, Pari SS, Zhang Z, Silverstein J, Kim AH and Patel B (2025) The evolution and application of multi-omic analysis for pituitary neuroendocrine tumors.
Front. Med. 12:1629621.
doi: 10.3389/fmed.2025.1629621

COPYRIGHT
© 2025 Pugazenthi, Pari, Zhang, Silverstein, Kim and Patel. 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 evolution and application of multi-omic analysis for pituitary neuroendocrine tumors

Sangami Pugazenthi[1,2], Shree S. Pari[1], Ziyan Zhang[1], Julie Silverstein[3,4], Albert H. Kim[1,4,5] and Bhuvic Patel[1,4,5]*

[1]Taylor Family Department of Neurological Surgery, Washington University School of Medicine, St. Louis, MO, United States, [2]Department of Neurological Surgery, University of Pittsburgh Medical Center, Pittsburgh, PA, United States, [3]Division of Endocrinology, Metabolism and Lipid Research, Washington University School of Medicine, St. Louis, MO, United States, [4]WashU Medicine Pituitary Center, Washington University School of Medicine, St. Louis, MO, United States, [5]The Brain Tumor Center, Siteman Cancer Center, Washington University School of Medicine, St. Louis, MO, United States

Pituitary neuroendocrine tumors (PitNETs) are a heterogeneous group of intracranial neoplasms that vary in hormonal activity, histological features, and clinical behavior. The rise of high-throughput sequencing and molecular profiling technologies has enabled multiomic approaches—including genomics, transcriptomics, epigenomics, proteomics, and metabolomics—to deepen our understanding of PitNET pathogenesis. These studies have identified key mutations, transcriptional lineages, epigenetic modifications, and proteomic features that contribute to tumor subtype classification, invasiveness, and treatment response. Integrative multi-omic analyses have further revealed distinct molecular subtypes, complex regulatory networks, and molecular profiles that can predict recurrence and therapeutic efficacy. These approaches hold strong potential for advancing personalized medicine in PitNETs, supporting patient-specific diagnosis, prognostication, and therapeutic strategies. Future directions include the application of emerging -omic technologies and the development of robust computational tools to integrate and translate multi-layered data into clinically actionable insights.

KEYWORDS

pituitary, PitNET, multiomics, molecular sequencing, transcriptomics, genomics, epigenomics, proteomics

Introduction

Pituitary tumors represent a diverse group of neoplasms that originate from the endocrine cells of the pituitary gland and account for approximately 17.8% of all intracranial tumors (1). Historically termed *pituitary adenomas*, these tumors have been considered largely benign and indolent. However, this perception has evolved significantly with advances in molecular pathology and clinical characterization. In 2022, the World Health Organization (WHO) officially reclassified these tumors as *pituitary neuroendocrine tumors* (PitNETs) to better reflect their neuroendocrine origin and biological spectrum (2). A pivotal aspect of the new WHO classification is the use of pituitary-specific transcription factors (TFs) to define tumor lineage more accurately than traditional hormonal immunostaining alone. The key TFs include pituitary-specific positive transcription factor 1 (*PIT1*), steroidogenic factor 1 (*SF1*), and T box transcription factor (*TPIT*), which correspond to the somatotroph/lactotroph/thyrotroph, gonadotroph, and corticotroph lineages, respectively (Figure 1). This molecular stratification helps distinguish morphologically similar but biologically distinct subtypes, thereby enhancing diagnostic precision and prognostic estimation (2). Despite these advances, the clinical management of PitNETs remains challenging due to a lack of robust biomarkers for tumor

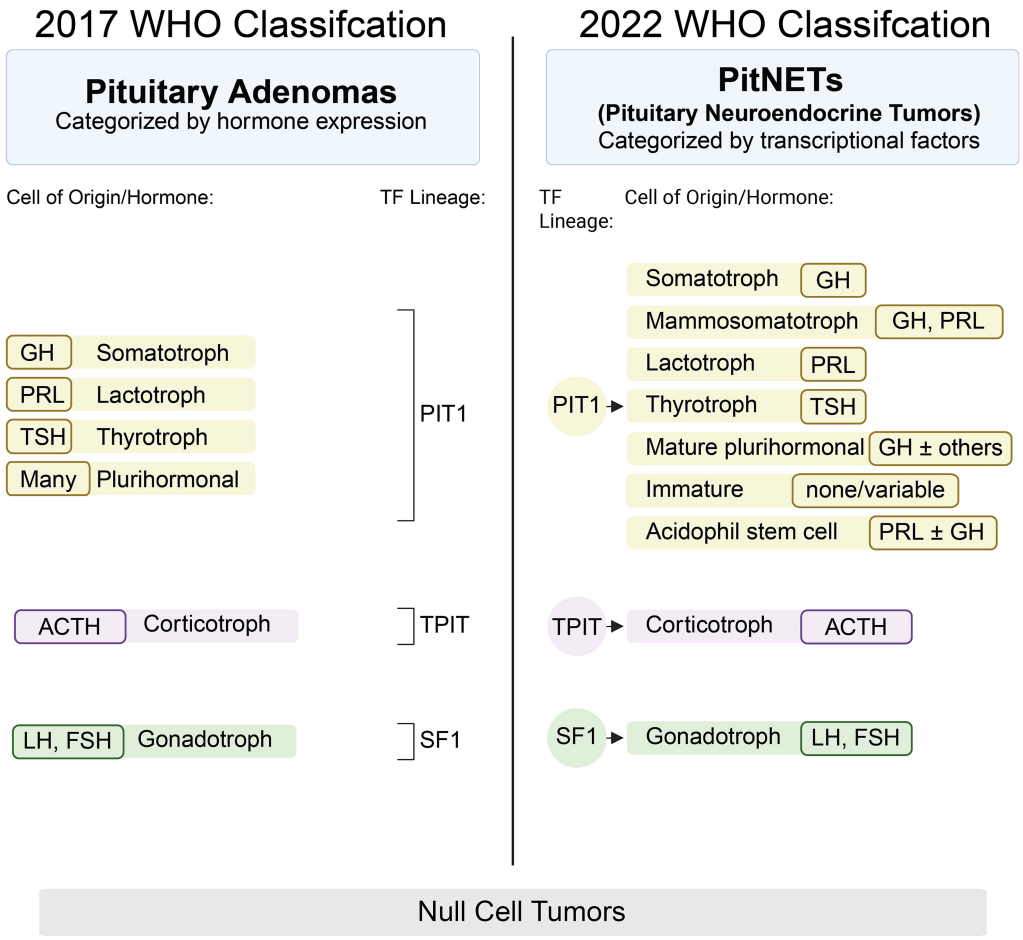


FIGURE 1
Comparison of 2017 and 2022 PitNET WHO classification schemes. PitNET, Pituitary neuroendocrine tumor; WHO, World Health Organization.

aggressiveness, treatment response, and recurrence risk. In this context, multiomic approaches—including genomic, transcriptomic, epigenomic, proteomic, and metabolomic profiling—offer powerful tools to dissect the complexity of PitNETs. Integrative multiomic analysis can provide a systems-level understanding of tumor biology, identify molecular subgroups, and uncover novel targets for therapy and early detection (3). This review summarizes the current landscape and emerging insights from multiomic studies in PitNETs, emphasizing their potential to revolutionize classification, prognosis, and individualized treatment strategies in pituitary tumor management.

Genomic analysis

Genomic analyses have played a crucial role in uncovering the molecular underpinnings of PitNETs, shedding light on both sporadic and hereditary forms. Although PitNETs display a relatively low mutational burden compared to other solid tumors, several recurrent somatic and germline alterations have been identified that contribute to tumor initiation, hormonal dysregulation, and progression.

The most well-characterized somatic mutations in PitNETs are subtype specific. Guanine nucleotide-binding protein, alpha stimulating (*GNAS*) mutations are frequently found in somatotroph tumors, promoting cyclic adenosine monophosphate (*cAMP*) signaling and growth hormone (GH) overproduction. PitNETs with *GNAS* mutations have been associated with smaller size and decreased invasiveness (4). In corticotroph tumors causing Cushing's Disease, ubiquitin carboxyl-terminal hydrolase 8 (*USP8*) mutations are present in up to 40% of cases and result in impaired degradation of epidermal growth factor receptor (*EGFR*), enhancing adrenocorticotropic hormone (ACTH) secretion and cellular proliferation (5, 6). Other mutations described in corticotroph PitNETs include ubiquitin specific peptidase 48 (*USP48*), B-Raf proto-oncogene, serine/threonine kinase (*BRAF*), and tumor protein p53 (*TP53*) (7, 8). Despite these discoveries, most PitNETs lack recurrent driver mutations, suggesting a significant role for epigenetic regulation, chromosomal instability, and post-transcriptional mechanisms in tumor biology. A subset of PitNETs arise in the context of hereditary tumor syndromes, most notably Multiple Endocrine Neoplasia type 1 (MEN1), caused by inactivating mutations in the *MEN1* gene, which encodes the tumor suppressor menin. Other inherited mutations involve cyclin-dependent kinase inhibitor 1B (*CDKN1B*) (associated with *MEN4*), aryl hydrocarbon receptor interacting protein (*AIP*), and succinate dehydrogenase (*SDHx*) (9–12). Patients harboring AIP mutations most commonly present with somatotropinomas, often at a younger age, with larger tumors and more growth hormone (GH) secretion (13, 14). Succinate

dehydrogenase complex iron sulfur subunit B/D (*SDHB/D*) mutations have been shown to be associated with combined paragangliomas, pheochromocytomas, and less frequently PitNETs, suggesting shared tumorigenesis pathways related to mitochondrial metabolism (12).

With the development of next-generation sequencing (NGS), including whole-exome sequencing (WES) and whole-genome sequencing (WGS), studies utilizing these methods provided a broader landscape of mutational events in PitNETs. In 2016, Song et al. examined the somatic mutational landscape of 125 PitNETs, identifying low mutational burden, confirming the presence of previously described mutations such as *GNAS*, *MEN1*, and *USP8*, identifying novel mutations such as kinesin heavy chain isoform 5A (*KIF5A*) and growth factor receptor-bound protein 10 (*GRB10*), and determining that 18% of tumors harbor copy number alterations (CNAs). Gene ontology analysis revealed that plurihormonal, GH-, prolactin (PR)-, and ACTH-secreting PitNETs were enriched for somatic mutations in overlapping molecular pathways as were TSH- and LH/FSH-secreting PitNETs (15). Subsequently, Bi et al. identified that 29% of PitNETs have CNAs, but novel somatic alterations in genes were infrequent and often non-recurrent. They found that the tumors with more disrupted genomes (higher CNA burden) were more likely to be functional PitNETs or null cell tumors compared to PitNETs with less disrupted genomes, which were more likely nonfunctional (16). Large-scale sequencing efforts continue to uncover novel candidate genes and low-frequency variants that may contribute to tumor biology but integration of genomic data with transcriptomic and epigenomic profiles is essential to elucidate the mechanistic impact of these mutations, and inclusion of phenotypic data is critical for clinical relevance.

To facilitate clinical interpretation, Table 1 summarizes key PitNET biomarkers identified across multiomic studies, specifically highlighting their functional roles, prognostic value, and therapeutic relevance. Even though many markers remain investigational, this framework may inform future biomarker-guided therapy trials.

Transcriptional profiling

Transcriptomic profiling using techniques such as bulk and single-cell RNA sequencing has emerged as a powerful approach to characterize PitNETs beyond histology and hormonal output, offering insights into their functional identity, heterogeneity, and aggressiveness. Unlike genomic alterations, which are relatively infrequent in PitNETs, transcriptional changes are widespread and reflect both lineage commitment and tumor behavior.

Transcriptomic profiling has had a significant impact on the field of pituitary tumors as this method was used to discover the relevance of TFs in the classification of PitNETs highlighted in the 2022 WHO guidelines. The use of transcription factors has been shown to be more reliable than previous methods using histology, immunochemistry, *in situ* hybridization, and hormone expression to identify and classify these tumors (2). The biological role of *PIT1*, *SF1* and *TPIT* in normal pituitary gland development and PitNET pathogenesis has also been investigated using bulk RNA sequencing (17, 18). In normal corticotroph development, *TPIT* along with paired like homeodomain 1 (*PITX1*) activate the proopiomelanocortin (*POMC*) gene (19, 20). On the other hand, suppression of *TPIT* causes pituitary neuroendocrine cells to differentiate into gonadotroph or thyrotroph cells (21). The *PIT1* TF lineage is positively regulated by paired-like homeobox 1 (*PROP1*) and negatively regulated by HESX homeobox 1 (*HESX1*) (22, 23). Each hormonal subtype of *PIT1* PitNETs have specific mechanisms through which *PIT1* is involved in pathogenesis. Gonadotrophs are part of the *SF1*-lineage of PitNETs; *SF1* transcription in part relies on the binding of estrogen-to-estrogen receptor alpha, which mediates chromatin remodeling of the *SF1* locus (24).

Invasive PitNETs have significant differences in their transcriptional profiles compared to noninvasive tumors, including differentially expressed genes related to the Nuclear Factor-kappa B (*NF-κB*) and antitumoral immune response (25, 26). Invasive prolactinomas exhibited significantly different transcriptional profiles compared to noninvasive prolactinomas (27). Compared to noninvasive corticotrophs, invasive corticotroph tumors exhibit upregulation of cyclin D2 (*CCND2*) and zinc finger protein 676 (*ZNF676*) and downregulation of death-associated protein kinase 1 (*DAPK1*) and tissue inhibitor of metalloproteinase 2 (*TIMP2*) (28). Additionally, in corticotroph tumors, RNA-sequencing showed a decrease in RNA expression of secreted frizzled-related protein 2 (*SFRP2*), which may promote tumorigenesis by upregulating *Wnt*

TABLE 1 Known biomarkers prognostic/therapeutic utility.

Biomarker	Subtype(s)	Alteration type	Functional role	Prognostic relevance	Therapeutic significance
GNAS	Somatotroph	Activating mutation	↑ cAMP signaling → GH hypersecretion	Smaller, less invasive tumors	Somatostatin analog sensitivity
USP8	Corticotroph	Gain-of-function mutation	↑ EGFR stability → ↑ ACTH secretion	Less aggressive, lower recurrence	EGFR inhibitors (experimental)
SF3B1	Lactotroph	Spliceosome mutation	Aberrant mRNA splicing	Potentially linked to aggressiveness	Still being investigated
HMGA2	Lactotroph	Overexpression/ epigenetic activation	Chromatin remodeling	Associated with invasiveness	HDAC inhibitors (preclinical)
TERT methylation	Multiple	Promoter methylation	Telomerase activation	Conflicting; may indicate poor prognosis	Still being investigated
ID2	Corticotroph, Lactotroph	Protein overexpression	EMT regulation	Linked to invasiveness	Potential EMT targeting (preclinical)

signaling (29). The transcriptional profile of lactotroph tumors showed activation of estrogen receptor signaling, oxidative phosphorylation signaling, and eukaryotic translation initiation factor (*EIF*) signaling. Network analysis of upstream regulators determined that potential pathogenic drivers may include early growth response 1 (*EGR1*), protein kinase cAMP-activated catalytic subunit Alpha (*PRKACA*), paired like homeodomain 2 (*PITX2*), cAMP responsive element binding protein 1 (*CREB1*), and Jun D (*JUND*) proto-oncogene, an AP-1 transcription factor subunit (30).

In addition to evaluating specific genes, pathways, and PitNET types, transcriptomic data has been used to cluster PitNETs based on molecular subtype. Consensus clustering of transcriptomic data from 117 PitNETs of all hormonal subtypes revealed three molecular subtypes of tumors defined by biological processes: Group I – signaling pathways, Group II – metabolic processes, and Group III – immune responses. Each group had different immune profiles, and Group III had the worst prognosis even though these tumors were smaller (31). Future investigation of the role of non-coding, long non-coding, micro, and circulating RNAs in PitNET biology represents a new frontier for transcriptional profiling of PitNETs (32).

Single cell RNA sequencing (scRNA-seq) has also been used to investigate biological pathways related to invasive PitNETs. Previous work has shown that silent corticotroph PitNETs have been associated with an invasive phenotype; scRNA-seq revealed that these tumors express epithelial to mesenchymal transition genes, which may be driving tumor invasion (20). scRNA-seq has also been utilized to more robustly identify the heterogeneous biology of PitNETs. For example, when analyzing tumor cells from *PIT1*-lineage tumors, expression of hormone-encoding genes represented the majority of variation between tumors. There were four major clusters of non-*PIT-1* tumor cells, and of the three clusters with majority *TPIT*-lineage tumor cells, one had significantly elevated Granzyme K (*GZMK*) expression, suggesting a possible novel subtype of corticotroph tumor. The fourth cluster of non-*PIT-1* tumor cells was predominantly composed of *SF-1* lineage cells with overexpression of follicle stimulating hormone subunit beta (*FSHB*). Additionally, within the tumor microenvironment, two distinct tumor-associated macrophage (TAM) clusters were enriched in PitNETs, one with pro-inflammatory M1 features and the other with immunosuppressive M2 marker upregulation (*SPP1*, *TREM2*, and *CX3CR1*). This finding suggests that depletion of TAMs or macrophage repolarization may be therapeutically relevant in PitNET treatment. In addition, stress response pathways were upregulated in T cells, suggesting functional exhaustion. This finding suggests that certain PitNET subtypes may be responsive to immune checkpoint blockade and other relevant tumor microenvironment modulating therapies (33).

Through the integration of scRNA-seq and single cell genomic sequencing, transcriptional profiles of normal endocrine cells (gonadotrophs, somatotrophs, and lactotrophs) to cognate tumor cells revealed several tumor-related genes such as adhesion molecule with Ig like domain 2 (*AMIGO2*), zinc finger protein 36 (*ZFP36*), BTG anti-proliferation factor 1 (*BTG1*), and disks large MAGUK scaffold protein 5 (*DLG5*) (34). Although 62% of tumors harbored CNAs, there was no significant intratumoral CNA heterogeneity (34). Although single cell molecular analyses have been utilized extensively to reveal the underlying biology and microenvironment of several cancer types and central nervous system tumors, there are only a few robust studies analyzing PitNETs at a single cell resolution. Further work in this area will likely lead to a more sophisticated understanding of PitNET tumorigenesis, especially with regard to differences between hormonal subtypes, tumor microenvironment, the immune landscape, and molecular drivers.

Epigenetic profiling

While genomic mutations in PitNETs are relatively uncommon, epigenetic dysregulation influencing gene expression, hormonal activity, and tumor behavior has emerged as a critical mechanism of PitNET pathogenesis (35). Epigenetic changes—such as in DNA methylation, histone modifications, and chromatin remodeling—are key modulators of transcriptional activity and cellular identity in both normal pituitary cells and tumors (35). Indeed, the activity of lineage-specific transcription factors such as *PIT1*, *SF1*, and *TPIT* is modulated by epigenetic marks, and clustering of PitNETs profiled by methylation array separated tumors by TF lineage (36).

Many studies have reported epigenetic changes in numerous genes associated with cell growth, cell signaling, and cell cycle signaling, including cyclin dependent kinase 1 (*CDK1*), cyclin dependent kinase inhibitor 1B (*CDKN1B*), cyclin dependent kinase inhibitor 2A (*CDKN2A*), cyclin dependent kinase inhibitor 2C (*CDKN2C*), growth arrest and DNA damage inducible gamma (*GADD45G*), Ras association domain family member 1 (*RASSF1A*), Ras association domain family member 3 (*RASSF3*), *DAPK*, pituitary tumor transforming gene 1 (*PTTG1*), maternally expressed 3 (*MEG3*), and fibroblast growth factor receptor 2 (*FGFR2*) (37–51). More aggressive PitNETs, defined by larger size and invasiveness, have been associated with the overexpression of DNA methyltransferases 1/3A (*DNMT1*/3A) and promoter hypermethylation of tumor suppressor genes (52). The first genome-wide methylation analysis of PitNETs in 2012 identified differentially methylated genes in nonfunctioning, GH-, and PRL-secreting PitNETs. Specifically, HHIP like 1 (*HHIPL1*) and transcription factor AP-2 epsilon (*TFAP2E*) were hypermethylated in nonfunctioning tumors (53). Multiple studies have shown that these nonfunctional tumors have global hypermethylation compared to hormonally active tumors (53–55). However, *invasive* nonfunctioning tumors have more hypomethylated cytosine-phosphate guanine (CpGs) sites compared to *noninvasive* nonfunctioning tumors (54), reminiscent of the global hypomethylation observed in many cancers (56). Biological pathways that were differentially methylated between invasive and noninvasive PitNETs included homophilic cell adhesion, cell–cell adhesion, and biological adhesion. The Polypeptide N-acetylgalactosaminyltransferase 9 (*GALNT9*) promoter was also found to be methylated with corresponding decreased RNA expression in invasive tumors, making *GALNT9* expression a potential therapeutic target (55).

Although telomerase reverse transcriptase (*TERT*) promoter mutation is a marker of aggressiveness in numerous cancers and central nervous system tumors, the role of *TERT* promoter alterations such as methylation has been debated in PitNETs. In 2018, a study with 101 patients found no relationship between *TERT* promoter mutation or methylation and outcomes in patients with PitNETs (57). However, in a 2019 study analyzing 70 patients, *TERT* promoter methylation was associated with disease progression and shorter progression free survival (58, 59). Other common epigenetic

biomarkers in brain tumors such as glioma include O6-methylguanine-DNA methyltransferase (*MGMT*) promoter methylation, which is related to response to temozolomide (TMZ) therapy. In contrast, in PitNETs the relationship between *MGMT* methylation status and prognosis or response to TMZ remains controversial (60–63).

Despite the ongoing debate surrounding prognostic epigenetic biomarkers like *MGMT* in PitNETs, the broader role of the epigenetic machinery itself presents a compelling target for therapeutic intervention. Importantly, DNA methyltransferase (DNMT) inhibitors and histone deacetylase (HDAC) inhibitors have demonstrated efficacy in other central nervous system tumors like glioblastoma and may be clinically relevant for the treatment of aggressive PitNETs (64, 65). While not yet clinically validated in PitNETs, DNMT and HDAC inhibitor therapies could be particularly beneficial when conventional therapies fail. Preclinical PitNET models will be essential in determinng whether modulation of the epigenetic landscape can suppress tumor proliferation, reduce hormonal hypersecretion, or enhance sensitivity to standard treatments such as temozolomide. As we further study PitNET epigenetics, targeted manipulation of regulators such as DNMTs and HDACs may emerge as a viable therapeutic strategy within a precision medicine framework.

Proteomic analysis

Proteomic analysis provides a direct readout of the functional state of cells by quantifying proteins and their post-translational modifications. In PitNETs, proteomic analyses offer unique insights into tumor activity, cellular heterogeneity, and treatment response.

Advanced mass spectrometry (MS)-based techniques, including tandem MS and data-independent acquisition (DIA), have enabled high-throughput profiling of PitNET proteomes and post-translational modifications. MS analysis reveals that nonfunctioning PitNETs have 2,000–6,000 differentially expressed proteins compared to normal pituitary glands (66, 67). Proteomic methods have also been used to identify the role of phosphorylation of proteins in nonfunctioning PitNETs. For example, phosphorylation of β-catenin at Serine552 is associated with aggressive disease characterized by invasion and recurrence (68). Meanwhile, comparison of nonfunctioning tumors to normal pituitary glands revealed 595 differentially phosphorylated proteins associated with biological pathways such as the spliceosome pathway, RNA transport pathway, and proteoglycans in cancer (69). Ubiquitination is another post-translational modification that has been investigated in PitNET biology. Ubiquitinated proteins in PitNETs were most involved in biological pathways such as the Phosphatidylinositol 3-kinase/protein kinase B (PI3K/AKT) pathway, Hippo (Hpo) pathway, ribosome signaling pathway, and nucleotide excision repair (70).

Alterations of specific protein abundances and functions have been investigated to identify their role in tumorigenesis in PitNETs. For example, hematopoietic cell signal transducer 1 (Hint1) is a protein marker that was found to have high expression in invasive PitNETs, especially those that expressed vascular endothelial growth factor (VEGF) and fetal liver kinase 1 (Flk1) (71). Invasive tumors were also found to have higher expression of cluster of differentiation 206 (CD206), a M2-macrophage marker, compared to noninvasive tumors based on immunohistochemical staining (72). Several protein components of the Notch pathway were altered in prolactinomas, in addition to increased expression of PIT1 and survival factor phosphoprotein associated with glycosphingolipid-enriched microdomains 1 (PAG1) and decreased expression of E-cadherin and N-cadherin (73).

Nitroproteomics is a subfield of proteomics that specifically studies nitropeptides and nitroproteins, which are often markers of oxidative damage and can be associated with tumorigenesis. In studies investigating nitroproteins in PitNETs, several nitroproteins and other proteins that interact with nitroproteins in nonfunctioning PitNETs were discovered using a nitrotyrosine affinity column (NTAC) (74, 75). Analysis of nitroproteins is important since identification of post-translational modifications such as nitrosylation may suggest potential new avenues for targeted therapy (76). Further work to identify the extent of the role of nitroproteomics in PitNET biology and tumorigenesis is warranted.

Metabolomics

Metabolomics—the comprehensive profiling of small-molecule metabolites in biological samples—provides a dynamic snapshot of cellular metabolism and its interaction with the tumor microenvironment. In PitNETs, metabolomic analysis has begun to uncover metabolic adaptations associated with hormone synthesis, tumor growth, and treatment resistance (77). Metabolomic methods such as matrix-assisted laser desorption/ionization (MALDI) mass spectrometry imaging have been used to confirm excess hormone production and classify PitNETs within 30 min (78). In patients with Cushing's disease, biomarkers such as pyridoxate, deoxycholic acid, and 3-methyladipate were altered in plasma samples (79). Urine metabolites were analyzed using gas chromatography mass spectrometry system in prolactinoma patients, which showed an elevation of urinary 17-ketosteroids and all estrogen metabolite concentrations, as well as the ratios of delta 5/delta 4-steroids and 5 beta/5 alpha- hydrogensteroids (80). These findings have implications for understanding tumor biology, the systemic effect of disease, and identification of measurable biomarkers. For instance, PitNETs are defined by a distinct metabolic profile with higher succinic and lactic acid (72). These finding suggest possible mechanisms of disease development and progression as well as identification of biomarkers for diagnosis and targeted therapy. Although still an emerging field in pituitary tumor research, metabolomics holds significant promise for identifying biomarkers and therapeutic vulnerabilities, particularly in combination with other -omic methods.

Integrative Multiomic analysis

The advent of high-throughput -omics technologies has revolutionized our understanding of PitNETs, enabling comprehensive analyses at multiple molecular levels. These technologies each offer distinct advantages and limitations in terms of resolution, sensititivty, sample input, cost, and use-case. Table 2 provides a comparative overview of commonly used technologies across omics layers in an effort highlight pragmatic and methodological constraints across PitNET research. Integrative -omic analysis provides a holistic view of the molecular landscape

TABLE 2 Omics technology comparison table.

Omics layer	Technology	Resolution	Noise/ Artifacts	Sample input	Cost	Use case in PitNETs
Genomics	Whole-Exome Sequencing (WES)	Coding regions only	Misses non-coding mutations	Low (DNA only)	Lower	Detects recurrent mutations (e.g., GNAS, USP8)
	Whole-Genome Sequencing (WGS)	Genome-wide	Higher data volume; difficult to interpret	Moderate to high	High	Detects CNAs, structural variants, non-coding mutations
Transcriptomics	Bulk RNA-seq	Average expression across all cells	Cell-type heterogeneity obscured	Moderate (bulk RNA)	Moderate	Captures bulk transcriptional signatures and TF expression
	Single-cell RNA-seq	Cell-level resolution	High dropout rate, technical variability	High quality single cells	High	Uncovers heterogeneity, subclonal expression, TME profiles
Epigenomics	Methylation Profiling	CpG-rich regions	Biased methylome coverage	Low (DNA)	Low to moderate	Differentiates TF-defined subtypes; correlates with RNA expression
Proteomics	Mass Spec based Proteomics	Protein-level, post translational	Stochastic sampling, high data volume	Moderate	Moderate to high	Identifies differentially expressed proteins and PTMs

of PitNETs, facilitates identification of biomarkers, elucidates complex regulatory networks, and uncovers potential therapeutic targets. Recent studies have demonstrated that such integrative analyses can reveal distinct molecular subtypes of PitNETs, improve correlations between molecular profiles and clinical outcomes, and provide insights into tumorigenesis and progression (Figure 2).

As Table 3 summarizes, each PitNET subtype is characterized by distinct molecular features across genomic, transcriptomic, epigenomic, proteomic, and metabolomic layers. Despite their differences, these multiomic signatures converge on shared biological pathways across subtype. For instance, somatotroph tumors exhibit GNAS mutations, PIT1-driven transcription, and enrichment of proteins in PI3K/AKT signaling, which collectively support growth hormone hypersecretion via cAMP signaling and metabolic reprogramming (81). Corticotroph tumors exhibit USP8 mutations, upregulation of proopiomelanocortin (POMC), transcriptomic changes in Wnt regulators like SFRP2, and proteomic changes in Galectin-3 and ID2, linking chromatin remodeling and epithelial-to-mesenchymal (EMT) transition with sustained ACTH hypersecretion (82–84). Finally, lactotroph tumors with FIPA or SF3B1 mutations and estrogen receptor activation display epigenetic change (HMGA regulation via chromatin architecture) and proteomic shifts in Galectin-3, HADH1, and ID2, linking genetic mutations and estrogen signaling to altered tumor epigenetics and protein expression patterns that drive tumor aggressiveness and treatment resistance (85, 86). These convergences evidently highlight shared mechanisms such as hormone hypersecretion, chromatin remodeling, biological pathway activation, and metabolic rewiring across tumor types, underscoring the translational value of integrative multiomic analysis in PitNET research. Additional molecular studies across different subtypes remain necessary, as certain subtypes such as Gonadotroph PitNETs lack any published molecular data (87).

Although genomic profiling suggests infrequent rates of somatic mutations in PitNETs, CNAs are common among all TF-lineage subtypes. Integrating analysis of methylation and transcriptional data suggests that hypomethylation of promoter regions is associated with increased RNA expression of *GH1* and Somatostatin Receptor subtype 5 (*SSTR5*) in GH-secreting PitNETs and *POMC* in ACTH-secreting PitNETs (88). In a 2020 multi-omic study, three molecular classes of PitNETs were identified by integrating somatic mutations, chromosomal alterations, and profiling of the miRNAome, methylome, and transcriptome (89). This classification scheme clustered PitNETs similar to the classification based on TF lineage. Prognostic analysis identified that *USP8* wildtype (WT) compared to *USP8* mutant corticotroph PitNETs were more aggressive with invasive properties (89). The transcriptome of these invasive corticotrophs was enriched for genes associated with epithelial-mesenchymal-transition, consistent with their invasive clinical behavior (89). Gene ontology analysis in a transcriptomic and proteomic integrated analysis of *GNAS* mutant vs. wildtype somatotrophs suggested that *GNAS* mutations may impact endocrine features through induction of G protein-coupled receptor (GPCR) pathways. Higher protein expression of WW and C2 domain-containing protein-3 (WWC3), serine incorporator 1 (SERINC1), and zinc finger AN1-type containing 3 (ZFAND3) was correlated with increased tumor volume after somatostatin analog treatment (90). Recurrence as a clinical marker of aggressive disease has also been investigated utilizing multiomic methodologies. A robust longitudinal study of primary and recurrent PitNETs from the same patient determined primary and recurrent PitNETs to have similar genomic profiles but divergent transcriptomic profiles (91). Interestingly,

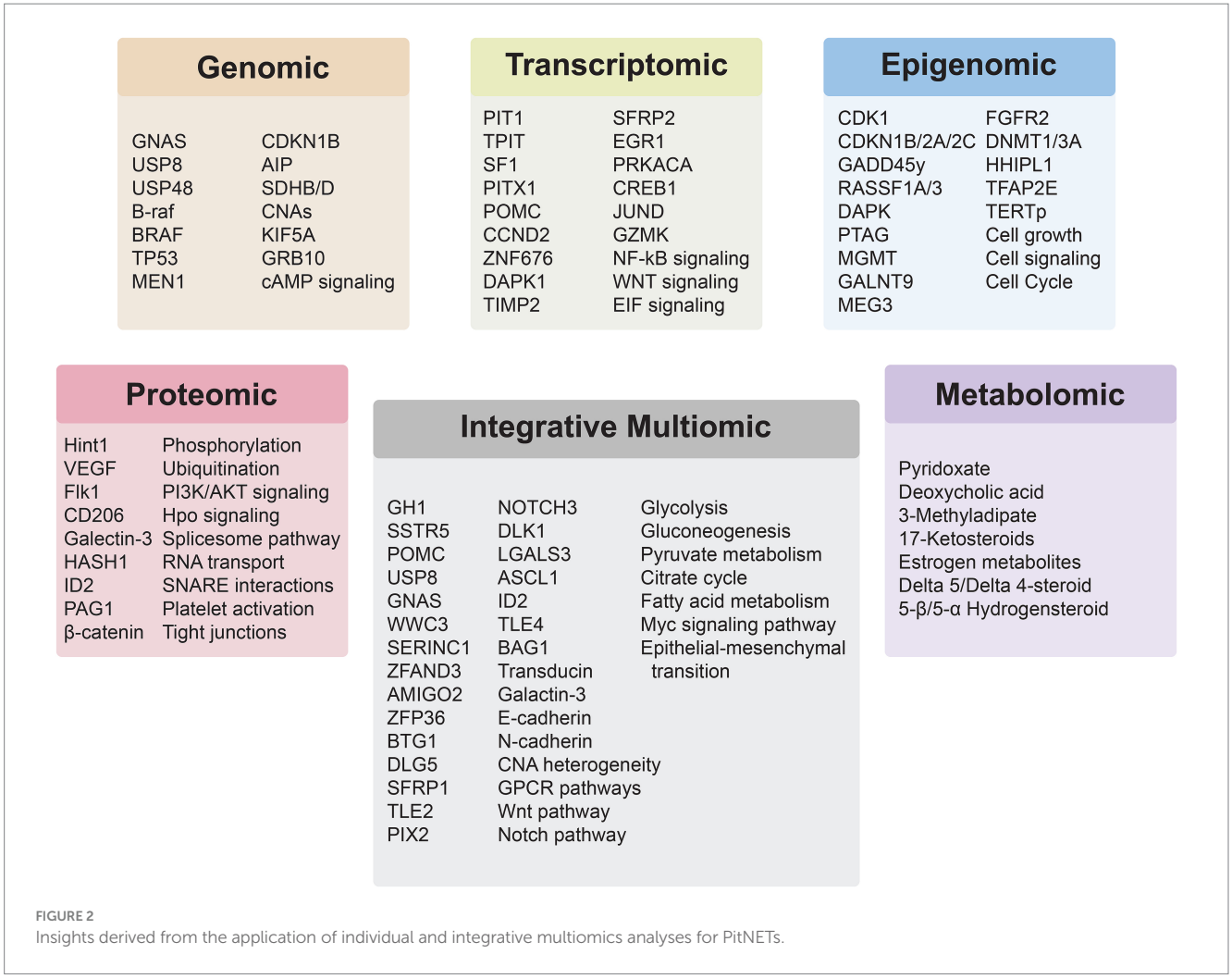


FIGURE 2
Insights derived from the application of individual and integrative multiomics analyses for PitNETs.

several metabolic pathways that were differentially expressed among primary and recurrent tumors based on transcriptional data did not seem to be regulated by methylation, raising the possibility of alternative regulatory mechanisms that warrant further investigation (91).

Multiomic analyses have also incorporated both proteomic and transcriptomic data to further understand PitNET biology. For example, nonfunctioning PitNETs had almost 300 differentially expressed genes and 50 differentially expressed proteins compared to controls including secreted frizzled-related protein 1 (*SFRP1*), transducin like enhancer of split 2 (*TLE2*), *PITX2*, Notch receptor 3 (*NOTCH3*), and delta like non-canonical Notch ligand 1 (*DLK1*) (92). These findings suggest potential critical molecular pathways implicated in this tumor type such as the Wnt and Notch pathways. Integrative proteomic and transcriptomic analysis has also been used to analyze metastatic PitNETs, which led to the identification of almost 5,000 differentially expressed genes, and the downregulation of beta-galactoside binding protein galactin-3. Other genes that may play important roles in metastatic PitNETs include lectin, galactoside-binding, soluble, 3 (*LGALS3*), achaete-scute family bHLH transcription factor 1 (*ASCL1*), *ID2*, and transducin like enhancer of split 4 (*TLE4*) (93). Lastly, transcriptomic and proteomic analysis of prolactinomas compared to normal pituitary glands identified a unique transcriptomic and proteomic profile. Notably, several components of the Notch pathway were altered in prolactinomas, along with increased expression of survival factor BCL2 associated athanogene 1 (BAG1) and decreased expression of E-cadherin and N-cadherin (73).

Metabolomics has been used alongside other -omic methods such as proteomics and lipidomics to delve further into the mechanisms of PitNET pathogenesis. In ACTH-secreting PitNETs, integrated analysis identified that these tumors were significantly enriched in protein-metabolite joint pathways such as glycolysis/gluconeogenesis, pyruvate metabolism, citrate cycle, and fatty acid metabolism (94). The Myc signaling pathway was also identified to have a significant role in the metabolic changes and tumorigenesis of these tumors (94). A broader study using desorption electrospray ionization (DESI-MS) derived phospholipid signals that differed between gray matter, white matter, gliomas, meningiomas and pituitary tumors. Principal component analysis of lipid and metabolite profiles from this analysis were able to separate different tumor types: gliomas, meningiomas, and pituitary tumors (95).

However, while these studies underscore the value of integrative multiomics, they also highlight the significant computational hurdles in merging heterogenous omic datasets. Despite the growing number of multi-omic studies in PitNETs, integration and standardization across datasets remain computationally challenging, as omics data is inherently heterogenous. Several bioinformatic tools have been

10.3389/fmed.2025.1629621

TABLE 3 Molecular features of PitNETs by hormonal expression.

Cell of origin	Somatotroph	Lactotroph	Thyrotroph	Corticotroph	Gonadotroph
Hormone	Growth Hormone	Prolactin	Thyroid stimulating hormone	Adrenocorticotrophic hormone	Luteinizing hormone/ follicle stimulating hormone
Transcription Factor	PIT1	PIT1	PIT1	TPIT	SF1
% of all PitNETs	11	40	0.2	6	43
Molecular features					
Genomics	GNAS AIP	FIPA SF3B1	MEN1 AIP ASTN2 CWH43 R3HDM2 SMOX STYL3 ZSCA23 CNAs	USP8 USP48 BRAF TP53	
Transcriptomics	Three transcriptional subtypes ODC BAG1	Activation of estrogen receptor, oxidative phosphorylation, and EIF signaling		SFRP2 Wnt signaling CCND ZN DAPK1 TIMP2	
Epigenomics		HMGA regulation via chromatin architecture	SLIT1		
Proteomics	IL-4 PDGF PTEN VEGF PI3K/AKT FAK	Galectin-3 HASH1 ID2		Galectin-3 HASH1 ID2	
Metabolomics		Urine 17-ketosteroids Succinic acid Lactic acid		Pyridoxate Deoxycholic acid 3-methyladipate	
Multiomics	GH1 SSTR5 GPCR pathway ATP2A2 ARID5B WWC3 SERINC1 ZFAND3	Notch pathway E-cadherin N-cadherin	Fatty acid metabolism Nitrogen metabolism Insulin PPAR HIPPO PIP5K1B NEK10	POMC Glycolysis Gluconeogenesis Pyruvate metabolism Citrate cycle Fatty acid metabolism Myc signaling	

developed to address these issues. Multi-omics factor analysis uses unsupervised latent factor modeling to identify hidden sources of variation across omics layers (96). Similarity network fusion constructs networks of samples and merges these networks effectively to discover subtypes (97). By contrast, iClusterPlus applies joint latent variable modeling to integrate multiple subtypes of genomic data for subtype identification (98). Unfortunately, these distinct data fusion techniques differ in scalability, handling of missing data, and interpretability. Moreover, these methods are rarely tailored to PitNET-specific datasets, which tend to be small and sparse.

Standardization of data in PitNET omics research faces similar issues. Batch effects, inconsistent normalization strategies, and variable bioinformatics pipelines undermine reproducibility of data. Transcriptomic analysis heavily relies on normalization and batch correction tools like ComBat or Harmony (99, 100). Proteomic and epigenomic analyses use quantile normalization and reference-based scaling to address technical variability (101). Collectively, these techniques' inconsistencies can complicate downstream integration efforts. Hence, adhering to data frameworks such as the NIH's Findable, Accessible, Interoperable, Reusable (FAIR) principles, standardizing pipelines, and reporting metadata in PitNET research would allow for increased reproducibility and comparability of data, facilitating the development of robust PitNET-specific computational pipeline that provide clinically meaningful data.

In parallel with efforts to integrate and standardize multiomic workflows, artificial intelligence (AI) and machine learning (ML) have emerged as powerful tools for analyzing complex multi-omic datasets. Although still in the nascent stages of adoption in PitNET research, these methods are beginning to prove extremely useful. Several studies have already utilized AI and ML to create robust PitNET classifiers for risk stratification and diagnosis. Wang et al. used LASSO regression and Support Vector Machine Recursive Feature Elimination to develop a Programmed Cell Death-associated index (PCDI) classifier that outperforms traditional prognostic models in identifying invasive PitNETs with a high degree of accuracy (102). In another study, Li et al. used radiomic features derived from T2-weighted MRI to construct a Gaussian process model capable of preoperatively predicting histological subtypes of PitNETs, such as prolactinoma (103). Despite these promising results, the translational potential of these approaches is limited by the paucity of PitNET datasets. Collaborative future modeling efforts may allow for more robust and accurate model construction and generalization.

Integrative multi-omics analyses have significantly advanced our understanding of PitNETs by revealing multiple molecular subtypes and the complex regulatory networks that underlie tumor behavior. Building upon these approaches, spatial omics technologies are emerging as vital tools for resolving tumor heterogeneity in its native context. Spatial transcriptomics and proteomics offer significant resolution advancement for characterizing intratumoral heterogeneity and tumor microenvironment architecture in PitNETs. For instance, spatial transcriptomics could distinguish between non-invasive and invasive PitNET phenotypes by localizing EMT markers. Similarly, spatial proteomic analysis could enable the visualization of PTMs throughout the invasive PitNET front. These tools have the potential to refine the current understanding of PitNET pathophysiology and support the development of spatially-informed, precision medicine strategies.

Translational gaps

While multiomics PitNET research has yielded invaluable biological insights, a significant gap remains between academic discovery and clinical translation. Cost and infrastructure requirements for generating and analyzing multi-layered omics data remains prohibitive, especially outside of academic centers. Governmental regulatory pathways for clinical grade omics assays are still evolving, with no PitNET omics-based biomarker panels still having received FDA clearance. Clinical trials for multiomic biomarker validation also remain rare and underpowered.

Conclusion

In conclusion, the integration of multi-omics technologies has profoundly advanced our understanding of PitNETs, offering a comprehensive view of their molecular landscape. By combining data from genomics, transcriptomics, proteomics, epigenomics, and metabolomics, researchers have identified distinct molecular subtypes, unveiled regulatory networks, and discovered novel biomarkers, thereby enhancing diagnostic precision and informing therapeutic strategies. Clinically, these integrative approaches hold promise for the development of personalized medicine in PitNET management, which is a critical need, in particular for recurrent tumors and tumors not cured by the current standard of care. The ability to correlate multiomic profiles with clinical outcomes facilitates more accurate prognostication and the potential for tailored treatment regimens. Looking forward, the continued evolution of computational tools and machine learning algorithms will be critical in managing the complexity of multiomic data, enabling real-time integration and interpretation in clinical settings. Advancements in single-cell and spatial omics technologies are expected to further define tumor heterogeneity and microenvironmental interactions, providing deeper insights into PitNET pathogenesis. Collectively, these developments herald a new era in PitNET management, where multiomic integration becomes central to patient-specific diagnosis, prognosis, and therapy.

Author contributions

SP: Conceptualization, Methodology, Visualization, Writing – original draft, Writing – review & editing. SSP: Methodology, Visualization, Writing – original draft, Writing – review & editing. ZZ: Methodology, Validation, Writing – original draft, Writing – review & editing. JS: Writing – original draft, Writing – review & editing, Project administration. AK: Conceptualization, Supervision, Writing – original draft, Writing – review & editing, Funding acquisition. BP: Conceptualization, Supervision, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. AK: Alvin J. Siteman Cancer Research Fund (GF0010215), the Christopher Davidson and Knight Family Fund, and the Duesenberg Research Fund.

Conflict of interest

AK is a consultant for Monteris Medical and has received nonrelated research grants from Stryker for study of dural substitute.

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.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

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

Publisher's note

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. Price M, Ballard C, Benedetti J, Neff C, Cioffi G, Waite KA, et al. Cbtrus statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2017–2021. *Neuro-Oncology*. (2024) 26:vi1–vi85. doi: 10.1093/neuonc/noae145

2. Asa SL, Mete O, Perry A, Osamura RY. Overview of the 2022 who classification of pituitary tumors. *Endocr Pathol*. (2022) 33:6–26. doi: 10.1007/s12022-022-09703-7

3. Mouchtouris N, Smit RD, Piper K, Prashant G, Evans JJ, Karsy M. A review of multiomics platforms in pituitary adenoma pathogenesis. *Front Biosci*. (2022) 27:77. doi: 10.31083/j.fbl2703077

4. Tang C, Zhong C, Zhu J, Yuan F, Yang J, Xu Y, et al. Gnas mutations suppress cell invasion by activating Meg3 in growth hormone-secreting pituitary adenoma. *Oncol Res*. (2024) 32:1079–91. doi: 10.32604/or.2024.046007

5. Reincke M, Sbiera S, Hayakawa A, Theodoropoulou M, Osswald A, Beuschlein F, et al. Mutations in the deubiquitinase gene Usp8 cause Cushing's disease. *Nat Genet*. (2015) 47:31–8. doi: 10.1038/ng.3166

6. Perez-Rivas LG, Theodoropoulou M, Ferraù F, Nusser C, Kawaguchi K, Stratakis CA, et al. The gene of the ubiquitin-specific protease 8 is frequently mutated in adenomas causing Cushing's disease. *J Clin Endocrinol Metab*. (2015) 100:E997–E1004. doi: 10.1210/jc.2015-1453

7. Perez-Rivas LG, Simon J, Albani A, Tang S, Roeber S, Assié G, et al. Tp53 mutations in functional Corticotroph tumors are linked to invasion and worse clinical outcome. *Acta Neuropathol Commun*. (2022) 10:139. doi: 10.1186/s40478-022-01437-1

8. Chen J, Jian X, Deng S, Ma Z, Shou X, Shen Y, et al. Identification of recurrent Usp48 and Braf mutations in Cushing's disease. *Nat Commun*. (2018) 9:3171. doi: 10.1038/s41467-018-05275-5

9. Dénes J, Korbonits M. The clinical aspects of pituitary tumour genetics. *Endocrine*. (2021) 71:663–74. doi: 10.1007/s12020-021-02633-0

10. Chin SO, Chik C, Tateno T. Pituitary neuroendocrine tumors in multiple endocrine neoplasia. *Endocrinol Metab (Seoul)*. (2025) 40:39–46. doi: 10.3803/EnM.2024.2074

11. Stratakis CA, Tichomirowa MA, Boikos S, Azevedo MF, Lodish M, Martari M, et al. The role of germline Aip, Men1, Prkar1a, Cdkn1b and Cdkn2c mutations in causing pituitary adenomas in a large cohort of children, adolescents, and patients with genetic syndromes. *Clin Genet*. (2010) 78:457–63. doi: 10.1111/j.1399-0004.2010.01406.x

12. Xekouki P, Stratakis CA. Succinate dehydrogenase (Sdhx) mutations in pituitary tumors: could this be a new role for mitochondrial complex II and/or Krebs cycle defects? *Endocr Relat Cancer*. (2012) 19:C33–40. doi: 10.1530/erc-12-0118

13. Daly AF, Tichomirowa MA, Petrossians P, Heliövaara E, Jaffrain-Rea ML, Barlier A, et al. Clinical characteristics and therapeutic responses in patients with germ-line Aip mutations and pituitary adenomas: an international collaborative study. *J Clin Endocrinol Metab*. (2010) 95:E373–83. doi: 10.1210/jc.2009-2556

14. Hernández-Ramírez LC, Gabrovska P, Dénes J, Stals K, Trivellin G, Tilley D, et al. Landscape of familial isolated and young-onset pituitary adenomas: prospective diagnosis in Aip mutation carriers. *J Clin Endocrinol Metab*. (2015) 100:E1242–54. doi: 10.1210/jc.2015-1869

15. Song Z-J, Reitman ZJ, Ma Z-Y, Chen J-H, Zhang Q-L, Shou X-F, et al. The genome-wide mutational landscape of pituitary adenomas. *Cell Res*. (2016) 26:1255–9. doi: 10.1038/cr.2016.114

16. Bi WL, Horowitz P, Greenwald NF, Abedalthagafi M, Agarwalla PK, Gibson WJ, et al. Landscape of genomic alterations in pituitary adenomas. *Clin Cancer Res*. (2017) 23:1841–51. doi: 10.1158/1078-0432.Ccr-16-0790

17. Oh JY, Osorio RC, Jung J, Carrete L, Choudhary N, Lad M, et al. Transcriptomic profiles of normal pituitary cells and pituitary neuroendocrine tumor cells. *Cancers (Basel)*. (2022) 15:110. doi: 10.3390/cancers15010110

18. Tebani A, Jotanovic J, Hekmati N, Sivertsson Å, Gudjonsson O, Edén Engström B, et al. Annotation of pituitary neuroendocrine tumors with genome-wide expression analysis. *Acta Neuropathol Commun*. (2021) 9:181. doi: 10.1186/s40478-021-01284-6

19. Lamolet B, Pulichino AM, Lamonerie T, Gauthier Y, Brue T, Enjalbert A, et al. A pituitary cell-restricted T box factor, Tpit, activates Pomc transcription in cooperation with Pitx Homeoproteins. *Cell*. (2001) 104:849–59. doi: 10.1016/s0092-8674(01)00282-3

20. Zhang D, Hugo W, Bergsneider M, Wang MB, Kim W, Vinters HV, et al. Single-cell Rna sequencing in silent Corticotroph tumors confirms impaired Pomc processing and provides new insights into their invasive behavior. *Eur J Endocrinol*. (2022) 187:49–64. doi: 10.1530/eje-21-1183

21. Pulichino AM, Vallette-Kasic S, Tsai JP, Couture C, Gauthier Y, Drouin J. Tpit determines alternate fates during pituitary cell differentiation. *Genes Dev*. (2003) 17:738–47. doi: 10.1101/gad.1065703

22. Zhu X, Gleiberman AS, Rosenfeld MG. Molecular physiology of pituitary development: signaling and transcriptional networks. *Physiol Rev*. (2007) 87:933–63. doi: 10.1152/physrev.00006.2006

23. Toda M, Tamura R, Toda M. Recent progress in stem cell research of the pituitary gland and pituitary adenoma. *Endocrine*. (2020) 1:49–57. doi: 10.3390/endocrines1010006

24. Zhao L, Bakke M, Krimkevich Y, Cushman LJ, Parlow AF, Camper SA, et al. Steroidogenic factor 1 (Sf1) is essential for pituitary Gonadotrope function. *Development*. (2001) 128:147–54. doi: 10.1242/dev.128.2.147

25. Jotanovic J, Tebani A, Hekmati N, Sivertsson Å, Lindskog C, Uhlèn M, et al. Transcriptome analysis reveals distinct patterns between the invasive and noninvasive pituitary neuroendocrine tumors. *J Endocr Soc*. (2024) 8:bvae040. doi: 10.1210/jendso/bvae040

26. Wierinckx A, Auger C, Devauchelle P, Reynaud A, Chevallier P, Jan M, et al. A diagnostic marker set for invasion, proliferation, and aggressiveness of prolactin pituitary tumors. *Endocr Relat Cancer*. (2007) 14:887–900. doi: 10.1677/erc-07-0062

27. Galland F, Lacroix L, Saulnier P, Dessen P, Meduri G, Bernier M, et al. Differential gene expression profiles of invasive and non-invasive non-functioning pituitary adenomas based on microarray analysis. *Endocr Relat Cancer*. (2010) 17:361–71. doi: 10.1677/erc-10-0018

28. de Araújo LJ, Lerario AM, de Castro M, Martins CS, Bronstein MD, Machado MC, et al. Transcriptome analysis showed a differential signature between invasive and non-invasive corticotrophinomas. *Front Endocrinol (Lausanne)*. (2017) 8:55. doi: 10.3389/fendo.2017.00055

29. Wu Q, Yin X, Zhao W, Xu W, Chen L. Downregulation of Sfrp2 facilitates Cancer Stemness and Radioresistance of glioma cells via activating Wnt/β-catenin signaling. *PLoS One*. (2021) 16:e0260864. doi: 10.1371/journal.pone.0260864

30. Hattori Y, Tahara S, Ozawa H, Morita A, Ishii H. Transcriptomic profiling of Lactotroph pituitary neuroendocrine tumors via Rna sequencing and ingenuity pathway analysis. *Neuroendocrinology*. (2024) 114:670–80. doi: 10.1159/000539017

31. Peng J, Yuan L, Kang P, Jin S, Ma S, Zhou W, et al. Comprehensive transcriptomic analysis identifies three distinct subtypes of pituitary adenomas: insights into tumor behavior, prognosis, and stem cell characteristics. *J Transl Med*. (2024) 22:892. doi: 10.1186/s12967-024-05702-w

32. Butz H. Circulating noncoding Rnas in pituitary neuroendocrine tumors-two sides of the same coin. *Int J Mol Sci*. (2022) 23:5122. doi: 10.3390/ijms23095122

33. Yan N, Xie W, Wang D, Fang Q, Guo J, Chen Y, et al. Single-cell transcriptomic analysis reveals tumor cell heterogeneity and immune microenvironment features of pituitary neuroendocrine tumors. *Genome Med*. (2024) 16:2. doi: 10.1186/s13073-023-01267-3

34. Cui Y, Li C, Jiang Z, Zhang S, Li Q, Liu X, et al. Single-cell transcriptome and genome analyses of pituitary neuroendocrine tumors. *Neuro-Oncology*. (2021) 23:1859–71. doi: 10.1093/neuonc/noab102

35. Hauser BM, Lau A, Gupta S, Bi WL, Dunn IF. The Epigenomics of pituitary adenoma. *Front Endocrinol (Lausanne)*. (2019) 10:290. doi: 10.3389/fendo.2019.00290

36. Mosella MS, Sabedot TS, Silva TC, Malta TM, Dezem FS, Asmaro KP, et al. DNA methylation-based signatures classify sporadic pituitary tumors according to Clinicopathological features. *Neuro-Oncology*. (2021) 23:1292–303. doi: 10.1093/neuonc/noab044

37. Lidhar K, Korbonits M, Jordan S, Khalimova Z, Kaltsas G, Lu X, et al. Low expression of the cell cycle inhibitor P27kip1 in normal corticotroph cells, corticotroph tumors, and malignant pituitary tumors. *J Clin Endocrinol Metab*. (1999) 84:3823–30. doi: 10.1210/jcem.84.10.6066

38. Zhou Y, Zhang X, Klibanski A. Genetic and epigenetic mutations of tumor suppressive genes in sporadic pituitary adenoma. *Mol Cell Endocrinol*. (2014) 386:16–33. doi: 10.1016/j.mce.2013.09.006

39. Kirsch M, Mörz M, Pinzer T, Schackert HK, Schackert G. Frequent loss of the Cdkn2c (P18ink4c) gene product in pituitary adenomas. *Genes Chromosomes Cancer*. (2009) 48:143–54. doi: 10.1002/gcc.20621

40. Hossain MG, Iwata T, Mizusawa N, Qian ZR, Shima SW, Okutsu T, et al. Expression of P18(Ink4c) is down-regulated in human pituitary adenomas. *Endocr Pathol*. (2009) 20:114–21. doi: 10.1007/s12022-009-9076-0

41. Simpson DJ, Hibberts NA, McNicol AM, Clayton RN, Farrell WE. Loss of Prb expression in pituitary adenomas is associated with methylation of the Rb1 Cpg island. *Cancer Res*. (2000) 60:1211–6.

42. Tateno T, Nakano-Tateno T, Ezzat S, Asa SL. Ng2 targets tumorigenic Rb inactivation in Pit1-lineage pituitary cells. *Endocr Relat Cancer*. (2016) 23:445–56. doi: 10.1530/erc-16-0013

43. Nakayama K, Ishida N, Shirane M, Inomata A, Inoue T, Shishido N, et al. Mice lacking P27(Kip1) display increased body size, multiple organ hyperplasia, retinal dysplasia, and pituitary tumors. *Cell*. (1996) 85:707–20. doi: 10.1016/s0092-8674(00)81237-4

44. Zhang X, Sun H, Danila DC, Johnson SR, Zhou Y, Swearingen B, et al. Loss of expression of Gadd45 gamma, a growth inhibitory gene, in human pituitary adenomas: implications for tumorigenesis. *J Clin Endocrinol Metab*. (2002) 87:1262–7. doi: 10.1210/jcem.87.3.8315

45. Carrier F, Georgel PT, Pourquier P, Blake M, Kontny HU, Antinore MJ, et al. Gadd45, a P53-responsive stress protein, modifies DNA accessibility on damaged chromatin. *Mol Cell Biol*. (1999) 19:1673–85. doi: 10.1128/mcb.19.3.1673

46. Bahar A, Bicknell JE, Simpson DJ, Clayton RN, Farrell WE. Loss of expression of the growth inhibitory gene Gadd45gamma, in human pituitary adenomas, is associated with Cpg Island methylation. *Oncogene*. (2004) 23:936–44. doi: 10.1038/sj.onc.1207193

47. Qian ZR, Sano T, Yoshimoto K, Yamada S, Ishizuka A, Mizusawa N, et al. Inactivation of Rassf1a tumor suppressor gene by aberrant promoter Hypermethylation in human pituitary adenomas. *Lab Investig*. (2005) 85:464–73. doi: 10.1038/labinvest.3700248

48. Simpson DJ, Clayton RN, Farrell WE. Preferential loss of death associated protein kinase expression in invasive pituitary Tumours is associated with either Cpg Island methylation or homozygous deletion. *Oncogene*. (2002) 21:1217–24. doi: 10.1038/sj.onc.1205195

49. Zhang X, Zhou Y, Klibanski A. Isolation and characterization of novel pituitary tumor related genes: a Cdna representational difference approach. *Mol Cell Endocrinol*. (2010) 326:40–7. doi: 10.1016/j.mce.2010.02.040

50. Zhao J, Dahle D, Zhou Y, Zhang X, Klibanski A. Hypermethylation of the promoter region is associated with the loss of Meg3 gene expression in human pituitary tumors. *J Clin Endocrinol Metab*. (2005) 90:2179–86. doi: 10.1210/jc.2004-1848

51. Zhu X, Lee K, Asa SL, Ezzat S. Epigenetic silencing through DNA and histone methylation of fibroblast growth factor receptor 2 in neoplastic pituitary cells. *Am J Pathol*. (2007) 170:1618–28. doi: 10.2353/ajpath.2007.061111

52. Ma HS, Wang EL, Xu WF, Yamada S, Yoshimoto K, Qian ZR, et al. Overexpression of DNA (Cytosine-5)-methyltransferase 1 (Dnmt1) and DNA (Cytosine-5)-methyltransferase 3a (Dnmt3a) is associated with aggressive behavior and hypermethylation of tumor suppressor genes in human pituitary adenomas. *Med Sci Monit*. (2018) 24:4841–50. doi: 10.12659/msm.910608

53. Duong CV, Emes RD, Wessely F, Yacqub-Usman K, Clayton RN, Farrell WE. Quantitative, genome-wide analysis of the DNA Methylome in sporadic pituitary adenomas. *Endocr Relat Cancer*. (2012) 19:805–16. doi: 10.1530/erc-12-0251

54. Ling C, Pease M, Shi L, Punj V, Shiroishi MS, Commins D, et al. A pilot genome-scale profiling of DNA methylation in sporadic pituitary macroadenomas: association with tumor invasion and histopathological subtype. *PLoS One*. (2014) 9:e96178. doi: 10.1371/journal.pone.0096178

55. Gu Y, Zhou X, Hu F, Yu Y, Xie T, Huang Y, et al. Differential DNA Methylome profiling of nonfunctioning pituitary adenomas suggesting tumour invasion is correlated with cell adhesion. *J Neuro-Oncol*. (2016) 129:23–31. doi: 10.1007/s11060-016-2139-4

56. Toraño EG, Petrus S, Fernandez AF, Fraga MF. Global DNA Hypomethylation in Cancer: review of validated methods and clinical significance. *Clin Chem Lab Med*. (2012) 50:1733–42. doi: 10.1515/cclm-2011-0902

57. Boresowicz J, Kober P, Rusetska N, Maksymowicz M, Goryca K, Kunicki J, et al. Telomere length and tert abnormalities in pituitary adenomas. *Neuro Endocrinol Lett*. (2018) 39:49–55.

58. Miyake Y, Adachi JI, Suzuki T, Mishima K, Araki R, Mizuno R, et al. Tert promoter methylation is significantly associated with Tert upregulation and disease progression in pituitary adenomas. *J Neuro-Oncol*. (2019) 141:131–8. doi: 10.1007/s11060-018-03016-8

59. Alzoubi H, Minasi S, Gianno F, Antonelli M, Belardinilli F, Giangaspero F, et al. Alternative lengthening of telomeres (alt) and telomerase reverse transcriptase promoter methylation in recurrent adult and primary pediatric pituitary neuroendocrine tumors. *Endocr Pathol*. (2022) 33:494–505. doi: 10.1007/s12022-021-09702-0

60. Kovacs K, Scheithauer BW, Lombardero M, McLendon RE, Syro LV, Uribe H, et al. Mgmt Immunoexpression predicts responsiveness of pituitary tumors to Temozolomide therapy. *Acta Neuropathol*. (2008) 115:261–2. doi: 10.1007/s00401-007-0279-5

61. Widhalm G, Wolfsberger S, Preusser M, Woehrer A, Kotter MR, Czech T, et al. O(6)-Methylguanine DNA methyltransferase Immunoexpression in nonfunctioning pituitary adenomas: are progressive tumors potential candidates for Temozolomide treatment? *Cancer*. (2009) 115:1070–80. doi: 10.1002/cncr.24053

62. Hirohata T, Asano K, Ogawa Y, Takano S, Amano K, Isozaki O, et al. DNA mismatch repair protein (Msh6) correlated with the responses of atypical pituitary adenomas and pituitary carcinomas to Temozolomide: the National Cooperative Study by the Japan Society for Hypothalamic and Pituitary Tumors. *J Clin Endocrinol Metab*. (2013) 98:1130–6. doi: 10.1210/jc.2012-2924

63. McCormack AI, McDonald KL, Gill AJ, Clark SJ, Burt MG, Campbell KA, et al. Low O6-methylguanine-DNA methyltransferase (Mgmt) expression and response to temozolomide in aggressive pituitary tumours. *Clin Endocrinol*. (2009) 71:226–33. doi: 10.1111/j.1365-2265.2008.03487.x

64. Kushihara Y, Tanaka S, Kobayashi Y, Nagaoka K, Kikuchi M, Nejo T, et al. Glioblastoma with high O6-methyl-guanine DNA methyltransferase expression are more immunologically active than tumors with low Mgmt expression. *Front Immunol*. (2024) 15:1328375. doi: 10.3389/fimmu.2024.1328375

65. Chen R, Zhang M, Zhou Y, Guo W, Yi M, Zhang Z, et al. The application of histone deacetylases inhibitors in glioblastoma. *J Exp Clin Cancer Res*. (2020) 39:138. doi: 10.1186/s13046-020-01643-6

66. Candy NG, Ramezanpour M, Bouras G, Chegeni N, Chataway T, Ormsby RJ, et al. Differential proteomic expression in non-functional pituitary neuroendocrine Tumours and pituitary glands. *Rhinology*. (2024) 62:750–8. doi: 10.4193/Rhin24.216

67. Cheng T, Wang Y, Lu M, Zhan X, Zhou T, Li B, et al. Quantitative analysis of proteome in non-functional pituitary adenomas: clinical relevance and potential benefits for the patients. *Front Endocrinol (Lausanne)*. (2019) 10:854. doi: 10.3389/fendo.2019.00854

68. Rai A, Yelamanchi SD, Radotra BD, Gupta SK, Mukherjee KK, Tripathi M, et al. Phosphorylation of β-catenin at Serine552 correlates with invasion and recurrence of non-functioning pituitary neuroendocrine Tumours. *Acta Neuropathol Commun*. (2022) 10:138. doi: 10.1186/s40478-022-01441-5

69. Li J, Wen S, Li B, Li N, Zhan X. Phosphorylation-mediated molecular pathway changes in human pituitary neuroendocrine tumors identified by quantitative Phosphoproteomics. *Cells*. (2021) 10:2225. doi: 10.3390/cells10092225

70. Qian S, Zhan X, Lu M, Li N, Long Y, Li X, et al. Quantitative analysis of Ubiquitinated proteins in human pituitary and pituitary adenoma tissues. *Front Endocrinol (Lausanne)*. (2019) 10:328. doi: 10.3389/fendo.2019.00328

71. Carrillo-Najar C, Rembao-Bojórquez D, Tena-Suck ML, Zavala-Vega S, Gelista-Herrera N, Ramos-Peek MA, et al. Comparative proteomic study shows the expression of Hint-1 in pituitary adenomas. *Diagnostics*. (2021) 11:330. doi: 10.3390/diagnostics11020330

72. Lin K, Zhang J, Lin Y, Pei Z, Wang S. Metabolic characteristics and M2 macrophage infiltrates in invasive nonfunctioning pituitary adenomas. *Front Endocrinol (Lausanne)*. (2022) 13:901884. doi: 10.3389/fendo.2022.901884

73. Evans CO, Moreno CS, Zhan X, McCabe MT, Vertino PM, Desiderio DM, et al. Molecular pathogenesis of human prolactinomas identified by gene expression profiling, Rt-Qpcr, and proteomic analyses. *Pituitary*. (2008) 11:231–45. doi: 10.1007/s11102-007-0082-2

74. Zhan X, Wang X, Desiderio DM. Pituitary adenoma Nitroproteomics: current status and perspectives. *Oxidative Med Cell Longev*. (2013) 2013:580710. doi: 10.1155/2013/580710

75. Zhan X, Desiderio DM. Nitroproteins from a human pituitary adenoma tissue discovered with a Nitrotyrosine affinity column and tandem mass spectrometry. *Anal Biochem*. (2006) 354:279–89. doi: 10.1016/j.ab.2006.05.024

76. Bhat FA, Mangalaparthi KK, Ding H, Jain A, Hsu JS, Peterson JA, et al. Exploration of Nitrotyrosine-containing proteins and peptides by antibody-based enrichment strategies. *Mol Cell Proteomics*. (2024) 23:100733. doi: 10.1016/j.mcpro.2024.100733

77. Pînzariu O, Georgescu B, Georgescu CE. Metabolomics-a promising approach to pituitary adenomas. *Front Endocrinol (Lausanne)*. (2018) 9:814. doi: 10.3389/fendo.2018.00814

78. Calligaris D, Feldman DR, Norton I, Olubiyi O, Changelian AN, Machaidze R, et al. Maldi mass spectrometry imaging analysis of pituitary adenomas for near-real-time tumor delineation. *Proc Natl Acad Sci USA*. (2015) 112:9978–83. doi: 10.1073/pnas.1423101112

79. Oklu R, Deipolyi AR, Wicky S, Ergul E, Deik AA, Chen JW, et al. Identification of small compound biomarkers of pituitary adenoma: a bilateral inferior petrosal sinus sampling study. *J Neurointerv Surg*. (2014) 6:541–6. doi: 10.1136/neurintsurg-2013-010821

80. Lee SH, Nam SY, Chung BC. Altered profile of endogenous steroids in the urine of patients with Prolactinoma. *Clin Biochem*. (1998) 31:529–35. doi: 10.1016/s0009-9120(98)00063-0

81. Yang Y, Yao Y, Deng K, Xing B, Lian W, You H, et al. Somatic Gnas mutations in acromegaly: prevalence, clinical features and gender differences. *Endocr Connect*. (2025) 14:266. doi: 10.1530/EC-24-0266

82. Huang C, Shi Y, Zhao Y. Usp8 mutation in Cushing's disease. *Oncotarget*. (2015) 6:18240–1. doi: 10.18632/oncotarget.4856

83. Paes T, Hofland LJ, Iyer AM, Feelders RA. Epigenetic implications in the pathogenesis of Corticotroph tumors. *Pituitary*. (2025) 28:51. doi: 10.1007/s11102-025-01522-3

84. Sumal AKS, Zhang D, Heaney AP. Refractory Corticotroph adenomas. *Pituitary*. (2023) 26:269–72. doi: 10.1007/s11102-023-01308-5

85. Biagetti B, Simo R. Molecular pathways in Prolactinomas: translational and therapeutic implications. *Int J Mol Sci*. (2021) 22:22 (20). doi: 10.3390/ijms222011247

86. Paes T, Buelvas Mebarak J, Magnotto JC, Stamatiades GA, Kuang Y, Paweletz CP, et al. Somatic activating Esr1 mutation in an aggressive Prolactinoma. *J Clin Endocrinol Metab*. (2025) 110:1166–76. doi: 10.1210/clinem/dgae615

87. Ilie MD, Vasiljevic A, Louvet C, Jouanneau E, Raverot G. Gonadotroph tumors show subtype differences that might have implications for therapy. *Cancers (Basel)*. (2020) 12:e1012. doi: 10.3390/cancers12041012

88. Salomon MP, Wang X, Marzese DM, Hsu SC, Nelson N, Zhang X, et al. The epigenomic landscape of pituitary adenomas reveals specific alterations and differentiates among acromegaly, Cushing's disease and endocrine-inactive subtypes. *Clin Cancer Res*. (2018) 24:4126–36. doi: 10.1158/1078-0432.Ccr-17-2206

89. Neou M, Villa C, Armignacco R, Jouinot A, Raffin-Sanson M-L, Septier A, et al. Pangenomic classification of pituitary neuroendocrine tumors. *Cancer Cell*. (2020) 37:123–34.e5. doi: 10.1016/j.ccell.2019.11.002

90. Yamato A, Nagano H, Gao Y, Matsuda T, Hashimoto N, Nakayama A, et al. Proteogenomic landscape and clinical characterization of Gh-producing pituitary adenomas/Somatotroph pituitary neuroendocrine tumors. *Commun Biol*. (2022) 5:1304. doi: 10.1038/s42003-022-04272-1

91. Taniguchi-Ponciano K, Hinojosa-Alvarez S, Hernandez-Perez J, Chavez-Santoscoy RA, Remba-Shapiro I, Guinto G, et al. Longitudinal multiomics analysis of aggressive pituitary neuroendocrine tumors: comparing primary and recurrent tumors from the same patient, reveals genomic stability and heterogeneous transcriptomic profiles with alterations in metabolic pathways. *Acta Neuropathol Commun*. (2024) 12:142. doi: 10.1186/s40478-024-01796-x

92. Moreno CS, Evans CO, Zhan X, Okor M, Desiderio DM, Oyesiku NM. Novel molecular signaling and classification of human clinically nonfunctional pituitary adenomas identified by gene expression profiling and proteomic analyses. *Cancer Res*. (2005) 65:10214–22. doi: 10.1158/0008-5472.Can-05-0884

93. Ruebel KH, Leontovich AA, Jin L, Stilling GA, Zhang H, Qian X, et al. Patterns of gene expression in pituitary carcinomas and adenomas analyzed by high-density oligonucleotide arrays, reverse transcriptase-quantitative Pcr, and protein expression. *Endocrine*. (2006) 29:435–44. doi: 10.1385/endo:29:3:435

94. Feng J, Zhang Q, Zhou Y, Yu S, Hong L, Zhao S, et al. Integration of proteomics and metabolomics revealed metabolite-protein networks in Acth-secreting pituitary adenoma. *Front Endocrinol (Lausanne)*. (2018) 9:678. doi: 10.3389/fendo.2018.00678

95. Jarmusch AK, Pirro V, Baird Z, Hattab EM, Cohen-Gadol AA, Cooks RG. Lipid and metabolite profiles of human brain tumors by desorption electrospray ionization-Ms. *Proc Natl Acad Sci USA*. (2016) 113:1486–91. doi: 10.1073/pnas.1523306113

96. Argelaguet R, Velten B, Arnol D, Dietrich S, Zenz T, Marioni JC, et al. Multi-omics factor analysis-a framework for unsupervised integration of multi-omics data sets. *Mol Syst Biol*. (2018) 14:e8124. doi: 10.15252/msb.20178124

97. Wang B, Mezlini AM, Demir F, Fiume M, Tu Z, Brudno M, et al. Similarity network fusion for aggregating data types on a genomic scale. *Nat Methods*. (2014) 11:333–7. doi: 10.1038/nmeth.2810

98. Shen R, Olshen AB, Ladanyi M. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics*. (2009) 25:2906–12. doi: 10.1093/bioinformatics/btp543

99. Arevalo J, Su E, Ewald JD, van Dijk R, Carpenter AE, Singh S. Evaluating batch correction methods for image-based cell profiling. *Nat Commun*. (2024) 15:6516. doi: 10.1038/s41467-024-50613-5

100. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with harmony. *Nat Methods*. (2019) 16:1289–96. doi: 10.1038/s41592-019-0619-0

101. Zhao Y, Wong L, Goh WWB. How to do quantile normalization correctly for gene expression data analyses. *Sci Rep*. (2020) 10:15534. doi: 10.1038/s41598-020-72664-6

102. Wang G, Yan S, Zhang L, Lin L, Liu R, Han Y, et al. Machine learning-driven Pcdi classifier for invasive Pitnets. *Curr Gene Ther*. (2025) 25:831. doi: 10.2174/0115665232399193250529074831

103. Li H, Liu Z, Li F, Xia Y, Zhang T, Shi F, et al. Identification of Prolactinoma in pituitary neuroendocrine tumors using Radiomics analysis based on multiparameter Mri. *J Imaging Inform Med*. (2024) 37:2865–73. doi: 10.1007/s10278-024-01153-3

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 24 September 2025
DOI 10.3389/fmed.2025.1664359

Check for updates

OPEN ACCESS

EDITED BY
HaiHui Huang,
Shaoguan University, China

REVIEWED BY
Changmin Peng,
George Washington University, United States
Asif Iqbal Khan,
Dow University of Health Sciences, Pakistan

*CORRESPONDENCE
Jiehua Li
953983654@qq.com

RECEIVED 11 July 2025
ACCEPTED 26 August 2025
PUBLISHED 24 September 2025

CITATION
Liu Y, Li Y and Li J (2025) Integrative analysis of serum microorganisms and serum metabolomics in osteoporosis patients based on 16S rDNA sequencing and UHPLC/MS-based metabolomics.
Front. Med. 12:1664359.
doi: 10.3389/fmed.2025.1664359

COPYRIGHT
© 2025 Liu, Li 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.

Integrative analysis of serum microorganisms and serum metabolomics in osteoporosis patients based on 16S rDNA sequencing and UHPLC/MS-based metabolomics

Yu Liu[1,2], Yun Li[3] and Jiehua Li[1]*

[1]Department of Geriatrics, The First Affiliated Hospital of Anhui Medical University, Hefei, China, [2]Department of Geriatrics, The Third Affiliated Hospital of Anhui Medical University, Hefei, China, [3]School of Biology and Pharmaceutical Engineering, Wuhan Polytechnic University, Wuhan, China

Introduction: Although significant progress has been made in the treatment and research of osteoporosis patients in recent years, the genetic mechanism of osteoporosis has not yet been fully elucidated.
Methods: We conducted a comprehensive analysis using 16S sequencing and UHPLC–MS/MS metabolomics data to characterize the microbial composition and metabolic composition in the serum of osteoporosis patients.
Results: At the phylum level, *Proteobacteria* are mainly present in Osteoporosis; In Normal, it is mainly *Bacteroidota*. At the genus level, *Cupriavidus* is the main species in Osteoporosis; In Normal, the main ones are *Blautia*, *Bacteroides*, *Alcaligenes* and *Pseudomonas*. Serum metabolomics revealed different metabolites (230 significantly differentially expressed metabolites) and lipid metabolism pathways (such as Glycerophospholipid metabolism) among the two groups. The combined serum microbiota and serum metabolomics datasets demonstrate a correlation reflecting the impact of microbiota on metabolic activity ($p < 0.05$).
Discussion: Our research findings indicate that microbiota and metabolomics analysis provide important candidate biomarkers. The correlation between these serum microbiota and host metabolism is of great significance for optimizing early diagnosis and developing personalized treatment strategies. This study elucidates the relationship between serum microbiota and metabolites in osteoporosis.

KEYWORDS

metabolomics, microbiome, serum, osteoporosis, biomarkers

1 Introduction

Osteoporosis is a common metabolic disorder, mainly characterized by reduced bone mass and abnormal bone tissue microstructure, decreased bone strength, increased bone fragility, and increased risk of fractures (1). Osteoporosis can be divided into primary osteoporosis, secondary osteoporosis, and other types of osteoporosis (2). The clinical manifestations of osteoporosis mainly include lower back pain, bone pain, spinal deformity, fractures, muscle weakness, fatigue, and worsening symptoms after activity. Psychological abnormalities may occur due to the impact of the disease on daily life, including fear, anxiety, depression, and loss of confidence (3). Despite the widespread use of medical therapies in the past decade, osteoporosis remains the leading cause of life-threatening conditions for the elderly, second

only to tumors and cardiovascular diseases (4). Osteoporosis is usually accompanied by severe bone pain, changes in bone density, and alterations in serum bone metabolism indicators (5) Therefore, identifying biomarkers of osteoporosis is of great significance for preventing the occurrence of osteoporosis.

Research has shown that the structure and composition of the microbiota in osteoporosis have undergone significant changes (6). Patients with low bone density have dysbiosis of the microbiota, and a decrease in the number of *bifidobacteria* and *lactobacilli* is associated with a decrease in bone density (7). Research has found that patients with dysbiosis but no osteoporosis have already experienced decreased bone density and abnormal bone metabolism, and the dysbiosis in the osteoporosis group is more severe, indicating that dysbiosis has already affected bone metabolism and bone density to some extent before osteoporosis occurs (8). Dysregulation of lipid metabolism plays an important role in the pathogenesis of osteoporosis (9, 10). In addition, various amino acids such as arginine, threonine, and tryptophan can affect bone density (11).

The changes in microbiota and metabolism may be related to the pathogenesis of osteoporosis (12–14), but the microbiota characteristics and metabolic profile of osteoporosis patients still need to be determined. In this study, we analyzed the microbiota and metabolic profiles of 18 osteoporosis patients and 18 healthy volunteers using high-throughput sequencing and non-targeted metabolomics. The combination of these two omics can reveal how microorganisms affect host metabolic processes and how metabolites regulate microbial growth and function by analyzing the correlation between microbial diversity and metabolite abundance. Based on multi-omics analysis, we identified specific characteristics of the microbiota and host metabolite profiles associated with osteoporosis, and further established these relationships, revealing the relationship between microbiota and serum metabolite functional modules. Our research reveals that the integration of metabolomics and 16S rRNA sequencing analysis may reveal the interactions occurring between hosts and microbial communities.

2 Materials and methods

2.1 Study population

The 2013 Helsinki Declaration is in compliance with this study, which has been approved by the Ethics Committee for Life Sciences at Hefei First People's Hospital. Prior to registration, written informed consent was provided by all participants. Among them, there were 18 healthy volunteers (Normal group) and 18 newly diagnosed osteoporosis patients (Osteoporosis group). Inclusion criteria for participants: no previous history of cancers; Participants who signed the informed consent form for the study. The exclusion criteria for participants are as follows: cancer patients; Participants who have not signed the informed consent form; patients had been treated with antibiotics in the past 6 months (15).

2.2 Sample collection and preparation

The collection of fasting blood from 36 participants was conducted during clinical examinations. Blood samples were collected from blood vessels using serum separation gel containing coagulants. After standing at room temperature for 60 min to coagulate, they were centrifuged at 3000 rpm for 10 min at 4 °C. 250 μL of supernatant was collected and divided into numbered and suitable 2 mL centrifuge tubes. After the samples are processed, they should be stored in a − 80 °C freezer to avoid repeated freezing and thawing of the collected samples (16).

2.3 DNA extraction and 16S rDNA sequencing

The genomic DNA of the sample was extracted by CTAB or SDS method and then the purity and concentration of DNA were detected by agarose gel electrophoresis. An appropriate amount of sample DNA was taken into a centrifuge tube, and the sample was diluted to 1 ng/μl with sterile water. Using diluted genomic DNA as a template and selecting sequencing regions, specific primers with barcode are used, corresponding to the following regions: 16S V3-V4(341F(CCTA YGGGRBGCASCAG) and 806R(GGACTACNNGGGTATCTAAT)). After mixing and purifying the PCR products, TruSeq was used ® The DNA PCR Free Sample Preparation Kit was used to construct a library. The constructed library was quantified using Qubit and Q-PCR, and after passing the test, it was sequenced using NovaSeq6000 (17).

2.4 Non-targeted metabolomics

The sample stored at −80 °C refrigerator was thawed on ice and vortexed for 10 s. 50 μL of sample and 300 μL of extraction solution (ACN: Methanol = 1:4, V/V) containing internal standards were added into a 2 mL microcentrifugetube. The sample was vortexed for 3 min and then centrifuged at 12000 rpm for 10 min (4 °C). 200 μL of the supernatant was collected and placed in −20 °C for 30 min, and then centrifuged at 12000 rpm for 3 min (4 °C). A 180 μL aliquots of supernatant were transferred for UHPLC (Vanquish, Thermo Scientific (Massachusetts, USA))-MS (Q Exactive HF-X, Thermo Scientific (Massachusetts, USA)) analysis. Selection of chromatographic columns: Waters ACQUITY Premier HSS T3 Column 1.8 μm, 2.1 mm*100 mm; Retention Time = 6.0 min. The ionization mode is electric spray ionization (ESI). All samples were for two ionization modes (ESI+, ESI-). During the detection process of metabolomics technology, quality control (QC) samples are used for method validation to ensure the stability of the entire analysis system. QC samples are obtained by mixing 100 μL of each sample. To reduce errors, sample testing is conducted randomly. Before analyzing the sample, run the QC sample 5 times to balance the system. During the sample testing process, run QC samples once every 3 normal samples to measure the stability of the system (18).

The raw data of the mass spectrometer was converted into mzXML format by ProteoWizard, and the XCMS program was used to extract and align the overall ion peaks of each substance to obtain the primary spectrum of metabolic ions. Further, the ion peaks of each fragment of metabolic ions were extracted to obtain the secondary spectrum of metabolic ions. Finally, the extracted primary and secondary spectra of metabolic ions were matched with the spectra of metabolites in online public databases, and qualitative information of metabolites was obtained using the metDNA method (19).

2.5 Microbial omics research

Separate each sample data from the offline data based on the barcode sequence and PCR amplification primer sequence and remove the barcode and primer sequences. Using Fastp Filter the original reads to obtain high-quality reads. The filtering method is to automatically detect and remove the joint sequence; Remove reads with a base number of 1 or more; Remove reads with low-quality bases (mass value<15) accounting for more than 40%; Deletion with an average mass of less than 20 within the 4 base window interval; Remove the polyG at the end; Delete reads with a length less than 150 bp. High quality dual end reads are concatenated using FLASH to obtain high-quality Tag data. The tag sequence is compared with the species annotation database using vsearch (v2.22.1) to detect chimeric sequences, and finally the chimeric sequences are removed to obtain the final valid data. Calculate alpha diversity and beta diversity analysis using the phylosseq and vegan packages of R software. $p < 0.05$ was considered to have significant. Perform LEfSe analysis on phylum and genera using R software. Phylogenetic Investigation of Communities by Reconstruction of Unobserved States 2 (PICRUSt2) were utilized to perform functional predication of the gut microbiota. Furthermore, Pathways that were significantly different between the Normal group and Osteoporosis group were identified by *t*-test. $p < 0.05$ was considered to have significant (20).

2.6 Metabolomics research

Perform principal component analysis (PCA) on metabolomic data using R software to outline the inherent similarities/dissimilarities within the dataset. Perform orthogonal partial least squares discriminant analysis (OPLS-DA) using qualitative orthogonal projection of metabolomic data onto latent structures and evaluate the quality of the model through model parameters such as Q2, which represents the predictability of the model, and R2, which represents the goodness of the model fit. The 7-fold cross validation method, CV-ANOVA, and permutation test (permutation number = 200) are used to evaluate the predictive performance of the model. The variable importance (VIP) value in the prediction reflects the importance of the terms in the model relative to Y (all responses) and relative to X (prediction). Finally, fold change (FC) and significant *p*-value calculations were performed, and metabolites with VIP > 1 and $p < 0.05$ were considered to have significant differences between groups. Based on the KEGG pathway, it is determined whether differential metabolites are significantly enriched in the KEGG metabolic pathway. The significantly enriched metabolic pathways indicate their significant importance in the biological processes studied, $p < 0.05$ was considered to have significant (21).

2.7 Omics association analysis

Further understand the pathogenesis of osteoporosis patients through multi-omics association analysis. Spearman rank correlation analysis uses Spearman correlation coefficient as an indicator to describe the correlation between two populations and uses rank correlation test to determine whether there is a statistically significant correlation between the two populations. The range of Spearman correlation coefficient [−1, 1], positive values indicating positive correlation and negative values indicating negative correlation. The correlation analysis was calculated using the cor function of R software, and the significance test of the correlation was calculated using the corPvalueStudent function of the WGCNA package in R software. The input differential metabolites (VIP > 1 and $p < 0.05$) are sorted in descending order of VIP, and the metabolites with the highest ranking are selected. Microorganisms are sorted in descending order based on the sum of relative quantitative values in all samples (22). Metabolites are fixed in the top 50, while microorganisms are assumed to be in the top 30. $p < 0.05$ is considered significant.

3 Results

3.1 Estimation of sequencing depth

The 16S rDNA sequencing of 36 samples was based on the NovaSeq6000 sequencing platform. Each sample's Raw Tags are greater than 50,000 reads. Based on noise reduction methods, a total of 1889 microorganisms were obtained (Supplementary Figure S1). The dilution curve shows that the curves of each sample have reached the plateau stage, indicating that the sequencing data volume is reasonable (Supplementary Figure S2A). The ranking richness curve reflects that the richness and evenness of each sample are high (Supplementary Figure S2B). The species accumulation box plot shows that as the sample size increases, species diversity gradually increases, and when the sample size reaches 36, the curve tends to flatten (Supplementary Figure S2C).

3.2 Alpha-diversity

Alpha diversity is used to analyze the diversity of microbial communities within a group. The evaluation of Shannon and Simpson showed significant changes in alpha diversity in osteoporosis group compared to Normal group ($p < 0.05$) (Figures 1A,B). $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$.

3.3 Beta-diversity

β-diversity is a comparative analysis of the composition of different microbial communities. PCA based on Euclidean distance can extract two axes that maximize the differences between samples, thereby reflecting the differences in multidimensional data on a two-dimensional coordinate graph (Figure 2A). The β-diversity index analyzed by Wilcox test showed significant differences between the Normal group and the osteoporosis group ($p < 0.05$) (Figure 2B). $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$.

3.4 Distribution of classification composition of microbial communities in patients

LEfSe is an analytical tool used to discover and interpret biomarkers in high-dimensional data, which can be used to

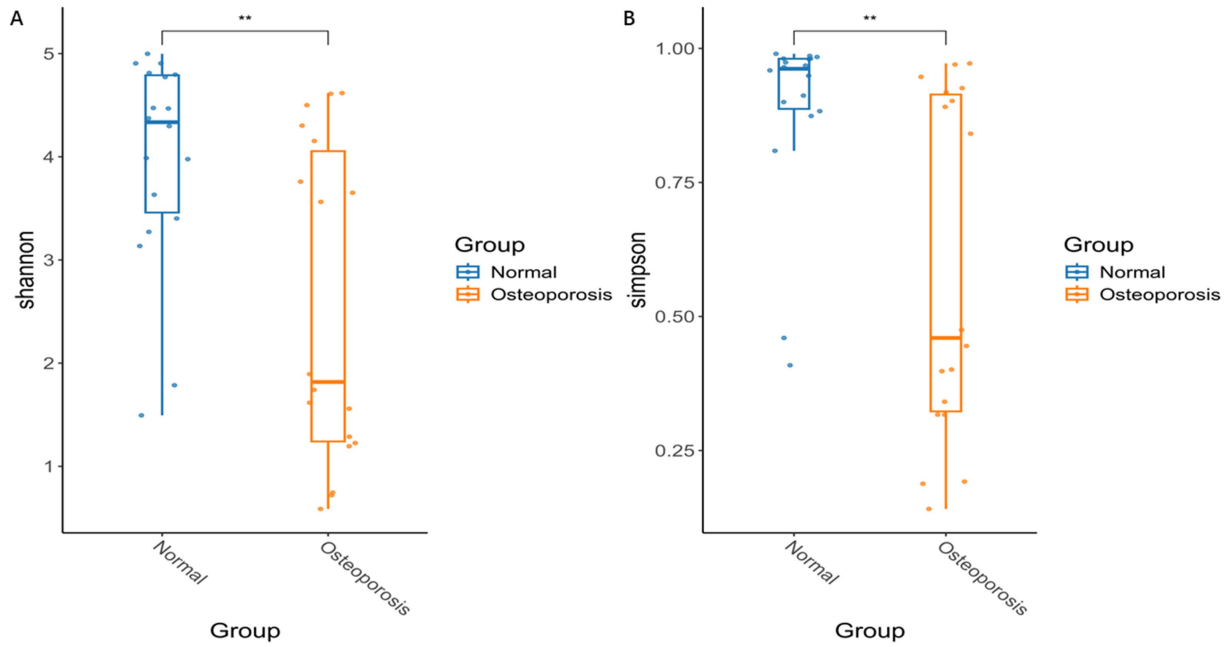


FIGURE 1
Differences in microbial community diversity between Osteoporosis group and Normal group. The evaluation of Shannon and Simpson showed significant changes in alpha diversity in osteoporosis group compared to Normal group (Normal) **(A,B)**. $p < 0.05$ is considered significant. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$.

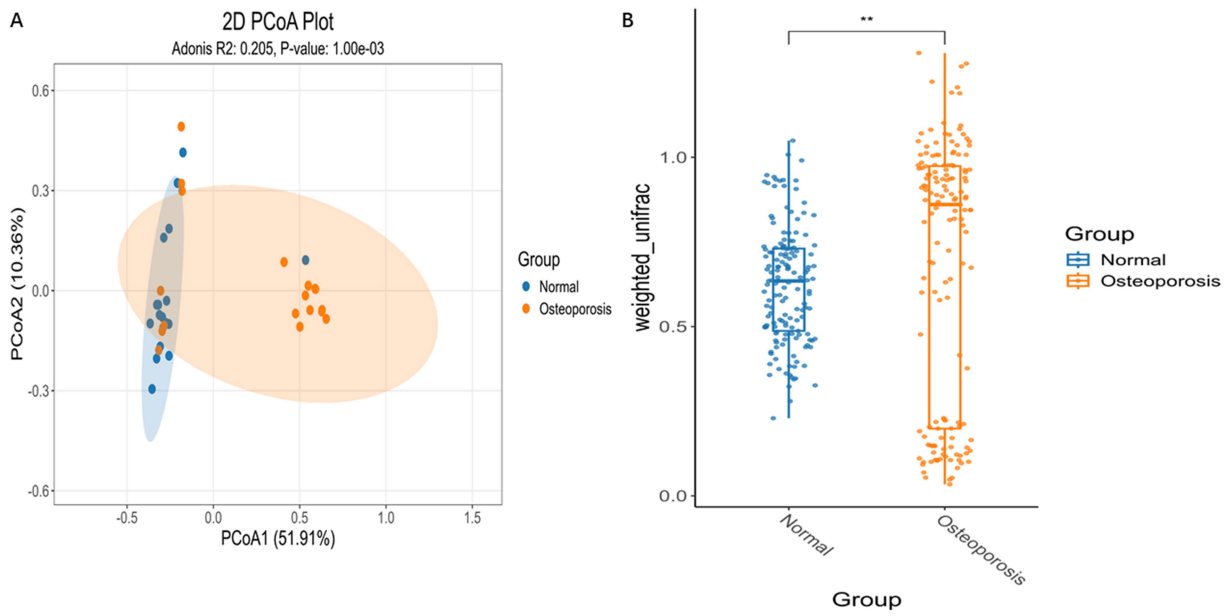


FIGURE 2
Differences in microbial community composition between Osteoporosis group and Normal group. PCA based on Euclidean distance can extract two axes that maximize the differences between samples, thereby reflecting the differences in multidimensional data on a two-dimensional coordinate graph **(A)**. The β-diversity index analyzed by Wilcox test showed significant differences between the Osteoporosis group and Normal group **(B)**. $p < 0.05$ is considered significant. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$.

compare two groups and find biomarkers with statistical differences between groups. As shown in Figures 3A,B, at the phylum level, *Proteobacteria* are mainly present in Osteoporosis; In Normal group, it is mainly *Bacteroidota*. At the genus level, *Cupriavidus* is the main species in Osteoporosis group; In Normal group, the main ones are *Blautia*, *Bacteroides*, *Alcaligenes* and *Pseudomonas*.

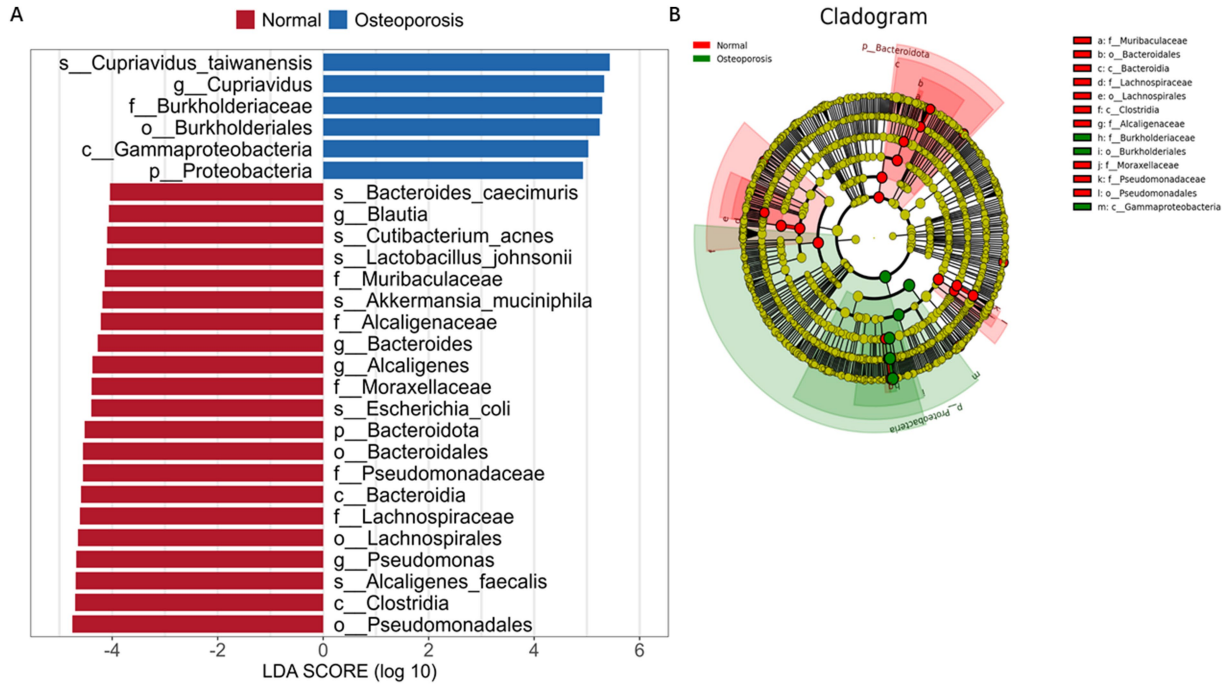


FIGURE 3
Screening of biomarkers with statistical differences in osteoporosis patients. LDA value distribution histogram **(A)**. Cladogram **(B)**.

3.5 PICRUSt2 predicts the functionality of the microbiome

The PICRUSt2 program predicts the functions of two groups (Osteoporosis group and Normal group) of microorganisms. At level 1, microorganisms within the osteoporosis group are involved in the Human Diseases and Cellular Processes ($p < 0.05$) (Supplementary Figure S3A). At level 2, microorganisms within the osteoporosis group are involved in the Amino acid metabolism and Signal transduction ($p < 0.05$) (Supplementary Figure S3B) 0.3.6. Multivariate Analysis of Metabolomics Data.

Based on mass spectrometry analysis of serum samples from 36 participants, the total ion chromatograms (TICs) of different quality control (QC) serum samples were overlaid. The results showed highly overlapping TIC curves for the detected metabolic molecules, indicating strong consistency in peak intensity and retention time of metabolic molecules. This demonstrates excellent instrumental stability when analyzing the same sample at different time points (Supplementary Figures S4A,B). The high stability of mass spectrometry ensures the reliability and reproducibility of the serum metabolomics data.

Multivariate principal component analysis (PCA) of two groups (Osteoporosis group and Normal group) showed significant differences between the Normal group and the Osteoporosis group (Figure 4A). OPLS-DA, a supervised pattern recognition method, was employed to visualize and characterize overall metabolic variations between groups. As shown in Figure 4B, each sample is represented as a point in the score plot, with clear separation between groups. A permutation test ($n = 200$) was conducted to validate the OPLS-DA model (Supplementary Figure S5A). An S-plot was used to identify differential metabolites (Supplementary Figure S5B). In the S-plot, each point represents a variable, and those farther from the origin contribute more significantly to the differences between the Osteoporosis group and Normal group.

3.6 Identification of differential metabolites in serum

Potential differential metabolites were selected based on the VIP derived from the OPLS-DA model and univariate analysis. Screening criteria included VIP > 1.0 and $p < 0.05$. In the Normal group vs. Osteoporosis group, 230 endogenous metabolites with robust differences across the two groups were identified as potential biomarkers (Figure 5A). Display of the top 20 metabolites with different multiples in group comparison (Supplementary Figure S6). The top three metabolites with significant upregulation are 4-Chloroaniline, Oleamide, and 1-Hexadecanoyl-2-docosanoyl-glycero-3-phosphorine. The top three metabolites with significant downregulation are PC(18:3(9Z,12Z,15Z)/18:3(9Z,12Z,15Z)), Astaxanthin, 1,2-Dipalmitoleoyl-sn-glycero-3-phosphoethanolamine.

3.7 Identification of differential metabolic pathways in serum

KEGG enrichment pathway analysis identified key metabolic pathways involved in metabolic reactions. In Normal vs. Osteoporosis, the significant differences in metabolic pathways are mainly enriched in Glycerophospholipid metabolism, Choline metabolism in cancer,

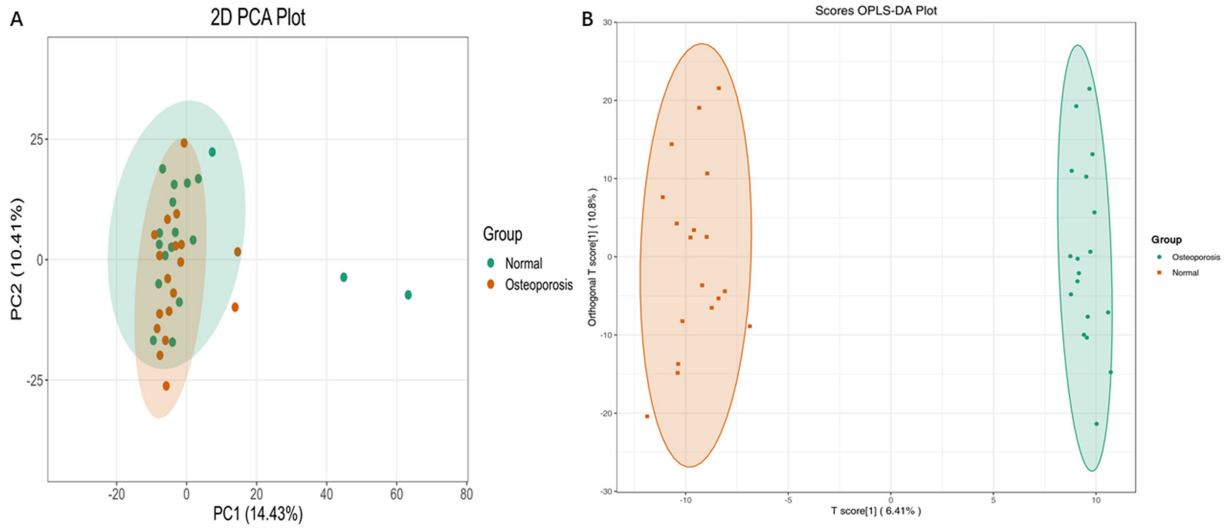


FIGURE 4
Multivariate Analysis of Metabolomics Data. Principal component analysis (PCA) **(A)**. OPLS-DA, a supervised pattern recognition method, was employed to visualize and characterize overall metabolic variations between groups **(B)**.

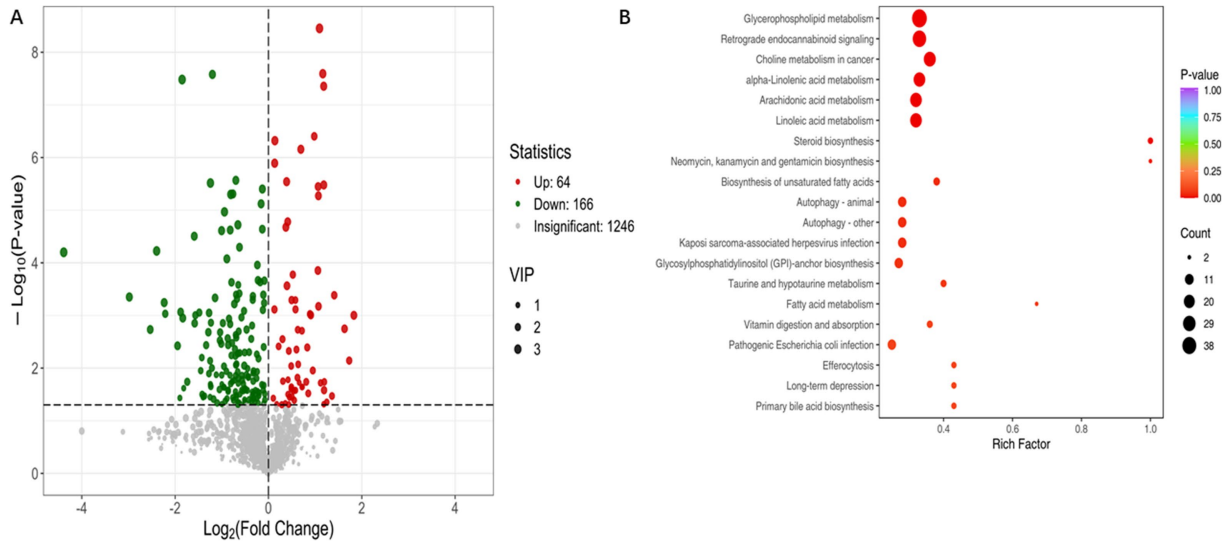


FIGURE 5
Metabolomics analysis is used to explore differential metabolites and biological pathways in osteoporosis patients. In the Normal vs. Osteoporosis, 230 endogenous metabolites with robust differences across the two groups were identified as potential biomarkers **(A)**. KEGG pathway analysis **(B)**.

Linoleic acid metabolism, Arachidonic acid metabolism ($p < 0.05$) (Figure 5B).

3.8 Omics association analysis

By conducting correlation analysis between the microbiome and metabolome, researchers calculated Spearman correlations between the top ranked differential genera and the top 50 differential metabolites in VIP rankings. In Normal group vs. Osteoporosis group, at the phylum level, *Proteobacteria* is significantly positively correlated with differential metabolites (PE-NMe2(20:4(8Z,11Z,14Z,17Z)/(16:0)), etc). At the genus level, *Cupriavidus* is significantly correlated with most of the differential metabolites among the top 50, *Cupriavidus* is significantly positively correlated with differential metabolites (PE-NMe2(20:4(8Z,11Z,14Z,17Z)/(16:0)), PA(22:2)(13Z,16Z)/22:2(13Z,16Z), PE(18:0/20:4(5Z,8Z,11Z,14Z)), etc) among the top 50 (Supplementary Figure S7). $p < 0.05$ is considered significant. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$.

4 Discussion

Although significant progress has been made in the treatment and research of osteoporosis patients in recent years (23), the genetic mechanism of osteoporosis has not yet been fully elucidated. Microorganisms have been reported to play an important role in the pathogenesis of osteoporosis (24). Our research shows that, at the phylum level, *Proteobacteria* are mainly present in Osteoporosis. At the genus level, *Cupriavidus* is the main species in Osteoporosis.

In osteoporosis research, it has been found that the abundance of *Proteobacteria* is negatively correlated with bone mass. When the number of *Proteobacteria* increases, bone mass may decrease, which may indicate that overgrowth or imbalance of *Proteobacteria* is related to the occurrence and development of osteoporosis (25). The increase in abundance of *Proteobacteria* is associated with enhanced inflammatory response. The release of inflammatory factors can affect the activity of osteoclasts and osteoblasts, promote bone resorption, inhibit bone formation and lead to bone loss and osteoporosis (26).

According to reports, *Cupriavidus* is associated with diseases (27). *Cupriavidus* is enriched in patients with high inflammatory response colon cancer (28). *Cupriavidus* is enriched in patients with duodenal bulb inflammation (29). In immune thrombocytopenia patients, *Cupriavidus* is positively correlated with lipid molecules (30). When the level of lipid molecules increases, lipid oxidation and accumulation occur in the bone. Oxidized lipids can inhibit osteoblast formation, induce osteoclast differentiation, and promote bone resorption. In addition, oxidized lipids can induce inflammatory reactions, produce cytokines (such as IL-6, TNF-α, etc.), further activate osteoclasts, lead to increased bone resorption, and inhibit osteoblast differentiation (10, 31). Our research found that *Cupriavidus* is associated with lipid molecules, and *Cupriavidus* may regulate the inflammatory response of osteoporosis patients through lipid molecules. However, the specific mechanism needs to be elucidated in further experiments. The differences in microbial community classification and composition demonstrated in the study provide a theoretical basis for future research that may improve osteoporosis patients.

Non-targeted metabolomics is a quantitative analysis of all endogenous metabolites in an organism, following the research ideas of proteomics and genomics. Metabolites help explain the mechanisms of disease occurrence and development (32). Non-targeted metabolomics analysis based on serum has been applied to identify biomarkers for early disease detection and treatment efficacy prediction, and to explore the pathological mechanisms of diseases in depth (33). In this study, we reported the metabolic profile differences between the Osteoporosis group and Normal group, and conducted multivariate analysis to elucidate the differences among the two groups. The results showed significant changes in the expression levels of metabolites in the two groups and identified the metabolic pathway with significant changes: Glycerophospholipid metabolism. Glycerophospholipid metabolism has been reported to be associated with the pathogenesis of many diseases (34). Glycerophospholipid metabolism metabolism is involved in energy metabolism regulation (35). The process of bone remodeling requires a large amount of energy, and abnormal energy metabolism can affect the activity of osteoblasts and osteoclasts, leading to osteoporosis (36). The oxidation of Glycerophospholipid metabolism can produce reactive oxygen species, triggering oxidative stress (37). During oxidative stress, a large amount of reactive oxygen species (ROS) are generated within cells. ROS can directly activate the nuclear factor kappa B (NF-κB) inflammatory signaling pathway. NF-κB is activated and enters the nucleus, promoting gene transcription of inflammatory factors such as tumor necrosis factor - α and interleukin-6, increasing their expression and release (38). Inflammatory cytokines such as interleukin-6 and tumor necrosis factor-α can promote osteoclast activity, inhibit osteoblast function, and lead to increased bone resorption and decreased bone formation (39). These studies may help to better understand the potential pathogenesis of osteoporosis patients and provide metabolic evidence for further research on osteoporosis patients.

The comprehensive analysis of the microbiome and non-targeted metabolome of diseased individuals has preliminarily revealed the correlation between differential microorganisms and differential metabolites, and indicated the main lipid metabolism pathways. Our multi-omics studies have demonstrated the correlation between differential bacterial genera and metabolites. Although the causes of these differentially expressed metabolites may come from changes in microbial community structure, they may also be related to the lipid metabolism homeostasis caused by the host microbial community (40). More and more evidence suggests that the metabolic products and structural components of microorganisms may promote the pathogenesis of osteoporosis (24). Our study provides aevidence for a deeper understanding of the mechanisms underlying osteoporosis, but significant limitations still exist. The sample size of this study is relatively small, a small sample size may lead to a decrease in the effectiveness of statistical testing, making it difficult to detect real differences or relationships. Insufficient sample size may lead to increased uncertainty in external validity, making it difficult to validate research results in other contexts, and future research with increased sample size is needed to further elucidate the roles of identified factors, lipids, and metabolic pathways in osteoporosis. In the future, targeted metabolomics and animal experiments will be used to conduct more in-depth mechanistic studies. In addition, due to the possibility of confounding variables (including age, race, diet, body mass index, and new drug intake), external validation queues will be needed in the future to validate the current research results.

5 Conclusion

In summary, there are differences in the relative abundance and structural composition of the microbiota in osteoporosis patients compared to Normal group. Understanding the role of microbiota may be helpful in disease mechanism understanding and the identification of biomarkers for diagnosis. Current metabolomics studies have shown identifiable differences in metabolites and lipid metabolism pathways between Osteoporosis group and Normal group. The identified metabolites contribute to the understanding of the pathophysiology of osteoporosis patients. Due to the heterogeneity and complexity of diseases, and with the rapid advancement of various detection technologies, treatment options for diseases have evolved from single target therapy to multi-target therapy. A comprehensive approach (microbiome and Non-targeted metabolomics) can provide multidimensional therapeutic targets for personalized treatment of osteoporosis. However, due to the limitations of the research, the next step requires larger external validation cohorts and an interventional study to confirm the relationships detected and potential biomarkers,

providing more reliable therapeutic targets for personalized treatment of osteoporosis patients.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1321014.

Ethics statement

The studies involving humans were approved by the Ethics Committee of the Hefei First People's Hospital (Approval Number: 2025-106-01). 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

YLiu: Formal analysis, Investigation, Methodology, Validation, Writing – original draft. YLi: Investigation, Methodology, Writing – original draft. JL: Funding acquisition, Writing – original draft.

Funding

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

Conflict of interest

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

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

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

Publisher's note

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/fmed.2025.1664359/full#supplementary-material

References

1. Subarajan P, Arceo-Mendoza RM, Camacho PM. Postmenopausal osteoporosis: a review of latest guidelines. *Endocrinol Metab Clin N Am*. (2024) 53:497–512. doi: 10.1016/j.ecl.2024.08.008

2. Mineta K, Nishisho T, Okada M, Kamada M, Sairyo K. Real-world effects, safety, and predictors of the effectiveness of romosozumab in primary and secondary osteoporosis: an observational study. *Bone*. (2024) 186:117164. doi: 10.1016/j.bone.2024.117164

3. Bhadouria N, Holguin N. Osteoporosis treatments for intervertebral disc degeneration and back pain: a perspective. *JBMR Plus*. (2024) 8:ziae048. doi: 10.1093/jbmrpl/ziae048

4. Johnston CB, Dagar M. Osteoporosis in older adults. *Med Clin North Am*. (2020) 104:873–84. doi: 10.1016/j.mcna.2020.06.004

5. Li C, Zang X, Liu H, Yin S, Cheng X, Zhang W, et al. Olink proteomics for the identification of biomarkers for early diagnosis of postmenopausal osteoporosis. *J Proteome Res*. (2024) 23:4567–78. doi: 10.1021/acs.jproteome.4c00470

6. Seely KD, Kotelko CA, Douglas H, Bealer B, Brooks AE. The human gut microbiota: a key mediator of osteoporosis and Osteogenesis. *Int J Mol Sci*. (2021) 22:9452. doi: 10.3390/ijms22179452

7. Guo M, Liu H, Yu Y, Zhu X, Xie H, Wei C, et al. *Lactobacillus rhamnosus* GG ameliorates osteoporosis in ovariectomized rats by regulating the Th17/Treg balance and gut microbiota structure. *Gut Microbes*. (2023) 15:2190304. doi: 10.1080/19490976.2023.2190304

8. Zhang YW, Wu Y, Liu XF, Chen X, Su JC. Targeting the gut microbiota-related metabolites for osteoporosis: the inextricable connection of gut-bone axis. *Ageing Res Rev*. (2024) 94:102196. doi: 10.1016/j.arr.2024.102196

9. Wang X, Zhang C, Zhao G, Yang K, Tao L. Obesity and lipid metabolism in the development of osteoporosis (review). *Int J Mol Med*. (2024) 54:61. doi: 10.3892/ijmm.2024.5385

10. Zhang J, Hu W, Zou Z, Li Y, Kang F, Li J, et al. The role of lipid metabolism in osteoporosis: clinical implication and cellular mechanism. *Genes Dis*. (2023) 11:101122. doi: 10.1016/j.gendis.2023.101122

11. Lv Z, Shi W, Zhang Q. Role of essential amino acids in age-induced bone loss. *Int J Mol Sci*. (2022) 23:11281. doi: 10.3390/ijms231911281

12. Lai J, Gong L, Liu Y, Zhang X, Liu W, Han M, et al. Associations between gut microbiota and osteoporosis or osteopenia in a cohort of Chinese Han youth. *Sci Rep*. (2024) 14:20948. doi: 10.1038/s41598-024-71731-6

13. Wang Y, Han X, Shi J, Liao Z, Zhang Y, Li Y, et al. Distinct metabolites in osteopenia and osteoporosis: a systematic review and Meta-analysis. *Nutrients*. (2023) 15:4895. doi: 10.3390/nu15234895

14. Lu Y, Cai X, Shi B, Gong H. Gut microbiota, plasma metabolites, and osteoporosis: unraveling links via Mendelian randomization. *Front Microbiol*. (2024) 15:1433892. doi: 10.3389/fmicb.2024.1433892

15. Halik A, Tilgner M, Silva P, Estrada N, Altwasser R, Jahn E, et al. Genomic characterization of AML with aberrations of chromosome 7: a multinational cohort of 519 patients. *J Hematol Oncol*. (2024) 17:70. doi: 10.1186/s13045-024-01590-1

16. Xu Y, Jiang Y, Lu J, Xu C, Li Q, Zhu H. Identification of serum metabolic markers in non-obese hypertensive patients using non-targeted metabolomics. *Sci Rep*. (2025) 15:18320. doi: 10.1038/s41598-025-02162-0

17. Feng X, Zhang Y, Feng J, Li Z, Zhang Z, Zhu L, et al. Exploring changes in metabolites and fecal microbiota of advanced gastric cancer based on plasma metabolomics and 16S rDNA sequencing. *Heliyon*. (2025) 11:e41715. doi: 10.1016/j.heliyon.2025.e41715

18. Zhao X, Liu X, Liu L, Chen R. Serum metabolome alterations in hyperhomocysteinemia based on targeted and non-targeted MS-platforms. *J Chromatogr B Analyt Technol Biomed Life Sci*. (2024) 1247:124336. doi: 10.1016/j.jchromb.2024.124336

19. Cai X, Pan S, Li M, Lu P, Guo X, Zheng S. Non-targeted metabolomics analysis of mother and infant in gestational diabetes mellitus and neonatal clinical characterization. *Clin Lab*. (2024) 70:527. doi: 10.7754/Clin.Lab.2023.230527

20. Yan S, Du R, Yao W, et al. Host-microbe interaction-mediated resistance to DSS-induced inflammatory enteritis in sheep. *Microbiome*. (2024) 12:208. doi: 10.1186/s40168-024-01932-8

21. Yang WY, Wang J, Li XH, Xu B, Yang YW, Yu L, et al. Analysis of non-targeted serum metabolomics in patients with chronic kidney disease and hyperuricemia. *Biotechnol Genet Eng Rev*. (2024) 40:4013–39. doi: 10.1080/02648725.2023.2204715

22. Yuan Y, Wang C, Wang G, Guo X, Jiang S, Zuo X, et al. Airway microbiome and serum metabolomics analysis identify differential candidate biomarkers in allergic rhinitis. *Front Immunol*. (2022) 12:771136. doi: 10.3389/fimmu.2021.771136

23. Ambrosio MR, Cattaneo CA, Gagliardi I, Carnevale A, Zatelli MC. Aetiology, diagnosis and treatment of thalassemia-associated osteoporosis of the adult. *J Endocrinol Investig*. (2025) 48:799–815. doi: 10.1007/s40618-024-02503-2

24. Li S, Zhang Y, Ding S, Chang J, Liu G, Hu S. Curcumin ameliorated glucocorticoid-induced osteoporosis while modulating the gut microbiota and serum metabolome. *J Agric Food Chem*. (2025) 73:8254–76. doi: 10.1021/acs.jafc.4c06689

25. Aurigemma NC, Koltun KJ, VanEvery H, Rogers CJ, de Souza MJ. Linking the gut microbiota to bone health in anorexia nervosa. *Curr Osteoporos Rep*. (2018) 16:65–75. doi: 10.1007/s11914-018-0420-5

26. Guss JD, Horsfield MW, Fontenele FF, Sandoval TN, Luna M, Apoorva F, et al. Alterations to the gut microbiome impair bone strength and tissue material properties. *J Bone Miner Res*. (2017) 32:1343–53. doi: 10.1002/jbmr.3114

27. Sung JJY, Coker OO, Chu E, Szeto CH, Luk STY, Lau HCH, et al. Gastric microbes associated with gastric inflammation, atrophy and intestinal metaplasia 1 year after *Helicobacter pylori* eradication. *Gut*. (2020) 69:1572–81. doi: 10.1136/gutjnl-2019-319826

28. Cao Y, Zheng X, Hu Y, Li J, Huang B, Zhao N, et al. Levels of systemic inflammation response index are correlated with tumor-associated bacteria in colorectal cancer. *Cell Death Dis*. (2023) 14:69. doi: 10.1038/s41419-023-05602-9

29. Fan HN, Zhu P, Zhang J, Zhu JS. Mucosal microbiome dysbiosis associated with duodenum bulb inflammation. *Microb Pathog*. (2021) 150:104711. doi: 10.1016/j.micpath.2020.104711

30. Yu X, Zheng Q, He Y, Yu D, Chang G, Chen C, et al. Associations of gut microbiota and fatty metabolism with immune thrombocytopenia. *Front Med*. (2022) 9:810612. doi: 10.3389/fmed.2022.810612

31. Chu W, Peng W, Lu Y, Liu Y, Li Q, Wang H, et al. PRMT6 epigenetically drives metabolic switch from fatty acid oxidation toward glycolysis and promotes osteoclast differentiation during osteoporosis. *Adv Sci*. (2024) 11:e2403177. doi: 10.1002/advs.202403177

32. Li J, Cheng B, Xie H, Zhan C, Li S, Bai P. Bladder cancer biomarker screening based on non-targeted urine metabolomics. *Int Urol Nephrol*. (2022) 54:23–9. doi: 10.1007/s11255-021-03080-6

33. Mu X, Ji C, Wang Q, Liu K, Hao X, Zhang G, et al. Non-targeted metabolomics reveals diagnostic biomarker in the tongue coating of patients with chronic gastritis. *J Pharm Biomed Anal*. (2019) 174:541–51. doi: 10.1016/j.jpba.2019.06.025

34. Shan S, Wu C, Shi J, Zhang X, Niu J, Li H, et al. Inhibitory effects of peroxidase from foxtail millet bran on colitis-associated colorectal carcinogenesis by the blockage of glycerophospholipid metabolism. *J Agric Food Chem*. (2020) 68:8295–307. doi: 10.1021/acs.jafc.0c03257

35. Chen X, Liu Y, Pu J, Gui S, Wang D, Zhong X, et al. Proteomics reveals mitochondrial dysfunction and energy metabolism disturbance of intestine in a nonhuman primate model of depression. *J Affect Disord*. (2023) 333:562–70. doi: 10.1016/j.jad.2023.04.031

36. Karsenty G, Khosla S. The crosstalk between bone remodeling and energy metabolism: a translational perspective. *Cell Metab*. (2022) 34:805–17. doi: 10.1016/j.cmet.2022.04.010

37. Li W, He E, Zhang P, Li Y, Qiu H. Multiomics analyses uncover nanoceria triggered oxidative injury and nutrient imbalance in earthworm *Eisenia fetida*. *J Hazard Mater*. (2022) 437:129354. doi: 10.1016/j.jhazmat.2022.129354

38. Yildirim C, Cangi S, Orkmez M, Yilmaz SG, Bozdayı MA, Yamaner H, et al. Sinapic acid attenuated cisplatin-induced cardiotoxicity by inhibiting oxidative stress and İnflammation with GPX4-mediated NF-kB modulation. *Cardiovasc Toxicol*. (2023) 23:10–22. doi: 10.1007/s12012-022-09773-3

39. Zhivodernikov IV, Kirichenko TV, Markina YV, Postnov AY, Markin AM. Molecular and cellular mechanisms of osteoporosis. *Int J Mol Sci*. (2023) 24:15772. doi: 10.3390/ijms242115772

40. Qin M, Huang Z, Huang Y, Huang X, Chen C, Wu Y, et al. Association analysis of gut microbiota with LDL-C metabolism and microbial pathogenicity in colorectal cancer patients. *Lipids Health Dis*. (2024) 23:367. doi: 10.1186/s12944-024-02333-4

frontiers | Frontiers in Systems Biology

TYPE Original Research
PUBLISHED 29 September 2025
DOI 10.3389/fsysb.2025.1649758

Check for updates

OPEN ACCESS

EDITED BY
HaiHui Huang,
Shaoguan University, China

REVIEWED BY
Xi Long,
Hong Kong University of Science and Technology, Hong Kong SAR, China
Eko Mugiyanto,
Muhammadiyah University of Pekajangan Pekalongan, Indonesia

*CORRESPONDENCE
Adrian Gu,
adrgu666@gmail.com

RECEIVED 19 June 2025
ACCEPTED 15 September 2025
PUBLISHED 29 September 2025

CITATION
Gu A and Chen JY (2025) GETgene-AI: a framework for prioritizing actionable cancer drug targets.
Front. Syst. Biol. 5:1649758.
doi: 10.3389/fsysb.2025.1649758

COPYRIGHT
© 2025 Gu and Chen. 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.

GETgene-AI: a framework for prioritizing actionable cancer drug targets

Adrian Gu[1]* and Jake Y. Chen[1,2]

[1]AlphaMind Club, Birmingham, AL, United States, [2]Systems Pharmacology AI Research Center, School of Medicine, The University of Alabama at Birmingham, Birmingham, AL, United States

Prioritizing actionable drug targets is a critical challenge in cancer research, where high-dimensional genomic data and the complexity of tumor biology often hinder effective prioritization. To address this, we developed GETgene-AI, a novel computational framework that integrates network-based prioritization, machine learning, and automated literature analysis to prioritize and rank potential therapeutic targets. Central to GETgene-AI is the G.E.T. strategy, which combines three data streams: mutational frequency (G List), differential expression (E List), and known drug targets (T List). These components are iteratively refined and ranked using the Biological Entity Expansion and Ranking Engine (BEERE), leveraging protein-protein interaction networks, functional annotations, and experimental evidence. Additionally, GETgene-AI incorporates GPT-4o, an advanced large language model, to automate literature-based ranking, reducing manual curation and increasing efficiency. In this study, we applied GETgene-AI to pancreatic cancer as a case study. The framework successfully prioritized high-priority targets such as PIK3CA and PRKCA, validated through experimental evidence and clinical relevance. Benchmarking against GEO2R and STRING demonstrated GETgene-AI's superior performance, achieving higher precision, recall, and efficiency in prioritizing actionable targets. Moreover, the framework mitigated false positives by deprioritizing genes lacking functional or clinical significance. While demonstrated on pancreatic cancer, the modular design of GETgene-AI enables scalability across diverse cancers and diseases. By integrating multi-omics datasets with advanced computational and AI-driven approaches, GETgene-AI provides a versatile and robust platform for accelerating cancer drug discovery. This framework bridges computational innovations with translational research to improve patient outcomes.

KEYWORDS

cancer, pancreatic cancer, network-based prioritization, computational biology and bioinformatics, drug target prioritization, drug target, network biology, gene prioritarization

1 Introduction

Traditional chemotherapeutic agents, which non-specifically target rapidly dividing cells (Gu et al., 2023; Sun et al., 2021), are contested with the promise of targeted therapies that disrupt specific molecular pathways governing cell survival and apoptosis (Sellers and Fisher, 1999; Lim et al., 2019). Drug target discovery is pivotal for advancing cancer therapies, yet traditional approaches face three critical limitations. First, manual curation of literature and static biomedical databases struggles to scale with the complexity of modern

multi-omics data (genomic, transcriptomic, proteomic), leading to incomplete or outdated target identification (Paananen and Fortino, 2020; Zhou et al., 2022; Trajanoska et al., 2023; Lindsay, 2003; Zhou and Zhong, 2017). Second, traditional network-based prioritization, which prioritize genes based on protein-protein interaction (PPI) network centrality, oversimplify biological context by ignoring tissue-specific genomic features such as mutation frequencies and differential expression profiles (Petti et al., 2020). These limitations contribute to high failure rates in translating preclinical discoveries to clinical therapies, particularly in genetically heterogeneous cancers like pancreatic cancer. Third, reliance on single-metric approaches like fold change or mutational frequency introduces variability due to arbitrary thresholds and sample bias (McCarthy and Smyth, 2009; Dinstag and Shamir, 2020; López-Cortés et al., 2018). These gaps contribute to high failure rates in translating preclinical discoveries to clinical therapies, particularly in genetically heterogeneous cancers like pancreatic cancer (Singh et al., 2023; Sun et al., 2022; Zhu et al., 2021; Somarelli et al., 2019).

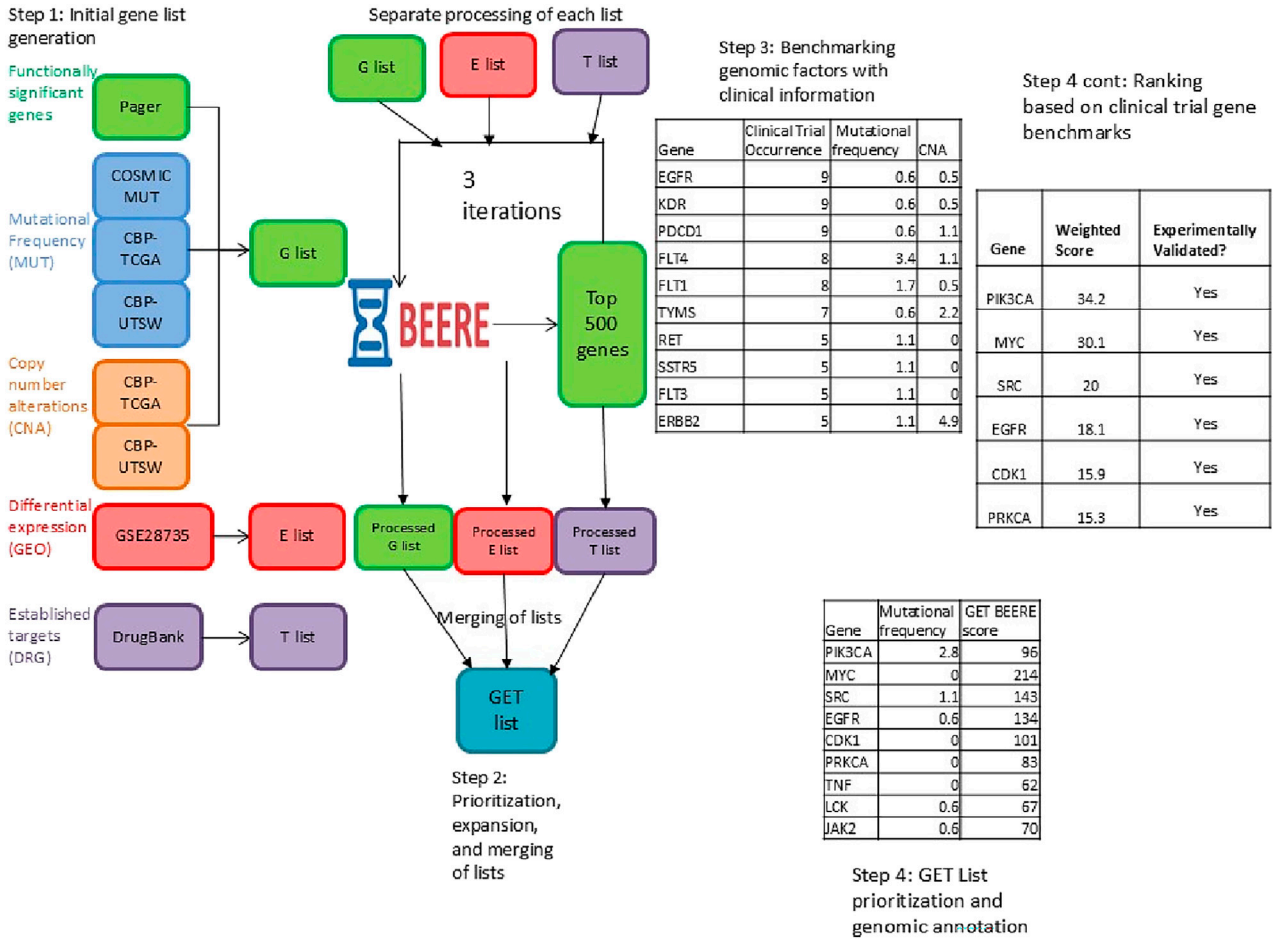
Computational advances address these challenges by integrating multi-omics data, network-based prioritization and AI-driven literature review, driving down costs, increasing precision, and expediting the development of effective therapies through *in silico* assessments (Sadybekov and Katritch, 2023; Sliwoski et al., 2014; Huan et al., 2010; Chen et al., 2006). The integration of multi-omics data contextualizes mutations within tissue-specific expression patterns, while network-based prioritization refines prioritization by mapping genes to functionally relevant pathways (Shim et al., 2015). Network-based prioritization enables researchers to analyze genomic datasets and identify critical regulatory genes implicated in cancer development (Chang et al., 2021; Sonehara and Okada, 2021). These methods prioritize disease-related genes by integrating data from PPI networks and known gene-drug associations (Mohsen et al., 2021; Zhang et al., 2021). Furthermore, network-based prioritization approaches provide the ability to efficiently process genomic information and derive meaningful insights is pivotal for identifying and visualizing relevant drug targets (Chen et al., 2013; Chen et al., 2009; Huan et al., 2010; Shim et al., 2015; Huang et al., 2012).

Differential gene expression is a critical method for identifying genes significantly altered between conditions, such as cancerous versus normal tissues (Bai et al., 2013; Van de Sande et al., 2023). A common approach involves calculating "fold change," which quantifies the ratio of gene expression levels between these states (Love et al., 2014; Mutch et al., 2002). GEO2R, a tool to determine differentially expressed genes, utilizes fold change to rank genes under experimental conditions (ie. tumor versus healthy tissue comparisons) (Barrett et al., 2013). However, the arbitrary selection of fold change thresholds can introduce variability into prioritization, compromising the reliability of target identification (McCarthy and Smyth, 2009). Separately, frequency-based prioritization methods focus on genes with elevated mutational rates in disease contexts, hypothesizing these as common therapeutic targets (Dinstag and Shamir, 2020; López-Cortés et al., 2018). Frequency-based prioritization methods for gene prioritization can be prone to bias, especially due to sample selection, which can skew results (Lazzeroni et al., 2014). To address these limitations, network centrality-based prioritization has emerged as a complementary strategy. This approach leverages gene connectivity within biological networks, offering a holistic framework for target selection by expanding gene lists and strengthening disease association metrics (Janyasupab et al., 2021; Magger et al., 2012).

Concurrently, AI-driven literature review (e.g., GPT-4) automates the synthesis of preclinical and clinical evidence, identifying targets with mechanistic and translational relevance (Liu et al., 2021; Oniani et al., 2024; Sallam, 2023; Tripathi et al., 2024). By combining these approaches, biases inherent to single-metric or fragmented datasets can be mitigated, yielding prioritized targets with mechanistic, functional, and translational relevance. (Somarelli et al., 2019; Zhu et al., 2021; Sadybekov and Katritch, 2023). LLMs can predict essential information about gene targets, including structural domains of proteins, protein structure, toxicity and adverse effects, functional significance, clinical and preclinical relevance, and treatment efficacy (Sallam, 2023; Tripathi et al., 2024). Furthermore, GPT-4 has demonstrated the ability to rival human performance in conducting literature reviews, thus streamlining the drug target prioritization process (Khraisha et al., 2024; Li et al., 2010).

In this study, we hypothesize that the utilization of network-based analysis, artificial intelligence, and biologically significant data will enable systemic prioritization of actionable therapeutic targets. Thus, we propose GETgene-AI, a framework which annotates network-based analysis with LLM enabled literature review, and biologically significant data. Central to GETgene-AI is the G.E.T. strategy, which integrates three key data streams: the G List (genes with genetic mutations, variations functionally implicated in genotype-to-phenotype association studies of the disease), the E List (disease target tissue-specific expressions of the candidate gene), and the T List (established drug targets based on reports from literature, patents, clinical trials, or existing approved drugs). Initial gene candidates are derived from heterogeneous biological datasets, including fold change, copy number alterations, and mutational frequency metrics. To mitigate biases inherent to fragmented or incomplete data, GETgene-AI employs a multi-dataset integration approach. The framework iteratively refines candidate lists through the network-based tool BEERE, which annotates and prioritizes genes with network-based centrality methods to create a high-quality, prioritized gene list. This iterative process expands and ranks candidates by evaluating their biological relevance, network centrality, and concordance with genomic aberrations, thereby improving target identification accuracy. GPT-4o is integrated into the process to improve literature review efficiency and further annotate the target list, enhancing the overall workflow. By combining traditional and *in silico* methods, GETgene-AI bridges gaps in drug discovery and facilitates the development of personalized cancer therapies.

The novel drug targets prioritized through our case study in pancreatic cancer not only offer insights into the unique molecular mechanisms driving this aggressive cancer but also present promising avenues for therapeutic intervention. While pancreatic cancer serves as a case study in this paper, the underlying methodology is adaptable to a wide range of cancers and diseases, thereby accelerating the discovery of therapeutic options.



Gene	Clinical Trial Occurrence	Mutational frequency	CNA
EGFR	9	0.6	0.5
KDR	9	0.6	0.5
PDCD1	9	0.6	1.1
FLT4	8	3.4	1.1
FLT1	8	1.7	0.5
TYMS	7	0.6	2.2
RET	5	1.1	0
SSTR5	5	1.1	0
FLT3	5	1.1	0
ERBB2	5	1.1	4.9

Gene	Weighted Score	Experimentally Validated?
PIK3CA	34.2	Yes
MYC	30.1	Yes
SRC	20	Yes
EGFR	18.1	Yes
CDK1	15.9	Yes
PRKCA	15.3	Yes

Gene	Mutational frequency	GET BEERE score
PIK3CA	2.8	96
MYC	0	214
SRC	1.1	143
EGFR	0.6	134
CDK1	0	101
PRKCA	0	83
TNF	0	62
LCK	0.6	67
JAK2	0.6	70

FIGURE 1
General overview of the GET list compilation and ranking process. Initial gene lists from each of the three subsets are compiled. 2,493 genes are compiled in the initial G list, 2000 genes are compiled in the initial E list, and 131 genes are compiled in the initial T list. Each list is iteratively prioritized using the BEERE network ranking and expansion tool, taking the top 500 genes each time and re expanding and ranking. The lists were then merged and annotated with biologically significant features. Separately, genes implicated in clinical trials related to treatment of pancreatic cancer were benchmarked to set the weights utilized for RP score ranking. Genes in the GET list were then ranked utilizing these weights.

2 Methods

In Figure 1, we show a general overview of the GETgene-AI framework.

The initial gene list is generated by employing a three-tiered strategy—comprising the Gene list (G list), Expression list (E list), and Target list (T list)—to integrate biological context into gene prioritization. The G list identifies genes with high mutational frequency, functional significance (e.g., pathway enrichment via the Kyoto Encyclopedia of Genes and Genomes (KEGG)), and genotype-phenotype associations. The E list focuses on genes exhibiting significant differential expression in pancreatic ductal adenocarcinoma (PDAC) compared to normal tissues, while the T list incorporates genes annotated as drug targets in clinical trials, patents, or approved therapies. To construct these lists, disease-specific genomic data were aggregated from public databases (e.g., TCGA, COSMIC, PAGER) and processed using GRIPPs (Gong and Chen, 2023), an iterative network-based approach that applies modality-specific thresholds to ensure robust inclusion criteria.

Following the initial gene list generation, the second step involves prioritizing and expanding these lists using the BEERE network-ranking tool. BEERE was selected for its demonstrated efficacy in filtering low-confidence data and enhancing prioritization accuracy (Yue et al., 2019), ensuring comprehensive and reliable gene sets.

A benchmark set of genes implicated in pancreatic cancer clinical trials (i.e., genes appearing as targets or biomarkers in registered interventional studies) was analyzed to evaluate which genomic and network features are most characteristic of clinically successful drug targets. This benchmark set is distinct from the T list, which consists only of genes targeted by FDA-approved drugs already indicated for pancreatic cancer. Genomic features considered included differential expression, mutation frequency, and copy number alterations, while network-based features included the BEERE scores of Gene, Expression, and Target lists. The benchmarking analysis did not alter the composition or scoring of the T list but instead provided interpretive context by identifying which factors were enriched among clinically validated targets. This analysis was further supplemented by a GPT-4–enabled literature

review, which added biological and clinical insights to the interpretation of results.

Finally, the GETgene-AI ranking is generated by integrating BEERE network rankings, annotated gene information, and insights derived from GPT-4. This multi-layered approach ensures a robust and contextually informed prioritization of potential drug targets.

Using PDAC as a case study—selected due to its poor prognosis and limited therapeutic options (Hu et al., 2021) —our framework produced quantitative data and novel insights into potential therapeutic targets, demonstrating its utility in advancing precision oncology.

2.1 Initial gene list generation

2.1.1 Compiling the gene list from genetic mutations

For the "GENE" component of our "GET" framework, we compiled three gene subsets: PAGER-NC, COSMIC-MUT, and CBP-CNA-MUT. The initial "GENE" list was compiled from the PAGER (Huang et al., 2012; Yue et al., 2018; 2022), cBioPortal (de Bruijn et al., 2023), and COSMIC (Tate et al., 2019) databases. To address potential sample biases and data incompleteness (e.g., studies failing to detect specific genes), we incorporated multiple datasets from these repositories when available. Genes associated with the term "Pancreatic Cancer" were manually curated from these databases. Empirical cutoffs were applied to prioritize genes with relevance to pancreatic cancer.

To integrate biological pathway context into gene prioritization, we utilized PAGER (Chowbina et al., 2009), which quantifies functional significance through pathway-based metrics. From PAGER, 844 candidate genes were selected heuristically using an nCoCo score threshold between 5 and 100. The nCoCo score, which measures gene set coherence by integrating co-citation and pathway data, with higher scores indicating stronger biological cohesion was constrained with a minimum of 5 (minimal coherence) and maximum of 100 (ubiquitous processes) (Huang et al., 2012; Yue et al., 2018; Yue et al., 2022).

For the cBioPortal and COSMIC databases, thresholds were defined by identifying points where mutational frequency no longer demonstrated cancer-specific significance in prior studies. From cBioPortal, 1,000 genes were selected using cutoffs of 8.2% for copy number alterations (CNA) and 2.8% for mutational frequency. The threshold for copy number alterations is significantly higher due to only 21 sets of copy number signitures being represented in 97% of tumor samples on The Cancer Genome Atlas (Steele et al., 2022). The 2.8% cutoff for mutational frequency is due to the fact that a limited amount of genes were found to be mutated in more than 5% of tumors (Sinkala, 2023). Most biologically relevant genes were found to be mutated at frequencies between 2%–20% (Lawrence et al., 2014). From COSMIC, 649 genes were compiled using a 20% mutational frequency cutoff according to the previously mentioned frequency range. Finally, candidate genes from PAGER, cBioPortal, and COSMIC were aggregated to form the "G list", comprising 2,493 genes in total.

Sensitivity analysis was performed by testing lower and higher cutoffs for both CNA and mutational frequency. For CNA, a lower threshold of 7.3% and a higher threshold of 9.2% were applied, while for mutational frequency, thresholds of 2.2% (lower) and 3.4% (higher) were used. For the COSMIC cancer database, a lower cutoff of 15% and a higher cutoff of 25% were applied. Genes within the top 250 of GETgene-AI were manually examined to identify those included or functionally related to genes falling within the lower and higher thresholds. The lower threshold did not identify any genes beyond those already present in the G list, whereas the higher threshold excluded the following genes: P3H2, P4HTM, PLOD3, PLOD2, P4HA1, PLOD1, PAM, PSMB5, C1QC, C1QA, and C1QB. All of these genes rank outside the top 150.

2.1.2 Compiling candidate genes for the "expression" subset

Candidate genes were prioritized by analyzing the GEO dataset GSE29735, titled "Pancreatic ductal adenocarcinoma tumor and adjacent non-tumor tissue" (Zhang et al., 2012; Zhang et al., 2013), using the GEO2R tool. Samples were categorized into tumor and non-tumor groups via the "Define groups" feature, with the tumor group defined as "human pancreatic tumor tissue patient samples" and the non-tumor group as "human pancreatic non-tumor patient samples". The dataset comprised of 90 patient samples, evenly distributed between 45 tumor and 45 non-tumor samples. Differentially gene expression analysis was performed using GEO2R's "analyze" function. The top 2,504 genes exhibiting logfc values over 0.25 were compiled into an initial "E list". A cutoff of 0.25 was determined based on the "FindAllMarker" function provided by the R package Seurat (Wang et al., 2024). The list was subsequently processed iteratively using the BEERE software in accordance with the GRIPPs method.

2.1.3 Compiling candidate genes for the "Target" subset

Incorporating pharmacology data with network-based prioritization is a well established approach (Huang et al., 2015; Huang et al., 2012b). Building on this methodolody, a set of 131 genes were identified using DrugBank (Wishart et al., 2018), a comprehensive drug and drug-target database. To extract relevant genes, the database was queried using the search terms "Pancreatic Cancer," "Pancreatic Ductal Adenocarcinoma," and "Neuroendocrine Pancreatic Cancer" within its drug repository. Drugs explicitly indicated for Pancreatic Cancer treatment were identified by reviewing their associated metadata, including summaries, background descriptions, indications, clinical trial references, and listed "Associated Conditions." Each drug's mechanism of action, therapeutic summary, and clinical trial references were manually evaluated to distinguish agents directly treating pancreatic cancer from those used for supportive care (e.g., chemotherapy relief, pain management, or sedation). For all drugs meeting the inclusion criteria, gene targets listed under their respective "Targets" section in DrugBank were compiled, resulting in 131 unique genes associated with pancreatic cancer therapeutics.

2.2 Prioritization and expansion of GET lists

To improve the specificity and biological relevance of our candidate gene lists, we implemented an iterative refinement process using the BEERE tool for prioritization and network-based expansion. The BEERE tool employs an initial ranking

algorithm and two iterative ranking algorithms—PageRank and an ant-colony algorithm—both of which have demonstrated success across diverse knowledge domains (Yue et al., 2019). Although both ranking algorithms use an iterative ranking process, they differ in how node importance weights are calculated. The PageRank algorithm assigns node importance directly from neighboring nodes. In the ant-colony algorithm nodes lose score when disseminating information and gain score upon receiving it. BEERE expands the gene list using the nearest-neighbor network constructed from protein-protein interactions in the HAPPI 2.0 database (Chen et al., 2009; 2017; Wu et al., 2012).

This workflow addresses the inherent limitations of single-dimensional analyses (e.g., relying solely on mutation or expression data) by integrating complementary biological evidence. Building on the GRIPPS framework (Gong and Chen, 2023), we developed a customized pipeline to systematically prioritize genes from three distinct categories: the combined GET list (genes ranked by aggregated mutational frequency, differential expression, and known drug-target status), the GT list (genes co-occurring in mutation and drug-target databases to highlight functionally relevant drivers), and a prioritized Expression (E) list (genes ranked exclusively by differential expression in pancreatic cancer).

The GET, GT, and E lists are expanded independently to preserve modality-specific signal during the BEERE prioritization phase. Combining them before expansion would dilute distinct biological features (e.g., mutation-specific drivers in G vs. expression-based biomarkers in E) and bias the expansion toward categories with larger initial representation, potentially overshadowing rare but high-impact genes. For example, MYC and TNF, identified through differential expression and drug-target overlap but not mutational frequency, would have been deprioritized if lists were merged prior to expansion. This systematic, modality-preserving approach enhanced the identification of potential therapeutic targets by ensuring that candidates from each evidence stream were equally represented in the final prioritization.

Each list underwent the same refinement workflow to balance comprehensiveness with specificity. First, BEERE expanded the initial gene sets by incorporating proximal interactors from protein-protein interaction (PPI) networks in the HAPPI 2.0 database, thereby capturing functionally related genes beyond those directly identified in our initial screens. Next, BEERE's network propagation and statistical ranking algorithms prioritized genes based on their network centrality and significance scores. To prevent overexpansion and maintain focus on high-confidence candidates, we empirically filtered each list to retain the top 500 genes after each prioritization cycle. This iterative process was repeated three times, as preliminary testing revealed that additional iterations caused excessive convergence of the lists, reducing their distinct biological relevance. Three iterations optimally preserved the unique profiles of each list while still enabling meaningful integration.

The independently expanded GET, GT, and E lists (each refined through three iterations of BEERE network expansion) were consolidated into an Initial GET List, which then underwent a final BEERE-based prioritization to generate the Final GET List. For comparative analysis, we also retained the previously defined Expression List (top differentially expressed genes) and the GT List (prioritized genes from mutation–drug target overlaps). These lists were not re-derived here but carried forward for side-by-side evaluation. This tiered approach ensured that our final candidate pool retained both mechanistic diversity (genes linked to distinct biological processes) and clinical relevance (genes with actionable potential as drug targets).

The refinement process was critical to address three key challenges: (1) mitigating the high false-positive rate inherent to mutational and expression screens in heterogeneous cancers like pancreatic adenocarcinoma, (2) reconciling discrepancies between genes prioritized by individual data types (e.g., highly mutated genes often lack expression changes, and *vice versa*), and (3) ensuring functional coherence by embedding candidates within PPI networks reflective of disease biology. By iteratively refining lists through network propagation and multi-evidence integration, we enhanced the biological plausibility of candidates while preserving distinct mechanistic hypotheses for downstream validation.

2.3 GPT-4o aided literature assessment

Recent research has demonstrated that GPT-4o performs "human-like" literature reviews, particularly in screening and analyzing scientific literature (Khraisha et al., 2024). For this study, abstracts related to pancreatic cancer genes and treatments were downloaded using PubMed's "save" feature. A total of 5,091 abstracts were collected and uploaded for analysis by GPT-4o through a custom GPTo interface. Due to the data processing limitations of GPT-4o, abstracts were filtered to include only meta-analyses, clinical trials, and systematic reviews on PubMed to ensure high-quality input data.

The custom GPTo model was configured with specific instructions to rank genes based on a scoring system with a maximum score of 400 points, distributed across four categories: functional significance in pancreatic cancer, research popularity, treatment effectiveness when targeting or inhibiting the gene, and protein structure. Each category was allocated 100 points, and the resulting metric was termed the GPT-4 score. To mitigate GPT-4o′s known issue of "hallucination" or the generation of inaccurate or nonexistent information, the model was explicitly instructed to base its rankings solely on the uploaded research database. Additionally, the model was required to cite articles referenced during the ranking process and provide explanations for the scores assigned to each gene in every category. GPT-4 outputs were manually verified against curated datasets to ensure biological relevance and mitigate hallucinations. Citations provided by GPT-4 were cross-referenced with PubMed to confirm validity. All cited articles were manually verified, and any errors or hallucinations were addressed by instructing the model to re-search the uploaded literature database for accurate mentions of the gene. Analyses involving database-derived information was performed on static datasets downloaded, ensuring that any subsequent database changes would not affect our reported results. Where possible, we provide accession numbers and dataset DOIs. This approach guarantees that the gene rankings and annotations presented here can be reproduced independently of future GPT-4 updates or changes to online resources.

2.4 Incorporation of clinically implicated genes and annotation of genes with factors relevant to drug target prioritization

Clinical trials are critical for evaluating the efficacy of therapeutic agents targeting specific genes. To assess the clinical relevance of prioritized genes, we quantified clinical trial activity by compiling the frequency of trials associated with each gene. Genes targeted by drugs investigated in pancreatic cancer treatment trials systemically identified through the following process: A search for the term "pancreatic cancer" was conducted on Clinicaltrials.gov, and all drugs listed in active or completed interventional trials for pancreatic cancer were extracted. Corresponding target genes for these drugs were then identified using DrugBank's "Targets" section, which provides genes targeted by the drug for pancreatic cancer treatment. This process yielded 357 drugs targeting 253 unique genes. These genes were annotated with BEERE scores derived from the previously described GET lists. To enhance biological validity, the analysis integrated quantitative genomic datasets. Mutation frequency data was obtained from cBioPortal (de Bruijn et al., 2023), while protein expression profiles across tissues relevant to therapeutic safety (e.g., brain, gastrointestinal tract, liver, and kidney) were sourced from the ProteinAtlas (Uhlén et al., 2015).

Following the prioritization of the GET list and identification of clinically trialed genes, we annotated these genes with functional genomic data. Mutational frequency—a key determinant in gene ontology ranking (Timar and Kashofer, 2020)—and Copy Number Alterations (CNA), a critical marker of genomic instability (Beroukhim et al., 2010), were evaluated. Mutation and CNA data were sourced from CBioPortal (de Bruijn et al., 2023) using two cohorts: the "Pancreatic Cancer (UTSW, Nat Commun 2015)" and "Pancreatic Adenocarcinoma (TCGA, PanCancer Atlas)" studies, both of which employed whole-exome sequencing for all samples. Network-based metric was also added through BEERE scores, namely the G-list score, GT-list score, E-list score, GET-list score, and the T-list score. The G, E, and T list scores are the BEERE prioritization scores derived from network-based expansion of the lists prioritized in step 2 of the methods. The GET list score is similarly from the merged GET-list detailed in step 2 of the methods. The GT-list score is a combination of the prioritized G and T scores, which aims to bring genes of higher mutational frequency into the network of the T list.

Tissue-specific expression is a vital factor in gene prioritization (Beroukhim et al., 2010). Genes with high expression in essential tissues—such as the heart, liver, gastrointestinal system, brain, and kidneys—pose a higher risk of adverse effects when targeted, necessitating their de-prioritization. Annotation of tissue expression was performed using the "RNA expression score" provided by ProteinAtlas (Uhlén et al., 2015), a comprehensive database mapping protein expression in various organs. This RNA expression score, manually calculated, measures the RNA expression levels of genes across different tissues.

2.5 GETgene-AI ranking

To unify these criteria, we developed a weighted RP score that integrates mutation frequency, copy number alterations (CNA), tissue expression, GET list scores (BEERE prioritization scores derived from network-based expansion), E list scores, GT list scores, and clinical trial activity. Clinical trial popularity was quantified as the number of registered interventional trials testing drugs targeting a given gene for cancer therapy. Modality weights were calibrated by Spearman rank correlation between each modality-specific ranking and two independent benchmarks of therapeutic relevance: (i) the number of associated clinical trials and (ii) the frequency of reported adverse events. The benchmark set used for this analysis consisted of genes implicated in pancreatic cancer clinical trials, independent of the GET and GT lists. Correlations with clinical trial count were used to assess genomic and network features (e.g., mutation frequency, CNA frequency, GET BEERE scores), while correlations with adverse event frequency were used to assess tissue expression features (e.g., expression in brain, liver, lung, and digestive system). Modalities showing stronger monotonic associations contributed proportionally more to the final RP score, while weaker associations retained smaller weights to preserve the potential for novel candidate discovery. Table 1 summarizes the relative weights of each factor in the RP score, ranked in descending order of contribution.

2.6 Mitigation of bias and false positives

To address potential sample biases and data incompleteness—such as studies failing to detect specific genes—multiple datasets from the same databases were utilized wherever possible. This redundancy ensured a more comprehensive analysis and minimized the impact of dataset-specific variability. For example, multiple studies within CBioPortal, such as "Pancreatic Cancer (UTSW, Nat Commun 2015)" and "Pancreatic Adenocarcinoma (TCGA, PanCancer Atlas)," were analyzed concurrently to increase the reliability of mutational frequency and CNA data.

Bias from literature frequency was mitigated by not using citation counts, publication frequency, or other literature-derived popularity metrics as a direct modality in the RP score. Instead, GETgene-AI rankings are based on cancer-type-specific genomic, transcriptomic, and drug-target evidence (mutation frequency, CNA, expression, and network centrality). While genes such as PIK3CA, EGFR, PRKCA, and TNF are indeed well known, their high ranks in our framework derive from pancreatic cancer–specific data rather than their prevalence in the broader cancer literature.

Sensitivity analysis was performed by testing lower and higher cutoffs for both CNA and Mutational Frequency. A lower threshold of 7.3% and a higher threshold of 9.2% was utilized for CNA, while a lower cutoff of 2.2% and a higher cutoff of 3.4% was utilized for mutational frequency. A lower cutoff of 15% and a higher cutoff of 25% was utilized for COSMIC cancer database. Manually searching for genes within the top 250 of GETgene-ai that were included or had functionally related genes within the lower and higher thresholds. A lower threshold did not yield any genes previously not found in the G list, while the higher threshold found P3H2, P4HTM, PLOD3, PLOD2, P4HA1, PLOD1, PAM, PSMB5, C1QC, C1QA, C1QB, to be genes excluded due to higher thresholding. These genes all rank outside of the top 150.

To further enhance the accuracy of the prioritization process, each gene within the top 250 ranked by RP score was manually verified through a literature review to confirm its role in cancer biology. This step was critical in identifying and eliminating false

TABLE 1 Weights each modality was assigned for calculation of the RP score in GETGENE-AI.

Modality of ranking	Weighted score
GT list score	0.329
CNA(CBIOPORTAL UTSW NAT COMMUN 2015)	0.201
Expression list score	0.088
GET list score	0.085
Mutation frequency (cBioporta lTCGA PanCancerAtlas)	0.079
CNA(CBIOPORTAL TCGA PANCANCERATLAS)	0.048
Mutation frequency (Cbioportal UTSW Nat Commun 2015)	−0.023
Brain expression score	−0.054
Kidney expression score	−0.081
Gastrointestinal expression score	−0.095
Liver expression score	−0.101

positives. Notably, no genes within the top 250 were found to be false positives, validating the robustness of the RP scoring methodology.

Additionally, hallucination errors from GPT-4o were mitigated through a structured training approach. The model was instructed to explicitly cite a source used in the calculation of each gene's ranking score. These citations were manually evaluated for accuracy and relevance, ensuring that the ranking process was grounded in verifiable scientific evidence. This dual-layered validation—automated scoring combined with manual review—was integral to maintaining the integrity and reliability of the gene prioritization framework.

2.7 Statistical methods

Spearman correlation coefficients were computed to assess the alignment of GPT-4o rankings with network-derived rankings. The Spearman correlation between the GPT-4 score and the Weighted Score was 0.291, indicating some significance. Interestingly, GPT-4 score is more strongly correlated with all BEERE list ranking scores, with 0.478 between GPT-4 score and Expression list score, 0.457 between GPT-4 score and Combined weighted score of all BEERE lists, 0.454 correlations between GPT-4 score and GET list score, and 0.444 between GPT-4 score and GT list score. These results indicate that the GPT-4 score is more similar to that of standard network prioritization techniques, which may be a result of the training data utilized.

2.8 Comparing research relevance to rank on GETgene-AI

To compare the popularity to the rankings of each gene in both the GPT-4 Score and the RP scores, the amount of results contained on PubMed when searching "Gene name Pancreatic Cancer" were compiled and used for the GPT-LIT score, and the RP-LIT score. The GPT-LIT score is the GPT4-score divided by the amount of publications on PubMed, while the RP-lit score is the RP-score divided by the amount of publications on PubMed. Genes with no functional relationship to cancer in any way were excluded from the rankings to remove false positives.

3 Results

3.1 GETgene-AI rankings and validations

We observe the highest ranked genes according to GETgene-AI in Table 2.

During the iterative ranking process, genes lacking functional relevance to cancer were systematically deprioritized. For instance, genes that ranked highly due to algorithmic artifacts but lacked experimental validation or literature support were ranked lower than genes with experimental validation or literature support. The final candidate set was defined as the top 250 genes ranked by RP score. This threshold was selected to enable manual literature verification for each gene, ensuring that all final candidates could be cross-checked for pancreatic cancer–specific evidence and therapeutic relevance. Expanding the list beyond this size would have substantially increased the manual verification burden without proportionally improving the quality of candidates for downstream analysis. This approach allowed us to maintain both methodological rigor and practical feasibility while focusing on the most highly ranked genes.

PIK3CA emerged as the highest-ranked gene on our list. It encodes the enzyme PI3K, which regulates critical cellular processes such as growth, metabolism, proliferation, and apoptosis (Conway et al., 2019). *PIK3CA* also modulates downstream effectors, including *AKT* and *mTOR* (Ala, 2022), and preclinical studies demonstrate that mutations in this gene sensitize cancers to dual *PI3K*/*mTOR* inhibitors (Zhang et al., 2021), underscoring its therapeutic potential. Notably, *PIK3CA*-null tumors exhibit heightened susceptibility to T-cell surveillance *in vitro* (Sivaram et al., 2019), while its inhibition in pancreatic cancer models initiates tumorigenesis (Payne et al., 2015), highlighting its dual role in progression and therapy.

MYC, the second highest-ranked gene, achieved its position due to its top GET list score, reflecting its network centrality among the 500 most expressed, clinically relevant, and frequently mutated genes. Overexpression of *c-MYC* is a hallmark of aggressive pancreatic cancer, where it binds promoter regions of oncogenic targets (Hayashi et al., 2021). Despite its pivotal regulatory role, *MYC*'s complex protein structure poses therapeutic challenges, resulting in a lower GT list score. Recent advances in small-molecule inhibitors, however, show preclinical promise.

SRC ranks as the third-highest gene on our list, driven by its high scores in both the GET list and Expression list modalities. Inhibition of *SRC* in pancreatic cancer has been shown to reverse chemoresistance to pyroptosis in both *in vitro* and *in vivo* studies (Su et al., 2023). Aberrant *SRC* activity promotes tumorigenesis and is frequently associated with poor prognosis in pancreatic ductal adenocarcinoma (PDAC) (Poh and Ernst, 2023). Several *SRC*-targeting therapies are currently under clinical investigation (Hilbig, 2008).

TABLE 2 Highest 20 genes ranked on GETGENE-AI. Weighted score is RP score, CHAT GPT score is GPT4 score.

Gene	RP score	CHATGPT score	GT list score	Mutation frequency (cBioportal TCGA PanCancer Atlas)	RP-LIT score	GPT-LIT score	GET list score	Expression list score
PIK3CA	34.8	310	58.7	2.8	0.199	1.771	96	97
MYC	30.1	330	9.5	0.0	0.032	0.349	214	210
SRC	20.0	320	0.0	1.1	0.044	0.711	143	144
EGFR	18.2	320	2.4	0.6	0.010	0.171	134	133
CDK1	15.9	305	15.3	65.4	0.134	2.563	30	7
PRKCA	15.3	305	3.0	0.0	1.702	25.556	101	102
TNF	12.1	270	2.4	0.0	0.013	0.292	83	86
LCK	11.5	220	1.7	0.0	1.274	24.444	62	60
JAK2	10.6	285	1.0	0.6	0.082	2.192	67	67
MAPK1	10.3	305	11.6	3.4	0.139	4.122	7	7
AURKB	9.1	295	0.0	0.6	0.008	0.246	70	70
KRAS	8.7	220	1.7	1.7	0.335	8.462	48	47
MAPK8	7.8	295	0.0	0.0	0.002	0.068	121	117
MTOR	7.1	220	1.7	0.0	0.588	18.333	52	52
ITGA4	6.9	220	4.3	0.6	2.298	73.333	40	37
TOP2A	6.9	310	10.2	1.1	0.215	9.688	0	0
CHEK1	6.7	220	1.7	0.0	0.128	4.231	46	45
BCL2	6.2	220	1.7	0.6	0.012	0.418	41	41
PRKCB	6.0	250	1.4	0.6	1.004	41.667	60	58
ERBB4	5.5	220	3.4	0.6	0.184	7.333	81	83

EGFR is the fourth highest-ranked gene, attributed to its high GET list and Expression list scores. *EGFR* is also implicated in tumorigenesis, particularly in lung and breast cancer (Sigismund et al., 2018). Anti-*EGFR* agents have shown significant clinical promise, despite associated adverse effects (Verma et al., 2020).

KRAS ranks 12th on our list, despite its prominence in pancreatic cancer research, with over 4,545 PubMed articles on *KRAS* mutations in pancreatic cancer. Its lower ranking is primarily due to a low expression score. The *KRAS* oncogene plays a critical role in the initiation and maintenance of pancreatic tumors (Luo, 2021). *KRAS* mutations are present in over 90% of PDAC cases, but therapeutic inhibition remains highly challenging, with effective inhibitors only recently being discovered (Bannoura et al., 2021).

CDK1 ranks fifth on our list, largely due to its high scores in both the GET and Expression lists. *CDK1* is strongly correlated with prognosis and is highly expressed in pancreatic cancer tissue, as well as in response to gemcitabine, an approved pancreatic cancer drug (Xu et al., 2023). Additionally, inhibition of *CDK1*, along with *CDK2* and *CDK5*, has been shown to overcome IFN-γ-triggered acquired resistance in pancreatic tumor immunity (Huang et al., 2021).

PRKCA ranks seventh on our list. It encodes protein kinase C and is mutated in various cancers. *PRKCA*'s high ranking is attributed to its strong GET and Expression list scores, as well as its extremely low organ expression score. It is strongly associated with the activation of the protein translation initiation pathway (Rosenberg et al., 2018) and is a hallmark mutation in chordoid gliomas (Jiang et al., 2019). *PRKCA* also contributes to susceptibility to pancreatic cancer through the peroxisome proliferator-activated receptor (*PPAR*) signaling pathway, which plays a key role in pancreatic cancer development and progression (Liu et al., 2020). Inhibition of *PRKCA* has demonstrated antitumor activity in patients with advanced non-small cell lung cancer (NSCLC) (Villalona-Calero et al., 2004).

TNF is the eighth highest-ranked gene on our list. Tumor Necrosis Factor (*TNF*) upregulation is associated with invasion and immunomodulation in pancreatic cancer (Wiedmann et al., 2023). *TNF*-mutated macrophages have also been shown to promote aggressive cancer behaviors through lineage reprogramming (Tu et al., 2021).

LCK ranks ninth on our list. This gene is expressed in tumor cells and plays a key role in T-cell development (Bommhardt et al., 2019). High *LCK* protein expression has been associated with improved patient survival in cancer (Cancer Genome Atlas Network, 2015). Despite its biological relevance, *LCK* has only four PubMed publications discussing its role in pancreatic cancer as of May 2024. Its identification as a high-priority target demonstrates GETgene-AI's ability to prioritize genes with strong biological relevance but limited literature prominence.

ITGA4 ranks 15th on our list. It has an extremely low organ expression score and only four PubMed articles discussing its role in pancreatic cancer. *ITGA4* has potential as an independent prognostic indicator for patient survival and has been linked to the *PI3K/AKT* pathway (Faleiro et al., 2021). Its identification as a high-priority target further highlights GETgene-AI's capability to prioritize genes with strong biological relevance despite limited literature attention.

KCNA ranks 34th on our list. Notably, there are no PubMed publications describing its relation to pancreatic cancer, and only three publications mention its role in cancer in general. The identification of *KCNA* as a high-priority target underscores GETgene-AI's ability to prioritize genes with strong biological relevance but minimal literature prominence. *KCNA* exhibits differentially high expression in stomach and lung cancers and is positively correlated with infiltrated immune cells and survival rates (Angi et al., 2023).

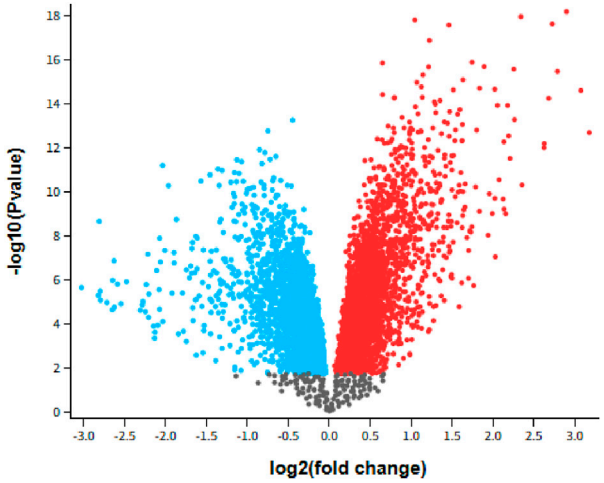


FIGURE 2
Volcano plot GSE28735: Microarray gene-expression profiles of 45 matching pairs of tumor vs. nontumor, Padj<0.05. Blue indicates downregulated while red indicates upregulated.

3.2 Comparing GETgene-AI to other frameworks

We benchmarked GETgene-AI against two other frameworks: one focused on differential expression analysis and the other on network-based gene prioritization. For the differential expression comparison, we selected GEO2R, utilizing the GSE28735 dataset, which was integrated into the 'Expression list' component of our GET lists. Genes were ranked based on their log-fold change (log-fc), representing the difference in gene expression between tumor and non-tumor groups. In the GEO2R list, the top-ranked genes were *PNLIPRP1* and *PNLIPRP2*, both of which encode pancreatic lipase-related proteins critical for digestion and fat absorption (Zhu et al., 2021). However, these genes are not considered viable targets for pancreatic cancer. The third-ranked gene, *IAPP* (Islet Amyloid Polypeptide), has been shown to lack tumor suppressor functionality, and loss of *IAPP* signaling is not associated with pancreatic cancer (Taylor et al., 2023). Among the top 50 genes identified by GEO2R, 30 were experimentally validated as relevant to pancreatic cancer. In contrast, GETgene-AI prioritized 49 experimentally validated targets within its top 50, representing a 38% improvement over GEO2R. GEO2R's limitations, including the absence of mutational frequency analysis, functional impact assessment, network-based analysis, and adverse effect evaluation, hinder its utility in drug target discovery. In comparison, GETgene-AI leverages statistical filtering and incorporates genomic information, significantly enhancing both the efficiency and quality of gene prioritization. Figure 2 presents a volcano plot illustrating the log2 (fold change) distributions for the analyzed genes.

For the network-based comparison, we employed STRING, a database that integrates protein-protein interaction data (Szklarczyk et al., 2023), focusing specifically on the KEGG pathway hsa0512 (Kanehisa and Goto, 2000; Kanehisa, 2019; Kanehisa et al., 2025). Genes were ranked based on node degree, a measure of the number of interactions a protein has within the network (Bozhilova et al., 2019). The highest-ranked gene in the STRING list was *AKT1*, a protein kinase known to stimulate cell growth and proliferation (Grassilli et al., 2020). However, *AKT1* has been shown to resist inhibition by shifting its metabolic activity from glycolysis to mitochondrial respiration (Arasanz et al., 2019). Additionally, it exhibits a low mutational frequency of only 1% in a cohort of 19,784 patients with various tumors (Millis et al., 2016). Due to its low mutational frequency and the challenges associated with its inhibition, *AKT1* was ranked 33rd by GETgene-AI. Among the top 50 genes prioritized by STRING, 46 were experimentally validated for relevance to pancreatic cancer, whereas GETgene-AI identified 49 experimentally validated genes within its top 50, demonstrating a 6% improvement over STRING. STRING's limitations, such as its inability to account for mutational frequency and other critical factors in drug target identification, result in a narrower focus, with only 81 targets prioritized compared to the more comprehensive analysis provided by GETgene-AI. Figure 3 illustrates the network constructed using STRING.

Comparing GETgene-AI to GEO2R and STRING, our framework demonstrates a 38% improvement over GEO2R and a 6% improvement over STRING in the rate of experimental validation of the top 50 genes on each list. In Figure 4, we observe the differences in the percentage of experimentally validated targets out of the top 50.

GETgene-AI was also compared to OpenTarget, an integrative AI-based prioritization platform (Koscielny et al., 2017). We compared GETgene-AI's rankings to those generated by OpenTargets for pancreatic cancer, focusing on the top 15genes from each tool. While there was overlap in high-confidence drivers (e.g., KRAS, TP53, SMAD4, BRCA2), several key differences emerged that highlight the value of GETgene-AI's multi-modal integration.

OpenTargets ranked genes such as POLE and POLD1 highly despite their low mutation frequency in pancreatic cancer datasets (POLE absent in one TCGA cohort; POLD1 <1% in UTSW CNA and mutation frequency). GETgene-AI deprioritized these genes due to the lack of mutational enrichment and limited pancreatic-specific evidence, avoiding inflation from literature-based or pathway-only associations.

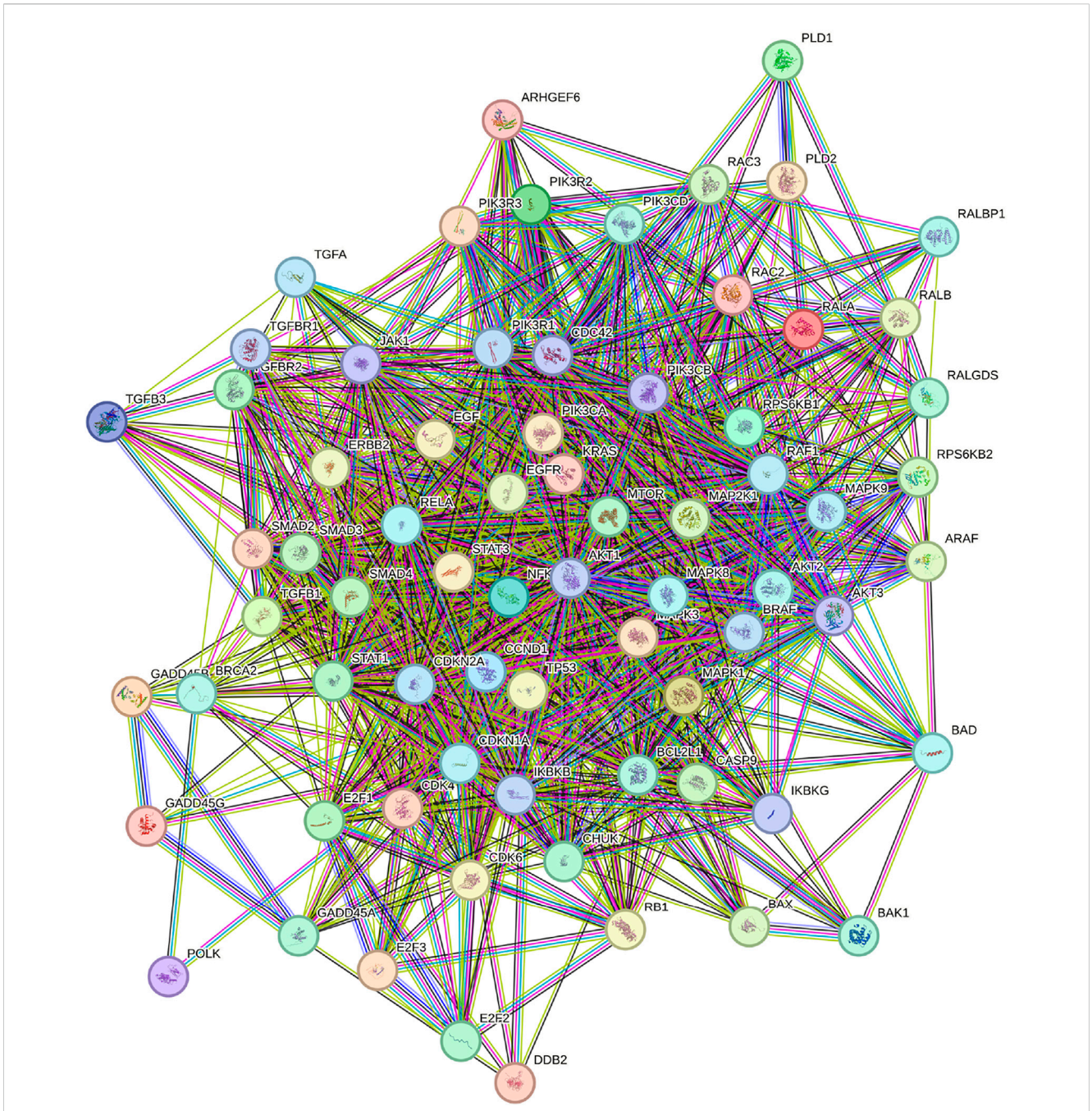


FIGURE 3
Network constructed by STRING utilizing the KEGG pathway HG0512. Content inside each node is known or predicted 3days structure of protein. Turquoise edges mean Protein-protein interactions from curated databases, purple means experimentally determined. Green, red, and dark blue edges indicate predicted Protein-protein interactions. Light green edges represent text mining, black represents co-expression, and light purple represents protein homology.

Conversely, GETgene-AI prioritized genes such as MYC, SRC, EGFR, and CDK1, which have strong differential expression and drug-target relevance in pancreatic cancer but were absent from OpenTargets' top list.

These differences indicate that OpenTargets may overweight generalized associations, whereas GETgene-AI incorporates cancer-type-specific genomic, transcriptomic, and therapeutic data, leading to rankings more aligned with the biological and clinical context of pancreatic cancer.

In Table 3, we observe the ranking overlap for the top 15 genes for all three frameworks. The top 15 highest ranked targets in both GETgene-AI and STRING have all been experimentally validated within pancreatic cancer, but 8 of the highest ranking targets in the GEO2R approach have not.

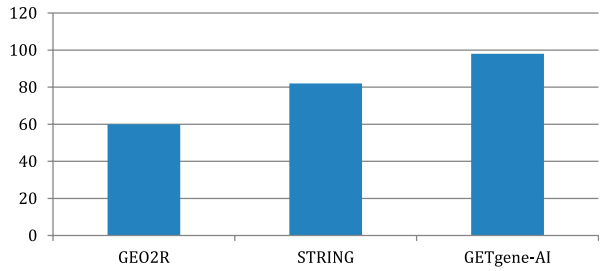


FIGURE 4
Bar graph displaying the percent of experimentally validated targets out of the top 50 genes with each framework.

3.3 Enhancement provided by AI

GPT-4o was utilized to conduct a comprehensive literature assessment for our gene list. Although its output was not incorporated into the final weighted score, the GPT-4o scores demonstrated strong correlations with both the weighted score and all three GET list scores. Notably, GPT-4o prioritized genes such as *MYC* and *SRC*, reflecting their well-documented prominence in the scientific literature. This complemented GETgene-AI's approach, which relies on network mutational analysis for gene prioritization. To minimize the inclusion of false positives in the GPT-4o scoring process, we instructed GPT-4o to directly cite articles from its internal database. While GPT-4o did not exhibit a higher rate of experimental validation compared to manual methods, it significantly reduced the time required for literature review by 80%. All cited articles were subsequently manually verified to ensure accuracy.

The RP-LIT score and GPT-4o score showed a high degree of correlation, with extremely similar rankings for each gene. Based on Spearman correlation analysis, the GPT-4o score (out of 400) exhibited a correlation coefficient of +0.457 with the weighted score, indicating a statistically significant relationship. Table 4 provides a detailed comparison of the ranking differences between the GPT-4o score and the GET ranking score, highlighting the alignment and discrepancies between the two approaches.

3.4 False positives and limitations

False positives are an inherent risk in large-scale computational analyses. The GETgene-AI framework addresses this challenge through iterative refinement and the systematic exclusion of genes lacking functional or experimental support. Future validation efforts will focus on further refining these rankings through targeted experimental studies. Additionally, the literature assessment provided by generative AI is expected to improve as AI technology advances and our model is trained on more experimental data, thereby minimizing inaccuracies or "hallucinations" in the generated outputs.

To mitigate false positives, genes without functional relevance to cancer were systematically excluded. For instance, genes that ranked highly due to algorithmic artifacts but lacked experimental validation or literature support were deprioritized. Examples include *ITGA4* and *PRKCB*, both of which have fewer than 10 PubMed articles discussing their role in pancreatic cancer. These genes were ranked lower than many well-established targets due to their low scores in the GET, GT, and Expression

TABLE 3 Top 15 genes from GETGENE-AI, STRING, and GEO2R and their status as experimentally validated drug targets.

GETGENE-AI top genes	Experimentally validated?	STRING top genes	Experimentally validated?	GEO2R top genes	Experimentally validated?
PIK3CA	Yes	AKT1	Yes	PNLIPRP1	No
MYC	Yes	TP53	Yes	PNLIPRP2	No
SRC	Yes	KRAS	Yes	IAPP	No
EGFR	Yes	PTEN	Yes	CTRC	No
CDK1	Yes	SRC	Yes	GP2	Yes
PRKCA	Yes	STAT3	Yes	CEL	No
TNF	Yes	EGFR	Yes	CPA2	Yes
LCK	Yes	MTOR	Yes	ALB	Yes
JAK2	Yes	BCL2	Yes	CUZD1	Yes
MAPK1	Yes	PIK3CA	Yes	ERP27	No
MTOR	Yes	CDKN2A	Yes	CLPS	Yes
AURKB	Yes	HRAS	Yes	SERPINI2	Yes
KRAS	Yes	CCND1	Yes	PLA2G1B	Yes
MAPK8	Yes	NFKB1	Yes	CELA2A	No
TOP2A	Yes	CDKN1A	Yes	CELA2B	No

TABLE 4 Top 20 highest ranked genes based off of GPT4 score compared to their ranks in GET and their status as experimentally validated drug targets.

Gene	GPT4-score ranking	GET ranking	Experimental validation?	Citation
MYC	1	2	Yes	Zhang et al. (2024)
SRC	2	3	Yes	Su et al. (2023)
EGFR	3	4	Yes	Wu et al. (2023)
TERT	4	27	Yes	Campa et al. (2015)
RRM2	5	21	Yes	Li et al. (2022)
PIK3CA	6	1	Yes	Payne et al. (2015)
TOP2A	7	16	Yes	Pei et al. (2018)
NTRK1	8	22	Yes	Cheng et al. (2013)
PTGS2	9	25	Yes	Hingorani et al. (2003)
EGF	10	30	Yes	Sheng et al. (2020)
CDK1	11	5	Yes	Huang et al. (2021)
MAPK1	12	10	Yes	Si et al. (2023)
KRAS	13	13	Yes	Timar and Kashofer (2020)
MTOR	14	11	Yes	Stanciu et al. (2022)
MSLN	15	37	Yes	Hu et al. (2024)
RET	16	28	Yes	Bhamidipati et al. (2023)
AKT1	17	31	Yes	Arasanz et al. (2019)
JAK2	18	9	Yes	Huang et al. (2022)
MET	19	34	Yes	Pothula et al. (2020)
PDCD1	20	38	Yes	Marabelle et al. (2020)

lists, which prioritize targets with robust experimental or literature support during the RP score calculation process.

This study has several limitations. First, the top-ranked targets identified by GETgene-AI require further experimental validation, which is a critical next step to confirm their biological and therapeutic relevance. Second, the reliance on publicly available datasets may introduce biases due to incomplete or inconsistent annotations. These limitations highlight the need for further experimental validation and the incorporation of more comprehensive datasets to enhance the accuracy and reliability of the framework.

3.5 Broader implications and generalizability

While the current study focuses on pancreatic cancer, the GETgene-AI framework can be readily adapted to other cancers or diseases with access to similar genomic and clinical data resources. Future studies will explore its application to breast and lung cancers by employing the same systematic process described in this work. The GETgene-AI framework integrates literature review, large-scale sequencing data, and network centrality scores, providing a comprehensive approach to drug target prioritization. Additionally, its reliance on computational methods for prioritization and the elimination of statistically insignificant data ensures that the framework is both scalable and efficient, making it suitable for broader applications in biomedical research.

4 Discussion

Through the application of GETgene-AI to pancreatic cancer, we have identified several promising drug targets, including PIK3CA, PRKCA, LCK, MAPK8, ITGA4, PRKCB, and KCNA1, warranting further investigation. These targets display strong pancreatic cancer-specific genomic and transcriptomic evidence, high network centrality in PPI analyses, and have not been extensively reported in the pancreatic cancer literature despite their biological relevance in our analysis.

GETgene-AI's approach to drug target prioritization integrates literature review, large-scale sequencing data, network-based centrality scoring, and assessment of potential adverse effects through organ expression scores. This multifaceted implementation offers a scalable and comprehensive framework for drug target prioritization, which can be readily adapted to other cancers with similar data availability. Furthermore, GETgene-AI's ability to systematically deprioritize genes with low mutational relevance underscores its superiority in efficiently narrowing down actionable and biologically relevant targets. Slight variations of cutoffs utilized for the compilation and prioritization of the GET lists did not result in significant variations of the final rankings or scores of the final GETgene-AI gene list.

In contrast to recent methods that rely largely on AI-driven network analysis alone (e.g., an AI-Driven Network Biology pipeline

TABLE 5 genes highlighted in the discussion section labeled by significance.

Gene	Significance
PIK3CA	Investigated in PDAC clinical trials
PRKCA	Investigated in PDAC preclinical models (*in vitro* or *in vivo*)
LCK	Investigated in PDAC preclinical models (*in vitro* or *in vivo*)
MAPK8	Investigated in PDAC preclinical models (*in vitro* or *in vivo*)
ITGA4	Novel and unstudied in PDAC
PRKCB	Novel and unstudied in PDAC
KCNA1	Novel and unstudied in PDAC

identifying SRC as a therapeutic target in pancreatic cancer) (Zhang and Chen, 2025), GETgene-AI offers a more automated and modular framework. Our approach not only evaluates protein–protein interaction networks but also incorporates tissue-specific gene expression and mutation frequency analyses, and integrates these modalities through distinct G, E, and T lists before merging. This enables multi-dimensional prioritization grounded in genomic, transcriptomic, and therapeutic evidence. In future extensions, the modular nature of GETgene-AI allows easy incorporation of additional evaluation modules—such as differential tissue analysis, motif-based mutation enrichment, or epigenetic regulation scores—each processed independently in their own list and then integrated via our weighted RP score. This design ensures adaptability and enables seamless expansion of the framework to accommodate new modalities as the data landscape evolves.

4.1 Contributions and limitations provided by GPT4o

GPT-4o significantly enhanced the efficiency of literature-based ranking by automating the review and prioritization of scientific abstracts. This approach increased the efficiency of literature review by over 80%. However, inherent challenges, such as the risk of hallucination, necessitated manual verification to ensure the accuracy of the results. While GPT-4o provides substantial value, its integration into research workflows should be approached cautiously, with safeguards implemented to mitigate potential errors. Additionally, training GPT-4o on more experimental data in the future will further improve its accuracy and reliability in prioritization tasks.

4.2 Future directions

While the current study focuses on cancer applications, future research will expand the scope of the GETgene-AI framework. We plan to validate its utility in additional cancer types, such as breast and lung cancer, and explore its applicability to non-cancerous disease contexts, including neurodegenerative disorders like Alzheimer's and Parkinson's. By integrating computational methods with large-scale genomic data, the GETgene-AI framework addresses critical gaps in drug discovery, accelerating the identification of actionable targets and advancing the development of personalized therapeutic strategies.

Future work will prioritize experimental validation of top-ranked targets, such as *PIK3CA* and *PRKCA*, using CRISPR-mediated knockouts in pancreatic cancer cell lines. Subsequent *in vitro* drug response assays will evaluate the therapeutic potential of these targets. Additionally, we aim to refine the framework by incorporating multi-omics datasets (e.g., proteomics, metabolomics) and enhancing its ability to predict adverse effects through improved organ expression profiling. ce of these targets.

5 Conclusion

The GET framework represents a significant advancement in computational drug discovery, integrating network-based prioritization with machine learning to prioritize actionable therapeutic targets efficiently. Genes highlighted through our case study in pancreatic cancer such as *PRKCA*, *LCK*, *ITGA4*, and *PRKCB* are novel targets that require further exploration. While this study focuses on pancreatic cancer, the GETGENE-AI framework is adaptable to other cancers and diseases, offering a modular and versatile approach for target discovery. GPT4o enhanced the efficiency and accuracy of literature-based ranking, reducing manual workload and aligning well with network-based rankings. However, its reliance on manual verification underscores the need for cautious integration into automated pipelines. By refining target discovery methods, the GETGENE-AI framework paves the way for personalized therapeutic strategies and accelerates the translational research in oncology. Future work will focus on expanding the framework to other cancers, improving ranking metrics, and integrating multi-omics datasets to enhance its predictive power. Future iterations of GETgene-AI aim to integrate multi-omics datasets, such as single-cell RNA-seq and metabolomics, to capture greater biological complexity. Table 5 indicates the significance of each gene labeled as novel.

Data availability statement

The data presented in the study are deposited in the https://github.com/alphamind-club/GETGENE-AI repository.

Author contributions

AG: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Resources, Validation, Visualization, Writing – original draft, Writing – review and editing. JC: Conceptualization, Data curation, Funding acquisition, Investigation, Resources, Software, Supervision, Validation, Writing – original draft, Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. JYC thanks the generous support of the startup fund to the Systems Pharmacology AI

Research Center at UAB and NIH common fund grant award U54-OD036472, which partially supported this research.

Acknowledgments

Both authors thank the administrative support of AlphaMind Club for making this mentored research possible. The authors acknowledge the use of ChatGPT in improving the structure and readability 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 author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. Generative AI utilized as part of the drug target analysis approach. Generative AI was also utilized to revise manuscript.

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

Publisher's note

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

Ala, M. (2022). Target c-Myc to treat pancreatic cancer. *Cancer Biol. Ther.* 23, 34–50. doi:10.1080/15384047.2021.2017223

Angi, B., Muccioli, S., Szabò, I., and Leanza, L. (2023). A meta-analysis study to infer voltage-gated K+ Channels prognostic value in different cancer types. *Antioxid. Basel Switz.* 12, 573. doi:10.3390/antiox12030573

Arasanz, H., Zuazo, M., Santamaría, E., Bocanegra, A. I., Gato-Cañas, M., Fernández-Hinojal, G., et al. (2019). Adaption of pancreatic cancer cells to AKT1 inhibition induces the acquisition of cancer stem-cell like phenotype through upregulation of mitochondrial functions. *Ann. Oncol.* 30, v11. doi:10.1093/annonc/mdz238.036

Bai, J. P. F., Alekseyenko, A. V., Statnikov, A., Wang, I.-M., and Wong, P. H. (2013). Strategic applications of gene expression: from drug discovery/development to bedside. *AAPS J.* 15, 427–437. doi:10.1208/s12248-012-9447-1

Bannoura, S. F., Uddin, Md. H., Nagasaka, M., Fazili, F., Al-Hallak, M. N., Philip, P. A., et al. (2021). Targeting KRAS in pancreatic cancer: new drugs on the horizon. *Cancer Metastasis Rev.* 40, 819–835. doi:10.1007/s10555-021-09990-2

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2013). NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res.* 41, D991–D995. doi:10.1093/nar/gks1193

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

Bhamidipati, D., Yedururi, S., Huse, J., Chinapuvvula, S. V., Wu, J., and Subbiah, V. (2023). Exceptional responses to selpercatinib in RET fusion-driven metastatic pancreatic cancer. *JCO Precis. Oncol.* 7, e2300252. doi:10.1200/PO.23.00252

Bommhardt, U., Schraven, B., and Simeoni, L. (2019). Beyond TCR signaling: emerging functions of lck in cancer and immunotherapy. *Int. J. Mol. Sci.* 20, 3500. doi:10.3390/ijms20143500

Bozhilova, L. V., Whitmore, A. V., Wray, J., Reinert, G., and Deane, C. M. (2019). Measuring rank robustness in scored protein interaction networks. *BMC Bioinforma.* 20, 446. doi:10.1186/s12859-019-3036-6

Campa, D., Rizzato, C., Stolzenberg-Solomon, R., Pacetti, P., Vodicka, P., Cleary, S. P., et al. (2015). TERT gene harbors multiple variants associated with pancreatic cancer susceptibility. *Int. J. Cancer* 137, 2175–2183. doi:10.1002/ijc.29590

Cancer Genome Atlas Network (2015). Genomic classification of cutaneous melanoma. *Cell* 161, 1681–1696. doi:10.1016/j.cell.2015.05.044

Chang, L., Ruiz, P., Ito, T., and Sellers, W. R. (2021). Targeting pan-essential genes in cancer: challenges and opportunities. *Cancer Cell* 39, 466–479. doi:10.1016/j.ccell.2020.12.008

Chen, J. Y., Shen, C., and Sivachenko, A. Y. (2006). Mining Alzheimer disease relevant proteins from integrated protein interactome data. *Pac. Symp. Biocomput. Pac. Symp. Biocomput.*, 367–378.

Chen, J. Y., Mamidipalli, S., and Huan, T. (2009). HAPPI: an online database of comprehensive human annotated and predicted protein interactions. *BMC Genomics* 10, S16. doi:10.1186/1471-2164-10-S1-S16

Chen, J. Y., Piquette-Miller, M., and Smith, B. P. (2013). Network medicine: finding the links to personalized therapy. *Clin. Pharmacol. Ther.* 94, 613–616. doi:10.1038/clpt.2013.195

Chen, J. Y., Pandey, R., and Nguyen, T. M. (2017). HAPPI-2: a comprehensive and high-quality map of human annotated and predicted protein interactions. *BMC Genomics* 18, 182. doi:10.1186/s12864-017-3512-1

Cheng, Y., Diao, D., Zhang, H., Song, Y.-C., and Dang, C.-X. (2013). Proliferation enhanced by NGF-NTRK1 signaling makes pancreatic cancer cells more sensitive to 2DG-induced apoptosis. *Int. J. Med. Sci.* 10, 634–640. doi:10.7150/ijms.5547

Chowbina, S. R., Wu, X., Zhang, F., Li, P. M., Pandey, R., Kasamsetty, H. N., et al. (2009). HPD: an online integrated human pathway database enabling systems biology studies. *BMC Bioinforma.* 10, S5. doi:10.1186/1471-2105-10-S11-S5

Conway, J. R., Herrmann, D., Evans, T. J., Morton, J. P., and Timpson, P. (2019). Combating pancreatic cancer with PI3K pathway inhibitors in the era of personalised medicine. *Gut* 68, 742–758. doi:10.1136/gutjnl-2018-316822

de Bruijn, I., Kundra, R., Mastrogiacomo, B., Tran, T. N., Sikina, L., Mazor, T., et al. (2023). Analysis and visualization of longitudinal genomic and clinical data from the AACR project GENIE biopharma collaborative in cBioPortal. *Cancer Res.* 83, 3861–3867. doi:10.1158/0008-5472.CAN-23-0816

Dinstag, G., and Shamir, R. (2020). PRODIGY: personalized prioritization of driver genes. *Bioinformatics* 36, 1831–1839. doi:10.1093/bioinformatics/btz815

Faleiro, I., Roberto, V. P., Demirkol Canli, S., Fraunhoffer, N. A., Iovanna, J., Gure, A. O., et al. (2021). DNA methylation of PI3K/AKT pathway-related genes predicts outcome in patients with pancreatic cancer: a comprehensive bioinformatics-based study. *Cancers* 13, 6354. doi:10.3390/cancers13246354

Gong, E., and Chen, J. Y. (2023). Prioritizing complex disease genes from heterogeneous public databases. doi:10.1101/2023.02.09.527562

Grassilli, S., Brugnoli, F., Lattanzio, R., Buglioni, S., and Bertagnolo, V. (2020). Vav1 down-modulates Akt2 expression in cells from pancreatic ductal adenocarcinoma: nuclear Vav1 as a potential regulator of akt related malignancy in pancreatic cancer. *Biomedicines* 8, 379. doi:10.3390/biomedicines8100379

Gu, L., Hickey, R. J., and Malkas, L. H. (2023). Therapeutic targeting of DNA replication stress in cancer. *Genes* 14, 1346. doi:10.3390/genes14071346

Hayashi, A., Hong, J., and Iacobuzio-Donahue, C. A. (2021). The pancreatic cancer genome revisited. *Nat. Rev. Gastroenterol. Hepatol.* 18, 469–481. doi:10.1038/s41575-021-00463-z

Hilbig, A. (2008). "Src kinase and pancreatic cancer," in *Pancreatic cancer* (Berlin, Heidelberg: Springer Berlin Heidelberg), 179–185. doi:10.1007/978-3-540-71279-4_19

Hingorani, S. R., Petricoin, E. F., Maitra, A., Rajapakse, V., King, C., Jacobetz, M. A., et al. (2003). Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. *Cancer Cell* 4, 437–450. doi:10.1016/s1535-6108(03)00309-x

Hu, J.-X., Zhao, C.-F., Chen, W.-B., Liu, Q.-C., Li, Q.-W., Lin, Y.-Y., et al. (2021). Pancreatic cancer: a review of epidemiology, trend, and risk factors. *World J. Gastroenterol.* 27, 4298–4321. doi:10.3748/wjg.v27.i27.4298

Hu, J., Wang, J., Guo, X., Fan, Q., Li, X., Li, K., et al. (2024). MSLN induced EMT, cancer stem cell traits and chemotherapy resistance of pancreatic cancer cells. *Heliyon* 10, e29210. doi:10.1016/j.heliyon.2024.e29210

Huan, T., Wu, X., and Chen, J. Y. (2010). Systems biology visualization tools for drug target discovery. *Expert Opin. Drug Discov.* 5, 425–439. doi:10.1517/17460441003725102

Huang, H., Wu, X., Pandey, R., Li, J., Zhao, G., Ibrahim, S., et al. (2012a). C²Maps: a network pharmacology database with comprehensive disease-gene-drug connectivity relationships. *BMC Genomics* 13, S17. doi:10.1186/1471-2164-13-S6-S17

Huang, H., Wu, X., Sonachalam, M., Mandape, S. N., Pandey, R., MacDorman, K. F., et al. (2012b). PAGED: a pathway and gene-set enrichment database to enable molecular phenotype discoveries. *BMC Bioinforma.* 13, S2. doi:10.1186/1471-2105-13-S15-S2

Huang, H., Nguyen, T., Ibrahim, S., Shantharam, S., Yue, Z., and Chen, J. Y. (2015). DMAP: a connectivity map database to enable identification of novel drug repositioning candidates. *BMC Bioinforma.* 16, S4. doi:10.1186/1471-2105-16-S13-S4

Huang, J., Chen, P., Liu, K., Liu, J., Zhou, B., Wu, R., et al. (2021). CDK1/2/5 inhibition overcomes IFNG-mediated adaptive immune resistance in pancreatic cancer. *Gut* 70, 890–899. doi:10.1136/gutjnl-2019-320441

Huang, B., Lang, X., and Li, X. (2022). The role of IL-6/JAK2/STAT3 signaling pathway in cancers. *Front. Oncol.* 12, 1023177. doi:10.3389/fonc.2022.1023177

Janyasupab, P., Suratanee, A., and Plaimas, K. (2021). Network diffusion with centrality measures to identify disease-related genes. *Math. Biosci. Eng.* 18, 2909–2929. doi:10.3934/mbe.2021147

Jiang, H., Fu, Q., Song, X., Ge, C., Li, R., Li, Z., et al. (2019). HDGF and PRKCA upregulation is associated with a poor prognosis in patients with lung adenocarcinoma. *Oncol. Lett.* 18, 4936–4946. doi:10.3892/ol.2019.10812

Kanehisa, M. (2019). Toward understanding the origin and evolution of cellular organisms. *Protein Sci. Publ. Protein Soc.* 28, 1947–1951. doi:10.1002/pro.3715

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi:10.1093/nar/28.1.27

Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y., and Ishiguro-Watanabe, M. (2025). KEGG: biological systems database as a model of the real world. *Nucleic Acids Res.* 53, D672–D677. doi:10.1093/nar/gkae909

Khraisha, Q., Put, S., Kappenberg, J., Warraitch, A., and Hadfield, K. (2024). Can large language models replace humans in systematic reviews? Evaluating GPT-4's efficacy in screening and extracting data from peer-reviewed and grey literature in multiple languages. *Res. Synth. Methods* 15, 616–626. doi:10.1002/jrsm.1715

Koscielny, G., An, P., Carvalho-Silva, D., Cham, J. A., Fumis, L., Gasparyan, R., et al. (2017). Open Targets: a platform for therapeutic target identification and validation. *Nucleic Acids Res.* 45, D985–D994. doi:10.1093/nar/gkw1055

Lawrence, M. S., Stojanov, P., Mermel, C. H., Garraway, L. A., Golub, T. R., Meyerson, M., et al. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 505, 495–501. doi:10.1038/nature12912

Lazzeroni, L. C., Lu, Y., and Belitskaya-Lévy, I. (2014). P-values in genomics: apparent precision masks high uncertainty. *Mol. Psychiatry* 19, 1336–1340. doi:10.1038/mp.2013.184

Li, J., Zhu, X., and Chen, J. Y. (2010). Discovering breast cancer drug candidates from biomedical literature. *Int. J. Data Min. Bioinforma.* 4, 241–255. doi:10.1504/IJDMB.2010.033519

Li, W., Chen, Q., Gao, W., and Zeng, H. (2022). ARID1A promotes chemosensitivity to gemcitabine in pancreatic cancer through epigenetic silencing of RRM2. *Pharm* 77, 224–229. doi:10.1691/ph.2022.1881

Lim, B., Greer, Y., Lipkowitz, S., and Takebe, N. (2019). Novel apoptosis-inducing agents for the treatment of cancer, a new arsenal in the toolbox. *Cancers* 11, 1087. doi:10.3390/cancers11081087

Lindsay, M. A. (2003). Target discovery. *Nat. Rev. Drug Discov.* 2, 831–838. doi:10.1038/nrd1202

Liu, X., Qian, D., Liu, H., Abbruzzese, J. L., Luo, S., Walsh, K. M., et al. (2020). Genetic variants of the peroxisome proliferator-activated receptor (PPAR) signaling pathway genes and risk of pancreatic cancer. *Mol. Carcinog.* 59, 930–939. doi:10.1002/mc.23208

Liu, Z., Roberts, R. A., Lal-Nag, M., Chen, X., Huang, R., and Tong, W. (2021). AI-based language models powering drug discovery and development. *Drug Discov. Today* 26, 2593–2607. doi:10.1016/j.drudis.2021.06.009

López-Cortés, A., Paz-y-Miño, C., Cabrera-Andrade, A., Barigye, S. J., Munteanu, C. R., González-Díaz, H., et al. (2018). Gene prioritization, communality analysis, networking and metabolic integrated pathway to better understand breast cancer pathogenesis. *Sci. Rep.* 8, 16679. doi:10.1038/s41598-018-35149-1

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

Luo, J. (2021). KRAS mutation in pancreatic cancer. *Semin. Oncol.* 48, 10–18. doi:10.1053/j.seminoncol.2021.02.003

Magger, O., Waldman, Y. Y., Ruppin, E., and Sharan, R. (2012). Enhancing the prioritization of disease-causing genes through tissue specific protein interaction networks. *PLoS Comput. Biol.* 8, e1002690. doi:10.1371/journal.pcbi.1002690

Marabelle, A., Le, D. T., Ascierto, P. A., Di Giacomo, A. M., De Jesus-Acosta, A., Delord, J.-P., et al. (2020). Efficacy of pembrolizumab in patients with noncolorectal high microsatellite instability/mismatch repair-deficient cancer: results from the phase II KEYNOTE-158 study. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* 38, 1–10. doi:10.1200/JCO.19.02105

McCarthy, D. J., and Smyth, G. K. (2009). Testing significance relative to a fold-change threshold is a TREAT. *Oxf. Engl.* 25, 765–771. doi:10.1093/bioinformatics/btp053

Millis, S. Z., Ikeda, S., Reddy, S., Gatalica, Z., and Kurzrock, R. (2016). Landscape of phosphatidylinositol-3-kinase pathway alterations across 19 784 diverse solid tumors. *JAMA Oncol.* 2, 1565–1573. doi:10.1001/jamaoncol.2016.0891

Mohsen, H., Gunasekharan, V., Qing, T., Seay, M., Surovtseva, Y., Negahban, S., et al. (2021). Network propagation-based prioritization of long tail genes in 17 cancer types. *Genome Biol.* 22, 287. doi:10.1186/s13059-021-02504-x

Mutch, D. M., Berger, A., Mansourian, R., Rytz, A., and Roberts, M.-A. (2002). The limit fold change model: a practical approach for selecting differentially expressed genes from microarray data. *BMC Bioinforma.* 3, 17. doi:10.1186/1471-2105-3-17

Oniani, D., Hilsman, J., Zang, C., Wang, J., Cai, L., Zawala, J., et al. (2024). Emerging opportunities of using large language models for translation between drug molecules and indications. *Sci. Rep.* 14, 10738. doi:10.1038/s41598-024-61124-0

Paananen, J., and Fortino, V. (2020). An omics perspective on drug target discovery platforms. *Brief. Bioinform.* 21, 1937–1953. doi:10.1093/bib/bbz122

Payne, S. N., Maher, M. E., Tran, N. H., Van De Hey, D. R., Foley, T. M., Yueh, A. E., et al. (2015). PIK3CA mutations can initiate pancreatic tumorigenesis and are targetable with PI3K inhibitors. *Oncogenesis* 4, e169. doi:10.1038/oncsis.2015.28

Pei, Y.-F., Yin, X.-M., and Liu, X.-Q. (2018). TOP2A induces malignant character of pancreatic cancer through activating β-catenin signaling pathway. *Biochim. Biophys. Acta Mol. Basis Dis.* 1864, 197–207. doi:10.1016/j.bbadis.2017.10.019

Petti, M., Bizzarri, D., Verrienti, A., Falcone, R., and Farina, L. (2020). Connectivity significance for disease gene prioritization in an expanding universe. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 17, 2155–2161. doi:10.1109/TCBB.2019.2938512

Poh, A. R., and Ernst, M. (2023). Functional roles of SRC signaling in pancreatic cancer: recent insights provide novel therapeutic opportunities. *Oncogene* 42, 1786–1801. doi:10.1038/s41388-023-02701-x

Pothula, S. P., Xu, Z., Goldstein, D., Pirola, R. C., Wilson, J. S., and Apte, M. V. (2020). Targeting HGF/c-MET Axis in pancreatic cancer. *Int. J. Mol. Sci.* 21, 9170. doi:10.3390/ijms21239170

Rosenberg, S., Simeonova, I., Bielle, F., Verreault, M., Bance, B., Le Roux, I., et al. (2018). A recurrent point mutation in PRKCA is a hallmark of chordoid gliomas. *Nat. Commun.* 9, 2371. doi:10.1038/s41467-018-04622-w

Sadybekov, A. V., and Katritch, V. (2023). Computational approaches streamlining drug discovery. *Nature* 616, 673–685. doi:10.1038/s41586-023-05905-z

Sallam, M. (2023). ChatGPT utility in healthcare Education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthc. Basel Switz.* 11, 887. doi:10.3390/healthcare11060887

Sellers, W. R., and Fisher, D. E. (1999). Apoptosis and cancer drug targeting. *J. Clin. Invest.* 104, 1655–1661. doi:10.1172/JCI9053

Sheng, W., Shi, X., Lin, Y., Tang, J., Jia, C., Cao, R., et al. (2020). Musashi2 promotes EGF-induced EMT in pancreatic cancer via ZEB1-ERK/MAPK signaling. *J. Exp. Clin. Cancer Res. CR* 39, 16. doi:10.1186/s13046-020-1521-4

Shim, J. E., Hwang, S., and Lee, I. (2015). Pathway-dependent effectiveness of network algorithms for gene prioritization. *PLOS ONE* 10, e0130589. doi:10.1371/journal.pone.0130589

Si, H., Zhang, N., Shi, C., Luo, Z., and Hou, S. (2023). Tumor-suppressive miR-29c binds to MAPK1 inhibiting the ERK/MAPK pathway in pancreatic cancer. *Clin. Transl. Oncol.* 25, 803–816. doi:10.1007/s12094-022-02991-9

Sigismund, S., Avanzato, D., and Lanzetti, L. (2018). Emerging functions of the EGFR in cancer. *Mol. Oncol.* 12, 3–20. doi:10.1002/1878-0261.12155

Singh, N., Vayer, P., Tanwar, S., Poyet, J.-L., Tsaioun, K., and Villoutreix, B. O. (2023). Drug discovery and development: introduction to the general public and patient groups. *Front. Drug Discov.* 3, 1201419. doi:10.3389/fddsv.2023.1201419

Sinkala, M. (2023). Mutational landscape of cancer-driver genes across human cancers. *Sci. Rep.* 13, 12742. doi:10.1038/s41598-023-39608-2

Sivaram, N., McLaughlin, P. A., Han, H. V., Petrenko, O., Jiang, Y.-P., Ballou, L. M., et al. (2019). Tumor-intrinsic PIK3CA represses tumor immunogenecity in a model of pancreatic cancer. *J. Clin. Invest.* 129, 3264–3276. doi:10.1172/JCI123540

Sliwoski, G., Kothiwale, S., Meiler, J., and Lowe, E. W. (2014). Computational methods in drug discovery. *Pharmacol. Rev.* 66, 334–395. doi:10.1124/pr.112.007336

Somarelli, J. A., Boddy, A. M., Gardner, H. L., DeWitt, S. B., Tuohy, J., Megquier, K., et al. (2019). Improving cancer drug discovery by studying cancer across the tree of life. *Mol. Biol. Evol.* 37, 11–17. doi:10.1093/molbev/msz254

Sonehara, K., and Okada, Y. (2021). Genomics-driven drug discovery based on disease-susceptibility genes. *Inflamm. Regen.* 41, 8. doi:10.1186/s41232-021-00158-7

Stanciu, S., Ionita-Radu, F., Stefani, C., Miricescu, D., Stanescu-Spinu, I.-I., Greabu, M., et al. (2022). Targeting PI3K/AKT/mTOR signaling pathway in pancreatic cancer: from molecular to clinical aspects. *Int. J. Mol. Sci.* 23, 10132. doi:10.3390/ijms231710132

Steele, C. D., Abbasi, A., Islam, S. M. A., Bowes, A. L., Khandekar, A., Haase, K., et al. (2022). Signatures of copy number alterations in human cancer. *Nature* 606, 984–991. doi:10.1038/s41586-022-04738-6

Su, L., Chen, Y., Huang, C., Wu, S., Wang, X., Zhao, X., et al. (2023). Targeting Src reactivates pyroptosis to reverse chemoresistance in lung and pancreatic cancer models. *Sci. Transl. Med.* 15, eabl7895. doi:10.1126/scitranslmed.abl7895

Sun, Y., Liu, Y., Ma, X., and Hu, H. (2021). The influence of cell cycle regulation on chemotherapy. *Int. J. Mol. Sci.* 22, 6923. doi:10.3390/ijms22136923

Sun, D., Gao, W., Hu, H., and Zhou, S. (2022). Why 90% of clinical drug development fails and how to improve it? *Acta Pharm. Sin. B* 12, 3049–3062. doi:10.1016/j.apsb.2022.02.002

Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., et al. (2023). The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* 51, D638–D646. doi:10.1093/nar/gkac1000

Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., et al. (2019). COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* 47, D941–D947. doi:10.1093/nar/gky1015

Taylor, A. J., Panzhinskiy, E., Orban, P. C., Lynn, F. C., Schaeffer, D. F., Johnson, J. D., et al. (2023). Islet amyloid polypeptide does not suppress pancreatic cancer. *Mol. Metab.* 68, 101667. doi:10.1016/j.molmet.2023.101667

Timar, J., and Kashofer, K. (2020). Molecular epidemiology and diagnostics of KRAS mutations in human cancer. *Cancer Metastasis Rev.* 39, 1029–1038. doi:10.1007/s10555-020-09915-5

Trajanoska, K., Bhérer, C., Taliun, D., Zhou, S., Richards, J. B., and Mooser, V. (2023). From target discovery to clinical drug development with human genetics. *Nature* 620, 737–745. doi:10.1038/s41586-023-06388-8

Tripathi, S., Gabriel, K., Tripathi, P. K., and Kim, E. (2024). Large language models reshaping molecular biology and drug development. *Chem. Biol. Drug Des.* 103, e14568. doi:10.1111/cbdd.14568

Tu, M., Klein, L., Espinet, E., Georgomanolis, T., Wegwitz, F., Li, X., et al. (2021). TNF-α-producing macrophages determine subtype identity and prognosis via AP1 enhancer reprogramming in pancreatic cancer. *Nat. Cancer* 2, 1185–1203. doi:10.1038/s43018-021-00258-w

Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. *Science* 347, 1260419. doi:10.1126/science.1260419

Van de Sande, B., Lee, J. S., Mutasa-Gottgens, E., Naughton, B., Bacon, W., Manning, J., et al. (2023). Applications of single-cell RNA sequencing in drug discovery and development. *Nat. Rev. Drug Discov.* 22, 496–520. doi:10.1038/s41573-023-00688-4

Verma, H. K., Kampalli, P. K., Lakkakula, S., Chalikonda, G., Bhaskar, L. V. K. S., and Pattnaik, S. (2020). A retrospective look at anti-EGFR agents in pancreatic cancer therapy. *Curr. Drug Metab.* 20, 958–966. doi:10.2174/1389200220666191122104955

Villalona-Calero, M. A., Ritch, P., Figueroa, J. A., Otterson, G. A., Belt, R., Dow, E., et al. (2004). A phase I/II study of LY900003, an antisense inhibitor of protein kinase C-alpha, in combination with cisplatin and gemcitabine in patients with advanced non-small cell lung cancer. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* 10, 6086–6093. doi:10.1158/1078-0432.CCR-04-0779

Wang, W., Li, C., Dai, Y., Wu, Q., and Yu, W. (2024). Unraveling metabolic characteristics and clinical implications in gastric cancer through single-cell resolution analysis. *Front. Mol. Biosci.* 11, 1399679. doi:10.3389/fmolb.2024.1399679

Wiedmann, L., De Angelis Rigotti, F., Vaquero-Siguero, N., Donato, E., Espinet, E., Moll, I., et al. (2023). HAPLN1 potentiates peritoneal metastasis in pancreatic cancer. *Nat. Commun.* 14, 2353. doi:10.1038/s41467-023-38064-w

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., et al. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 46, D1074–D1082. doi:10.1093/nar/gkx1037

Wu, X., Huang, H., Wei, T., Pandey, R., Reinhard, C., Li, S. D., et al. (2012). Network expansion and pathway enrichment analysis towards biologically significant findings from microarrays. *J. Integr. Bioinforma.* 9, 213. doi:10.2390/biecoll-jib-2012-213

Wu, F., He, J., Deng, Q., Chen, J., Peng, M., Xiao, J., et al. (2023). Neuroglobin inhibits pancreatic cancer proliferation and metastasis by targeting the GNAI1/EGFR/AKT/ERK signaling axis. *Biochem. Biophys. Res. Commun.* 664, 108–116. doi:10.1016/j.bbrc.2023.04.080

Xu, X., Ding, Y., Jin, J., Xu, C., Hu, W., Wu, S., et al. (2023). Post-translational modification of CDK1–STAT3 signaling by fisetin suppresses pancreatic cancer stem cell properties. *Cell Biosci.* 13, 176. doi:10.1186/s13578-023-01118-z

Yue, Z., Zheng, Q., Neylon, M. T., Yoo, M., Shin, J., Zhao, Z., et al. (2018). PAGER 2.0: an update to the pathway, annotated-list and gene-signature electronic repository for Human Network Biology. *Nucleic Acids Res.* 46, D668–D676. doi:10.1093/nar/gkx1040

Yue, Z., Willey, C. D., Hjelmeland, A. B., and Chen, J. Y. (2019). BEERE: a web server for biomedical entity expansion, ranking and explorations. *Nucleic Acids Res.* 47, W578–W586. doi:10.1093/nar/gkz428

Yue, Z., Slominski, R., Bharti, S., and Chen, J. Y. (2022). PAGER web APP: an interactive, online gene set and network interpretation tool for functional genomics. *Front. Genet.* 13, 820361. doi:10.3389/fgene.2022.820361

Zhang, A., and Chen, J. Y. (2025). AI-driven network biology identifies SRC as a therapeutic target in metastatic pancreatic adenocarcinoma. *Intell. Oncol.* 1 (3), 233–243. doi:10.1016/j.intonc.2025.06.004

Zhang, G., Schetter, A., He, P., Funamizu, N., Gaedcke, J., Ghadimi, B. M., et al. (2012). DPEP1 inhibits tumor cell invasiveness, enhances chemosensitivity and predicts clinical outcome in pancreatic ductal adenocarcinoma. *PloS One* 7, e31507. doi:10.1371/journal.pone.0031507

Zhang, G., He, P., Tan, H., Budhu, A., Gaedcke, J., Ghadimi, B. M., et al. (2013). Integration of metabolomics and transcriptomics revealed a fatty acid network exerting growth inhibitory effects in human pancreatic cancer. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* 19, 4983–4993. doi:10.1158/1078-0432.CCR-13-0209

Zhang, H., Ferguson, A., Robertson, G., Jiang, M., Zhang, T., Sudlow, C., et al. (2021). Benchmarking network-based gene prioritization methods for cerebral small vessel disease. *Brief. Bioinform.* 22, bbab006. doi:10.1093/bib/bbab006

Zhang, H., Sun, Y., Wang, Z., Huang, X., Tang, L., Jiang, K., et al. (2024). ZDHHC20-mediated S-palmitoylation of YTHDF3 stabilizes MYC mRNA to promote pancreatic cancer progression. *Nat. Commun.* 15, 4642. doi:10.1038/s41467-024-49105-3

Zhou, S.-F., and Zhong, W.-Z. (2017). Drug design and discovery: principles and applications. *Molecules* 22, 279. doi:10.3390/molecules22020279

Zhou, Y., Zhang, Y., Lian, X., Li, F., Wang, C., Zhu, F., et al. (2022). Therapeutic target database update 2022: facilitating drug discovery with enriched comparative data of targeted agents. *Nucleic Acids Res.* 50, D1398–D1407. doi:10.1093/nar/gkab953

Zhu, G., Fang, Q., Zhu, F., Huang, D., and Yang, C. (2021a). Structure and function of pancreatic lipase-related protein 2 and its relationship with pathological states. *Front. Genet.* 12, 693538. doi:10.3389/fgene.2021.693538

Zhu, L., Jiang, M., Wang, H., Sun, H., Zhu, J., Zhao, W., et al. (2021b). A narrative review of tumor heterogeneity and challenges to tumor drug therapy. *Ann. Transl. Med.* 9, 1351. doi:10.21037/atm-21-1948

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 10 October 2025
DOI 10.3389/fmed.2025.1640961

Check for updates

OPEN ACCESS

EDITED BY
HaiHui Huang,
Shaoguan University, China

REVIEWED BY
Dejun Kong,
Capital Medical University, China
Xi Wang,
Fujian Medical University, China

*CORRESPONDENCE
Yin Liu
✉ 17751495685@163.com

RECEIVED 04 June 2025
ACCEPTED 22 September 2025
PUBLISHED 10 October 2025

CITATION
Jin L, Yuan Z-z and Liu Y (2025) Multi-omics integration to identify immune-associated biomarkers and potential therapeutics in periodontitis.
Front. Med. 12:1640961.
doi: 10.3389/fmed.2025.1640961

COPYRIGHT
© 2025 Jin, Yuan 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.

Multi-omics integration to identify immune-associated biomarkers and potential therapeutics in periodontitis

Ling Jin[1], Zhong-zheng Yuan[2] and Yin Liu[2]*

[1]The First Outpatient Department, Stomatological Hospital of Jilin University, Changchun, China, [2]Wuxi Stomatological Hospital, Wuxi, Jiangsu, China

Background: Periodontitis, a chronic inflammatory disease of periodontal tissues, is linked to immune response and epigenetic modifications, with DNA methylation playing a crucial role. This study integrates transcriptomic and DNA methylation profiles from periodontitis patients to explore the immune microenvironment and identify potential biomarkers and therapeutic targets.
Methods: Transcriptomic and methylation profiles from 24 periodontitis patients were analyzed to evaluate the immune microenvironment and identify related abnormal genes. WGCNA was used to identify immune cell-associated genes. Subsequently, machine learning algorithms identified diagnostic biomarkers for periodontitis, which then validated in two cohorts with 247 and 310 periodontitis patients, respectively. Finally, network pharmacology analysis identified potential targeted drugs for the candidate genes.
Results: We obtained 23,528 differentially methylated sites and 1,641 differential expressed genes. Immune cell analysis identified eight abnormal cell types in periodontitis, and WGCNA highlighted two gene modules linked to these immune alterations. Machine learning with random forest and SVM identified nine key genes (ATP2C2, FAM43B, FOXA3, HSPA12A, KIF1C, NCS1, PGM1, RASSF6, SH2B2) with diagnostic efficacy, achieving high AUC scores across validation datasets. Network pharmacology analysis identified three drugs—bisphenol A, acetaminophen, and valproic acid—as potential regulators of these genes, offering new treatment avenues.
Conclusion: Through integrating s transcriptomic and DNA methylation profiles, nine genes have been filtered as potential diagnostic biomarkers of periodontitis. Drugs targeting these genes may serve as potential therapeutics for periodontitis. These findings reveal valuable insights into immune and epigenetic mechanisms in periodontitis, presenting new biomarkers and therapeutic options that may enhance clinical diagnosis and treatment of the disease and provide unique insights for further exploration of the pathogenesis of periodontitis and the development of related therapeutic drugs.

KEYWORDS

periodontitis, immune microenvironment, DNA methylation, machine learning, diagnostic biomarkers

Introduction

Periodontal disease is considered to be the most common disease in humans. The prevalence of periodontal disease is showing a significant increase (1), and globally, the prevalence of severe periodontal disease is 11%, affecting 743 million people (2). Epidemiologic surveys have shown that the leading cause of tooth loss worldwide is periodontitis, which is associated with a reduced quality of life and may cause a variety of other systemic health problems (3). Periodontitis is a chronic inflammatory condition affecting the tissues that support teeth, initiated by plaque buildup. This process results in progressive tissue destruction, formation of periodontal pockets, loss of attachment, and resorption of alveolar bone, ultimately causing tooth mobility, gum recession, and eventually, tooth loss (4). Previous studies have reported the complex molecular mechanisms of this periodontitis (5). However, the specific roles of genes, cell types, and cellular mechanisms in the development of periodontitis remain unclear, and there are currently no reliable early diagnostic markers or therapeutic targets available (6, 7). For instance, researchers found that chronic injury may alter transglutaminase gene expression, potentially playing a crucial role in remodeling and adaptation (8); It has been found that a significant link between miRNA in gingival sulcus fluid and the risk of periodontitis (9).

While bacteria are essential in initiating periodontitis, disease progression largely relies on the host's immune response. An excessive or imbalanced immune reaction to these microorganisms can speed up both the onset and advancement of periodontitis (10), accompanied by the release of various inflammatory mediators and cytokines (11). For example, prostaglandin E_2 (PGE_2), interleukin-1β (IL-1β), tumor necrosis factor-α (TNF-α) (12), IL-8 (13), and interferon-γ (IFN-γ) (14). Thus, the immune response of the host, particularly the cellular immune response, is crucial in regulating the equilibrium between the repair and damage of periodontal tissues (15). Therefore, current research on periodontitis focuses on understanding how the immune system and immunomodulatory factors influence periodontal inflammation and alveolar bone degradation, as well as the role of molecular regulatory networks in immune cell activation and differentiation (4).

To further elucidate the mechanisms underlying periodontitis, it is important to consider not only the immune response but also the epigenetic factors that regulate gene expression. Epigenetics refers to changes in gene expression that do not involve alterations to the underlying DNA sequence. Key epigenetic processes include DNA methylation, histone modifications, and chromatin remodeling. Recent studies suggest that chronic inflammatory conditions, such as periodontitis, can induce epigenetic changes, thereby modulating the immune response and contributing to disease progression. Growing evidence indicates that these epigenetic changes are linked to the development of periodontitis (15). In particular, epigenetic modifications occur in periodontal tissues during the periodontitis process. Currently, DNA methylation is the most studied epigenetic modification associated with periodontitis (16). DNA methylation is a widespread epigenetic alteration in eukaryotic cells, involving the attachment of methyl groups to cytosine residues within CpG dinucleotides. This modification can be either hypermethylation or hypomethylation, leading to the repression or activation of certain genes (17). DNA methylation of cytokine-encoding genes has been found in periodontal tissues of patients with periodontitis (18). For instance, the IL6 gene expression in the gingival tissues of patients with periodontitis was elevated compared to healthy controls (19). In addition, DNA methylation affects genes encoding interferons and chemokines (20). Recently, researchers investigated CpG methylation of 22 inflammatory candidate genes (21). These findings may provide some new insights into the relationship between altered methylation of encoded genes and periodontitis.

In this study, we hypothesize that integrating transcriptomic and DNA methylation profiles will reveal novel immune-related biomarkers and mechanistic links in periodontitis. To confirm it, we systematically integrated periodontitis-associated transcriptome and DNA methylation data to explore the immune microenvironment of periodontitis. We aimed to identify key immune biomarkers in multiple omics dimensions using a range of bioinformatics approaches (Figure 1). These findings may offer new insights for the development of diagnostic and therapeutic biomarkers for periodontitis.

Methods

Data source

DNA methylation and corresponding mRNA expression data from periodontitis patients were retrieved from the GEO database under accession numbers GSE173081 (DNA methylation, $N_{total} = 24$, $N_{periodontitis} = 12$, and $N_{healthy} = 12$) and GSE173078 (mRNA expression, $N_{total} = 24$, $N_{periodontitis} = 12$, and $N_{healthy} = 12$). Two additional independent datasets, GSE16134 (testing dataset1, $N_{total} = 310$, $N_{periodontitis} = 241$, and $N_{healthy} = 69$) and GSE10334 (testing dataset2, $N_{total} = 247$, $N_{periodontitis} = 183$, and $N_{healthy} = 64$), were used for testing. All FPKM expression values were normalized using a log2 transformation. All these datasets are publicly available and unrestricted re-use is permitted via the open license of GEO database.

DNA methylation profiles

The Illumina Human Methylation EPIC Array was used to analyze the methylation status of periodontitis patients ($N_{total} = 24$, $N_{periodontitis} = 12$, and $N_{healthy} = 12$). This bead chip covers more than 810,000 methylation sites per sample. The raw data were processed by the following steps: firstly, probes with a null value and located in sex chromosomes were removed. Then, probes that mapped to multiple genes or were not mapped to genes or containing SNPs were removed.

The *minfi* R package was used for the normalization of the raw Methylation EPIC Array data. Probes with a p-value < 0.05 and absolute detabeta ($|\Delta\beta|$) > 0.1 were considered differentially methylated.

Immune microenvironment analysis

The *xcell* R package was employed to estimate the abundance of 64 immune cell types in periodontitis patients, including various T-cell subtypes and other immune cells such as B cells, NK cells, monocytes, and macrophages. The abundance of immune cells in periodontitis patients was compared to that in healthy individuals to identify distinctive features for further investigation.

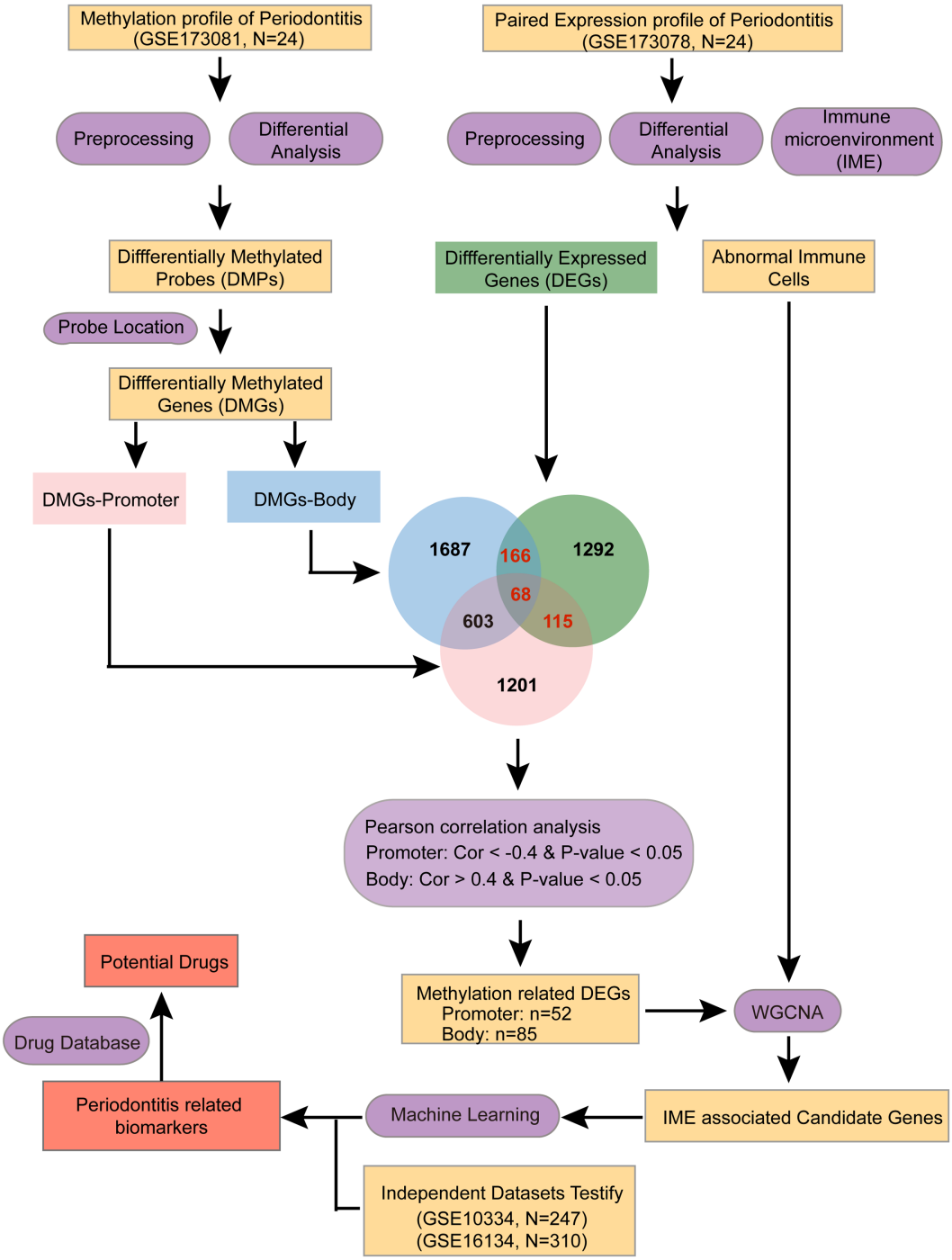


FIGURE 1
Workflow of the study.

Differential expression analysis

The *limma* R package was employed to analyze gene expression differences between periodontitis and control groups. Differentially expressed mRNAs were identified with an adjusted p-value < 0.05 and an absolute log2 fold change ≥ 0.263 (22). Subsequently, Pearson correlation analysis was conducted to assess the relationship between DNA methylation levels and gene expression. Only correlations with an absolute Pearson coefficient above 0.4 and a p-value below 0.05 were considered significant.

WGCNA

Co-expression networks were constructed using the WGCNA R package to analyze candidate genes showing correlated patterns in both methylation and expression levels, alongside abnormal immune

cell types in periodontitis patients. In this study, hierarchical clustering was used to group genes with similar expression patterns. These gene clusters were then linked to the altered immune cells in patients, and the most relevant genes within these clusters were selected for further investigation.

Machine learning

The *randomForest* R package was employed to build a periodontitis prediction model using the random forest method, which involved training and testing categorical models to identify gene combinations with high discriminatory power for distinguishing periodontitis from normal groups. The key genes were identified using an SVM algorithm with the *e1071* R package to construct an optimal diagnostic model.

Function enrichment analysis

We extracted all differentially expressed genes (DEGs) and differentially methylated genes (DMGs) for further functional enrichment analysis using the Metascape webserver. Enrichment analysis was conducted for KEGG pathways and Hallmark gene sets, with functions selected based on a false discovery rate of less than 0.05.

Statistical analyses were performed using R software (version 4.3.2). A *t*-test was used to assess differences between the two groups, and a *p*-value of less than 0.05 was considered statistically significant.

Results

Differently expressed and differentially methylated genes are associated with inflammatory and immune-related pathways in periodontitis

We first assessed methylation levels in patients with periodontitis. First, we performed differential analysis of the EPIC methylation array and obtained a total of 23,528 differentially methylated sites ($p < 0.05$, $|\Delta\beta| > 0.1$). Subsequently, we categorized the differentially methylated probes into promoter region probes (TSS200, TSS1500, 1stExon) and body region probes based on their location in the genome. Among them, there are 5,152 differentially methylated promoter region probes distributed on 2,489 genes and 4,814 differentially methylated body region probes, which fell on 2,784 genes (Figure 2A). Subsequently, we performed enrichment analysis of these differentially methylated genes. The results showed that the differentially methylated genes in the body region were mainly enriched in the Calcium signaling pathway, Wnt signaling pathway and other inflammation-related pathways (Figure 2B), while the differentially methylated genes in the promoter region were mainly enriched in the cMAP signaling pathway, the PI3K-Akt signaling pathway, and the Cytokine-cytokine receptor interaction. Receptor interaction and other immune-related pathways (Figure 2C).

To further elucidate the functional impact of these epigenetic modifications, we next examined the gene expression profiles in periodontitis patients. We analyzed the gene expression data of periodontitis patients to screen for genes abnormally expressed in periodontitis ($|\log2FC| > 0.263$, $p < 0.05$). We screened a total of 1,641 differential expressed genes, of which 398 were abnormally down-regulated and 1,243 were abnormally up-regulated (Figure 2D). Enrichment analysis of these differential genes showed that periodontitis-associated aberrantly expressed genes were mainly enriched in pathways such as Cytokine-cytokine receptor interaction, NF-kappa B signaling pathway and HIF-1 signaling pathway (Figure 2E). The NF-kappa B signaling pathway, in particular, plays a pivotal role in orchestrating inflammatory responses. Activation of NF-kappa B leads to the transcription of a variety of cytokines and chemokines that mediate inflammation, which is critical in the progression of periodontitis. This pathway can contribute to the persistence of inflammation, thereby exacerbating tissue destruction and bone resorption observed in periodontitis.

In our integrated analysis, we identified 349 genes that were both differentially methylated and differentially expressed. Notably, several key genes involved in inflammation and immune regulation were among these 349 genes. For example, MMP9, a matrix metalloproteinase known for its role in tissue remodeling and inflammatory processes, has been implicated in periodontal tissue degradation. Similarly, CD86, a critical co-stimulatory molecule involved in T-cell activation, and PTPRC (CD45), a regulator of immune cell signaling, underscore the immune involvement in periodontitis. Other genes such as IL2RA and IL21R are central to immune cell differentiation and activation, while FAM43B and FOXA3 have emerged as potential diagnostic markers in our analysis. These gene-specific findings reinforce the biological relevance of our integrated analysis and suggest that the dysregulation of these key genes may contribute significantly to the pathogenesis of periodontitis (Figure 2F).

Altered immune cells in periodontitis linked to differentially expressed gene modules regulated by aberrant methylation

We evaluated the immune microenvironment of periodontitis patients based on the xcell algorithm. The results showed that the abundance of immune cells such as Astrocytes, Granulocyte-Macrophage Progenitor (GMP), Hepatocyte, Monocyte, Neutrophil, Plasma cells, and Prevacidocytes was significantly increased while Platelets were significantly decreased in periodontitis patients (Figure 3A; Supplementary Figure 1). It is important to note that the detection of hepatocyte signatures in gingival tissue is unexpected. This may be due to the inherent limitations of the xCell algorithm, which relies on gene expression profiles that can sometimes overlap among different cell types. The "Hepatocyte" signal observed might represent a similar cell population with a related expression profile rather than true hepatocytes. Further experimental validation is needed to clarify this observation. It should be noted that the detection of "hepatocyte" and "platelet" signatures likely reflects algorithmic limitations of bulk transcriptomic deconvolution rather than the true presence of these cell types in gingival tissue.

These alterations in immune cell composition suggest an imbalance in immune regulation, potentially driven by underlying epigenetic changes. Subsequently, we further screened the aberrant genes regulated by methylation, and the screening criteria were, body region differentially methylated genes, whose methylation level was

FIGURE 2
Transcriptome- and DNA mathylation-based screening of periodontitis. **(A)** Volcano plot of differentially methylated probes (periodontitis vs. control). **(B)** Enrichment analysis of differentially methylated probes which located at gene body region. **(C)** Enrichment analysis of differentially methylated probes which located at gene promoter region. **(D)** Volcano plot of differentially expressed genes (periodontitis vs. control). **(E)** Enrichment analysis of differentially expressed genes. **(F)** Venn plot showed the overlap of differentially methylated probes and differentially expressed genes.

positively correlated with the expression level ($R > 0.4$, $p < 0.05$), and promoter region differentially methylated genes, whose methylation level was negatively correlated with the expression level ($R < -0.4$, $p < 0.05$). These correlation thresholds were chosen to ensure a moderate to strong association between methylation changes and gene expression regulation while minimizing false positives. Previous studies have used similar cutoffs to establish meaningful methylation-expression relationships in disease contexts (23, 24). Finally, we screened 132 eligible candidate genes (Figure 3B).

To connect these findings with the observed immune cell alterations, we investigated whether the aberrantly methylated genes might drive changes in immune cell profiles. We performed WGCNA

10.3389/fmed.2025.1640961

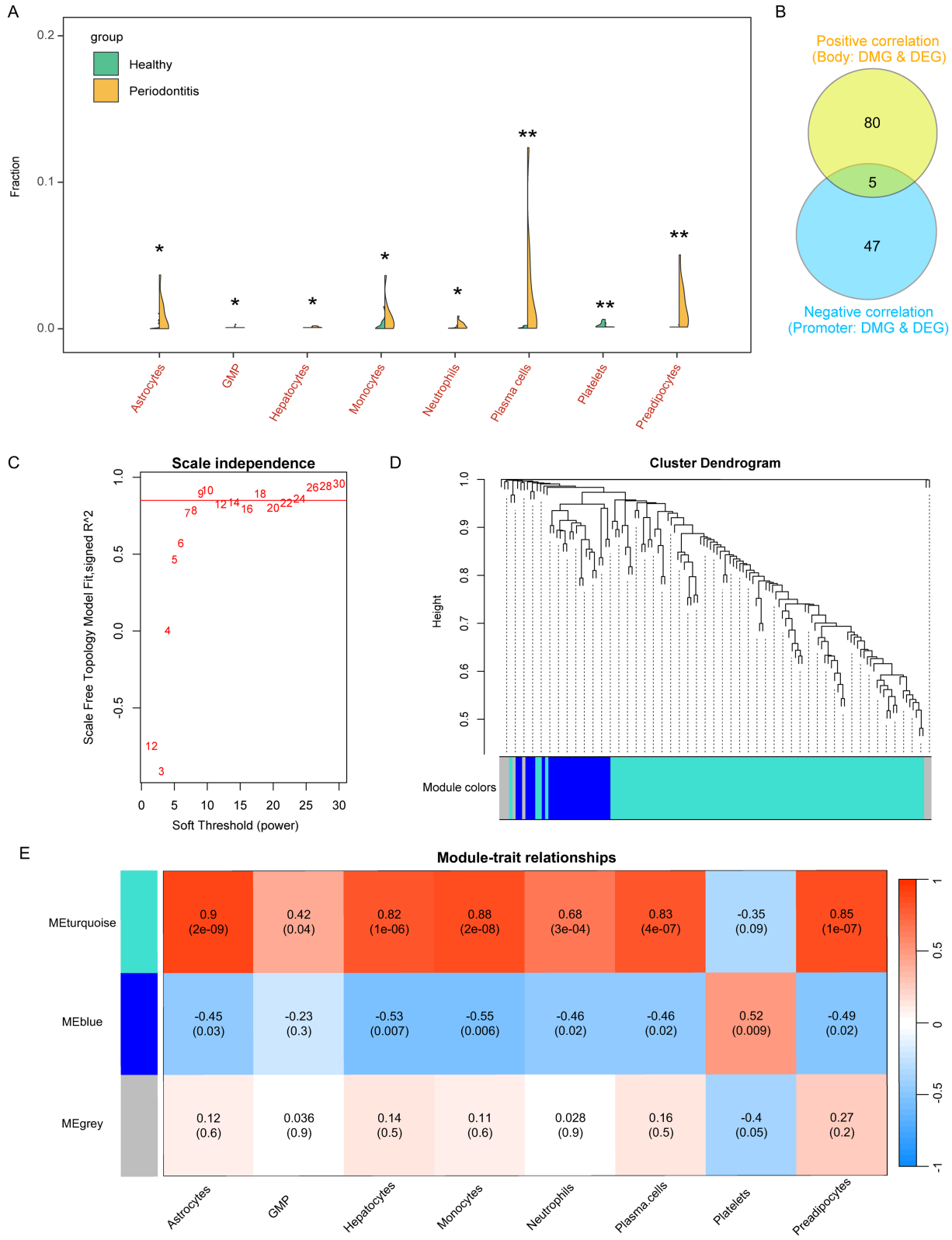


FIGURE 3
Assessment of the immune microenvironment for periodontitis and screening of relevant gene modules. **(A)** Violin plot showed the altered immune cells of periodontitis. GMP, Granulocyte-Macrophage Progenitor. **(B)** Correlation ship between differentially methylated probes and differentially expressed genes. **(C)** Power of WGCNA co-expression network. **(D)** Cluster dendrogram of candidate genes which were correlated in DNA methylation and gene expression in periodontitis. **(E)** Immune cells associated co-expression modules in periodontitis.

analysis to screen the co-expression modules of altered immune cells based on the expression levels of these 132 candidate genes with the abundance of the above mentioned 8 altered immune cells associated with periodontitis. The results showed that there were three expression patterns of these candidate genes (Figures 3C,D), among which, the MEturquoise module (contains 100 candidate genes) was significantly associated with abnormally elevated immune cells such as Astrocytes, GMP, Hepatocyte, Monocyte, Neutrophil, Plasma cells, and Previpocytes. correlated, while MEblue (contains 25 candidate genes) significantly correlated with Astrocytes, Hepatocyte, Monocyte, Neutrophil, Plasma cells, Previpocytes and Platelets (Figure 3E; Supplementary Table S1). These findings indicate that epigenetic regulation, as reflected by aberrant methylation, may influence immune cell composition by modulating the expression of gene modules relevant to immune functions.

Machine learning-based screening for multi-omics diagnostic biomarkers in periodontitis patients

Through WGCNA analysis, we found that MEblue and MEturquoise are associated with altered immune cells in periodontitis patients. Among them, MEblue contains 25 candidate genes while MEturquoise contains 100 candidate genes. These gene modules exhibited significant correlations with immune cell types that are dysregulated in periodontitis, including monocytes, neutrophils, and plasma cells (Figure 3E). The enrichment analysis of these genes revealed their involvement in immune-related pathways (such as leukocyte activation, toll-like receptor 2 signaling pathway), further underscoring their biological relevance (Supplementary Figure 2). The strong correlation between these genes and immune cell alterations suggests their potential role in immune dysregulation and inflammation in periodontitis. Thus, we selected these genes for machine learning analysis to identify the most informative biomarkers for disease classification. We first performed random forest modeling for the 25 genes in the MEblue module. The results show that the random forest model has the optimal classification efficacy when the number of genes in the model reaches 3 (Figure 4A). Subsequently, we show the gene scores for each node in the random forest and select the top3 genes (HSPA12A, ATP2C2, and NCS1) (Figure 4B). These three genes have been previously implicated in periodontitis-related processes. For instance, HSPA12A is known to regulate inflammatory responses (25), ATP2C2 plays a critical role in immune microenvironment (26), and NCS1 is associated with immunotherapy and prognosis of cancer (27). Next, we construct a classification model based on SVM for these top3 genes. The results show that in the training set, the classification efficiency of this 3-gene model reaches 0.826 (AUC = 0.826, Figure 4C), while in the testing dataset1, the AUC of this model is also as high as 0.775 (Figure 4D), and in the testing dataset2, the AUC value is 0.752 (Figure 4E). This suggests that this 3-gene model has a better diagnostic efficacy for periodontitis patients. In the MEturquoise module, we found that the 6-gene random forest model had the best classification efficacy (Figure 5A). We then ranked the genes in the model based on importance and selected the top6 genes (PGM1, RASSF6, KIF1C, SH2B2, FOXA3, and FAM43B) (Figure 5B). The six genes selected from the MEturquoise module are also intricately linked to immune and inflammatory pathways. For example, PGM1 and RASSF6 are associated with macrophage (28). KIF1C could regulate the podosome dynamics in macrophages (29). The immunologic significance of SH2B2 is related to the invasion of colon adenocarcinoma (30). FOXA3 is a transcriptional activator that is associated with signal transduction in tumors (31). Additionally, FAM43B could repress the cell proliferation and is regulated by DNA methylation (32). Similarly, we construct SVM classifiers based on these 6 candidate genes. The results show that this model has AUC = 0.819 (Figure 5C) in the training set, while in testing dataset1 and testing dataset2, the AUC is 0.860 (Figure 5D) and 0.816 (Figure 5E), respectively. Detailed performance metrics are also summarized in Supplementary Table S2. These results suggesting that these models have effective efficacy for periodontitis diagnosis.

Identification of target drugs for periodontitis patients based multi-omics diagnostic biomarkers

Based on epigenome and transcriptomics, we screened 9 periodontitis diagnostic genes in the periodontitis immune microenvironment. Subsequently, we further explored potential target drugs for these 9 genes. We constructed a drug-targeting network for these genes based on the CTD database and identified 345 drugs/compounds targeting these 9 genes (Figure 6). Through further network analysis, we screened out 3 drugs/compounds targeting all 9 genes simultaneously: bisphenol A, Acetaminophen and Valproic Acid (Supplementary Table S3). Bisphenol A, though primarily considered an environmental contaminant, has been implicated in immune modulation and inflammatory responses (33). Acetaminophen is widely used as an analgesic and has been shown to modulate oxidative stress pathways, which are relevant in periodontitis pathology (34). Valproic acid, a histone deacetylase inhibitor, has demonstrated anti-inflammatory effects and potential benefits in immune-related conditions (35). These insights support the relevance of these drugs in the context of periodontitis and highlight their possible regulatory roles in disease-associated pathways. Among them, Acetaminophen and Valproic Acid are FDA-approved drugs with better results in analgesia. In this study, we found for the first time that they are associated with periodontitis-related targets, which provides a new idea for the subsequent screening of potential periodontitis-related drugs. Their repurposing for periodontitis could offer advantages such as well-characterized pharmacokinetics and widespread clinical availability. However, the potential off-target effects and adverse reactions-such as hepatotoxicity for Acetaminophen and the broad systemic effects associated with Valproic Acid-necessitate further investigation in the context of periodontitis. Additional preclinical studies and clinical trials are warranted to optimize dosing, evaluate long-term safety, and establish their efficacy as adjuncts in periodontitis management.

Discussion

While numerous studies have highlighted the immune microenvironment's involvement in the development of periodontitis, the exact mechanisms through which it affects the onset and progression of the disease are still not fully understood (21–23). It has

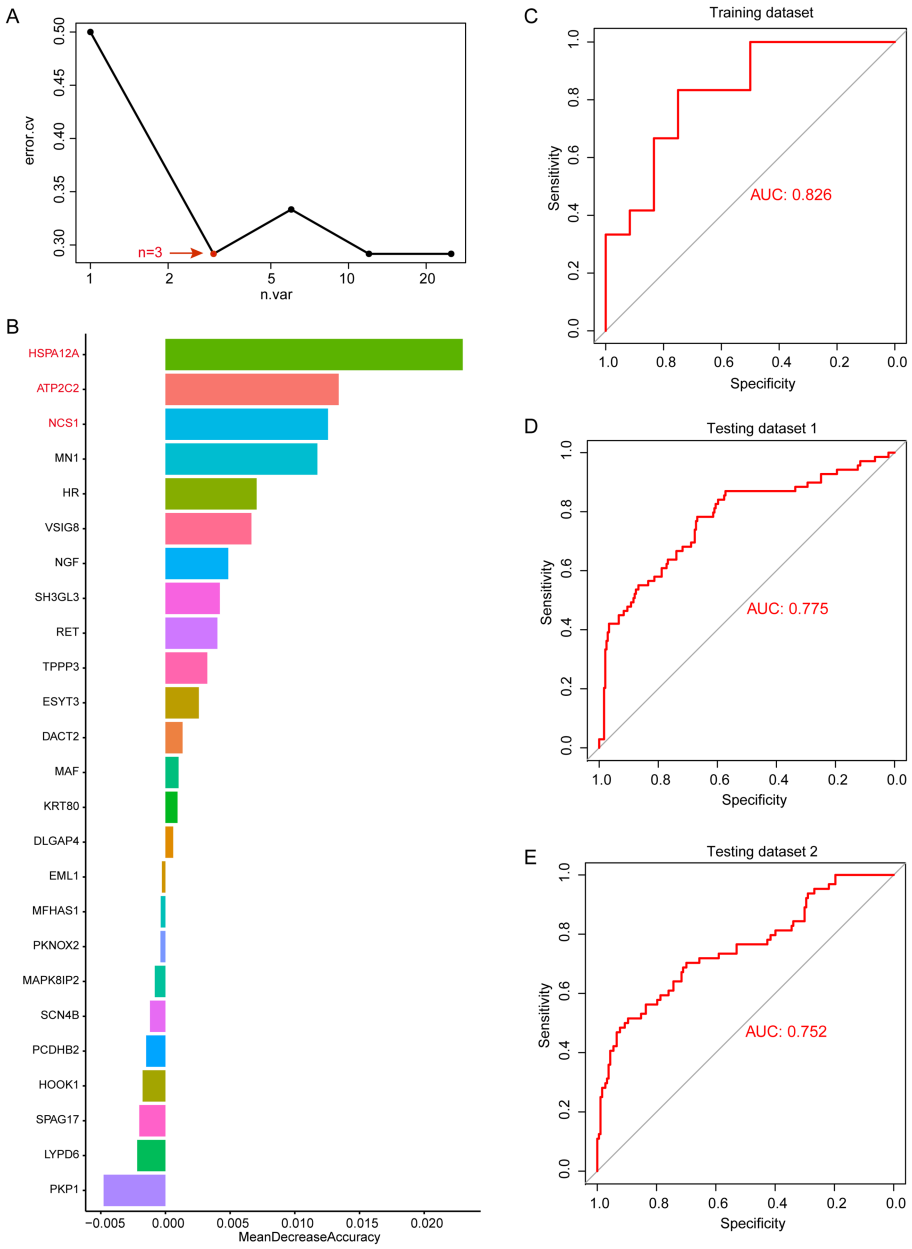


FIGURE 4
Random forest analysis in MEblue. **(A)** error.cv. plot of random forest analysis. Here, the random forest model has the optimal classification efficacy when the number of genes in the model reaches 3. **(B)** Mean decrease accuracy of random forest model. **(C)** SVM performance in training datasets. **(D)** SVM performance in testing datasets 1. **(E)** SVM performance in testing datasets 2.

been found that periodontitis is not only affected by the transcriptional level but also involves epigenetic alterations. Nevertheless, there are limited studies that have identified immune-related genes linked to periodontitis across various histological layers, which could potentially serve as important clinical biomarkers for the disease. In this study, we first investigated the immune microenvironment of periodontitis and identified candidate genes that showed both abnormal methylation and expression patterns in periodontitis samples, utilizing epigenomic and transcriptomic approaches. Subsequently, through a multi-dataset machine learning algorithm, we further narrowed down these candidate genes to nine key genes with diagnostic efficacy for periodontitis, namely, ATP2C2, FAM43B, FOXA3, HSPA12A, KIF1C, NCS1, PGM1, RASSF6, and SH2B2. We then further analyzed these genes by network pharmacology to screen for their potential drug targets. This study revealed the association of key genes related to the immune microenvironment with periodontitis at the epigenetic and transcriptional levels, and screened for drug targets that could regulate these key genes through the drug target network. Our research offers significant insights into the potential use of these key genes as diagnostic and therapeutic markers for improving the clinical management of periodontitis.

ATP2C2 is involved in calcium transmembrane transport, intracellular calcium ion homeostasis, and manganese ion transport (36). FAM43B has been found to control innate immunity through

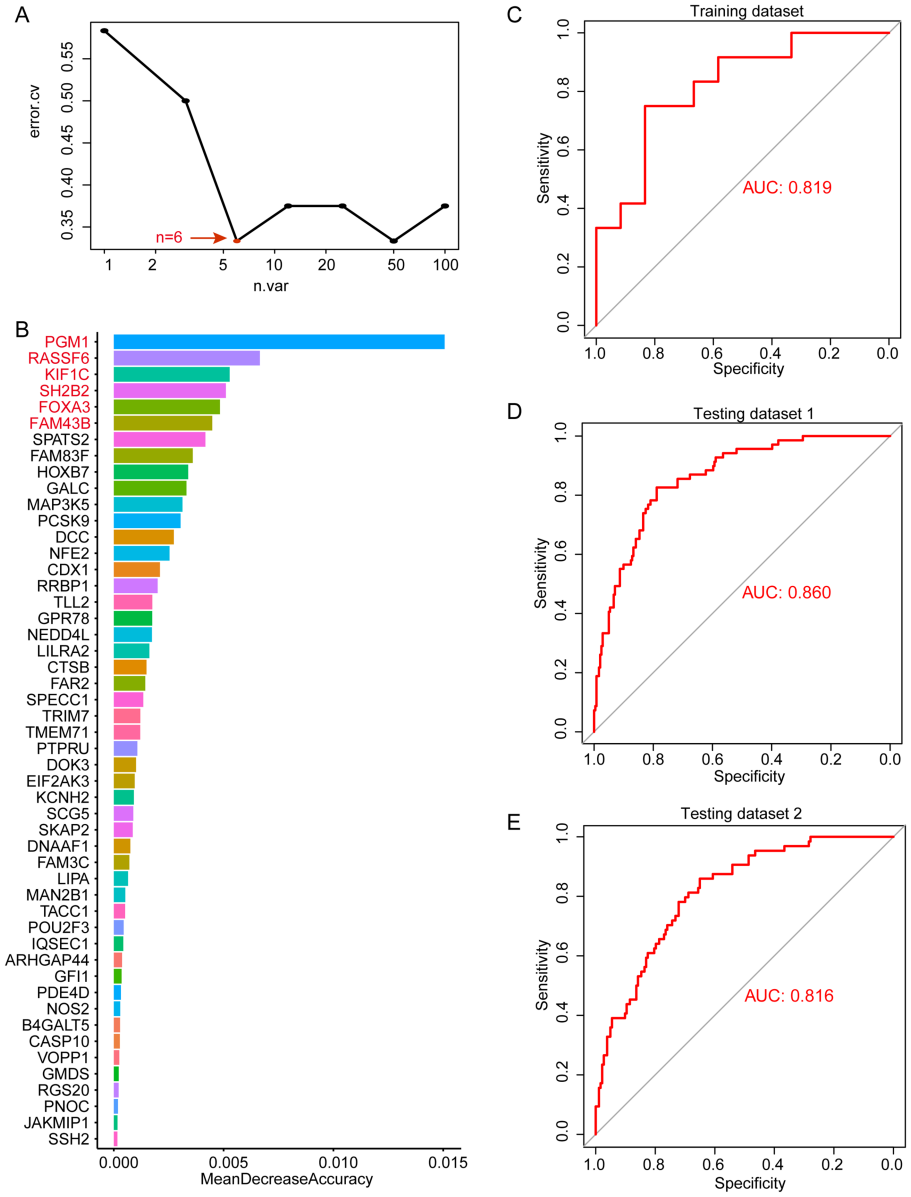


FIGURE 5
Random forest analysis in MEturquoise. **(A)** error.cv. plot of random forest analysis. Here, the random forest model has the optimal classification efficacy when the number of genes in the model reaches 6. **(B)** Mean decrease accuracy of random forest model. **(C)** SVM performance in training datasets. **(D)** SVM performance in testing datasets 1. **(E)** SVM performance in testing datasets 2.

Epigenetic Regulation (37). FOXA3 encodes a forkhead-like DNA-binding protein that interacts with chromatin. It also plays a role in the regulation of metabolism as well as organ differentiation. FOXA3 methylation has been found to cause dedifferentiation and sorafenib resistance in hepatocellular carcinoma (38). HSPA12A, which is predicted to have ATP-binding activity and is located in extracellular exosomes, was found to promote nuclear PKM2-mediated polarization of M1 macrophages (39). The protein encoded by KIF1C belongs to a family of kinesin-like proteins that transport APC-dependent mRNAs to cellular protrusions (40) and can re-localize GLUT4 to immune-modification-positive cell sites (41). NCS1, a member of the neuronal calcium sensor gene family, is a key Ca2 + −binding protein thought to play a role in cell proliferation and immune infiltration (27). The protein encoded by this gene is an isoform of phosphoglucomutase (PGM) and is associated with M2 macrophages and TFH cells and their surface markers CD163 and CXCR5 (42). RASSF6 encodes a member of the Ras-associated structural domain family (RASSF), and the protein encoded by this gene is a Ras effector protein that induces apoptosis. In acute lymphoblastic leukemia (ALL), there is a high prevalence of aberrant RASSF6 promoter methylation, and its DNA methylation status has the potential to serve as a biomarker for assessing MRD levels in ALL patients (43). SH2B2 encodes a protein expressed in B lymphocytes that undergoes tyrosine phosphorylation in response to B cell receptor stimulation and plays a role in signaling in the Shc/Grb2 pathway (44).

In conclusion, through our integration of DNA methylation profiles and transcriptomes of periodontitis patients, we assessed the immune microenvironment of periodontitis patients and screened nine diagnostic

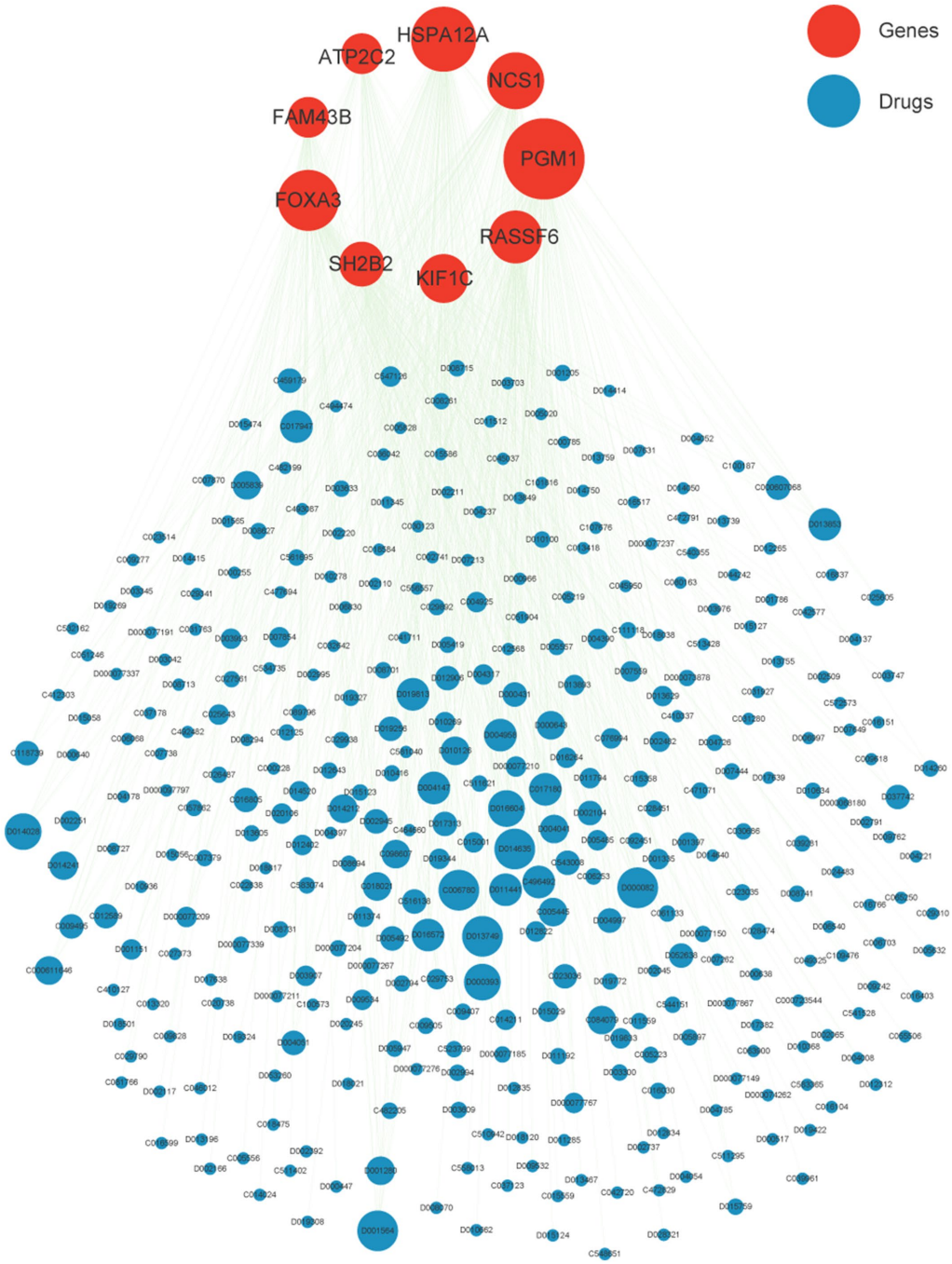


FIGURE 6
Drug network of nine crucial genes. Red nodes represent nine crucial gene while blue is potential target drugs. The node size indicates the degree of the network.

markers related to periodontitis patients based on machine-learning algorithms, and screened for relevant targeted drugs. This finding will provide new insights for subsequent diagnosis and treatment of periodontitis. Furthermore, our study builds on previous research using machine learning to identify biomarkers in immune-related diseases. For instance, studies on interactomic hub gene prediction in PBMCs for type 2 diabetes mellitus, dyslipidemia, and periodontitis have demonstrated the potential of network-based approaches in identifying key regulatory genes (45). Additionally, machine learning models for predicting rheumatoid arthritis based on ACPA autoantibody development in the presence of non-HLA gene polymorphisms highlight the utility of such methods in complex diseases (46). Similarly, the prediction of interactomic hub genes in rheumatoid arthritis using peripheral mononuclear cells underscores the importance of transcriptomic and network-based analyses in understanding immune-related pathologies (47). Our study contributes to this growing body of research by

identifying key diagnostic genes and their potential drug interactions in periodontitis. Meanwhile, our diagnostic models, with AUC values ranging from 0.75 to 0.86, compare favorably with existing periodontitis biomarkers, which often rely on single-parameter assessments such as probing depth, clinical attachment loss, or inflammatory mediators in gingival crevicular fluid (48). The integration of epigenetic and transcriptomic data in our models not only improves diagnostic accuracy but also captures the complexity of the disease's molecular basis. This multi-omics approach allows for a more comprehensive evaluation of the disease state and may facilitate the development of personalized treatment strategies. The identification of these drugs through a multi-omics approach presents a novel strategy for periodontitis therapy. In terms of efficacy, the FDA-approved drugs Acetaminophen and Valproic Acid have well-documented pharmacological profiles that may enhance their potential as adjunct therapies. They offer the possibility of modulating key molecular mechanisms underlying periodontitis, such as oxidative stress and immune regulation. However, while conventional therapies focus on bacterial control and symptomatic relief, these drugs may provide benefits by directly impacting the disease's molecular drivers. Regarding safety, current standard therapies generally have minimal systemic side effects but may not fully address the inflammatory and tissue-degradative components of periodontitis. In contrast, the off-target effects of Acetaminophen (e.g., hepatotoxicity) and Valproic Acid (e.g., gastrointestinal and metabolic disturbances) require careful dosing and monitoring.

Overall, these findings underscore the clinical and biological significance of integrating multi-omics data to identify potential therapeutic agents. The approach not only enhances our understanding of periodontitis pathogenesis but also opens new avenues for developing targeted interventions that may complement existing treatment modalities.

Despite the promising findings of our study, several limitations should be acknowledged. First, our analyses relied on publicly available datasets with relatively small sample sizes, which may limit the generalizability of the results. The lack of detailed demographic information, such as age and gender, may also introduce selection bias and restrict applicability across broader populations. Future large-scale studies with demographically matched cohorts are warranted to address these concerns. Second, while the use of multi-dataset machine learning improved robustness, potential confounders (including patient demographics, disease severity, and sample processing) could still influence the outcomes. Integrating additional omics layers, such as proteomics and metabolomics, may provide a more comprehensive understanding of periodontitis pathogenesis.

Moreover, the current study provides predictive insights into immune alterations in periodontitis based on bulk transcriptomic deconvolution. However, bulk analyses cannot fully capture the complexity of the immune microenvironment, which ideally requires single-cell transcriptomic and spatially resolved approaches. Importantly, all conclusions are computationally derived without protein-level or *in vivo* validation. Future research should therefore include experimental confirmation, such as immunohistochemistry, flow cytometry, and animal models, to validate the biological and therapeutic relevance of the identified biomarkers and drug candidates. Specifically, preclinical testing of acetaminophen and valproic acid will be crucial to determine their mechanistic roles and feasibility as adjunctive therapies for periodontitis.

Conclusion

In conclusion, through our integration of DNA methylation profiles and transcriptomes of periodontitis patients, we assessed the immune microenvironment of periodontitis patients and screened nine diagnostic markers related to periodontitis patients based on machine-learning algorithms, and screened for relevant targeted drugs. This finding will provide new insights for subsequent diagnosis and treatment of periodontitis.

Data availability statement

The datasets used and analyzed during the current study are downloaded from GEO database under accession numbers: GSE173081, GSE173078, GSE10334, and GSE16134.

Ethics statement

The studies involving humans were approved by Ethics Committee of the Wuxi Stomatological Hospital. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin because this study is based on public data sources and did not require any intervention or experiment related to patients.

Author contributions

LJ: Writing – review & editing, Writing – original draft, Formal analysis, Data curation, Methodology. Z-zY: Resources, Writing – original draft, Visualization. YL: Project administration, Supervision, Writing – review & editing, Conceptualization, Writing – original draft.

Funding

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

Conflict of interest

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

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial

intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

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/fmed.2025.1640961/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Assessment of the immune microenvironment for periodontitis. Violin plot showed the abundance of immune cells in periodontitis. DC, Dendritic Cell; CLP, Common Lymphoid Progenitor; CMP, Common Myeloid Progenitor; GMP, Granulocyte-Macrophage Progenitor; HSC, Hematopoietic Stem Cell; MEP, Megakaryocyte-Erythroid Progenitor; MPP, Multipotent Progenitor; MSC, Mesenchymal Stem Cell; NKT, Natural Killer T Cell.

SUPPLEMENTARY FIGURE 2
Enrichment analysis of genes in MEturquoise and MEblue. **(A)** Enrichment analysis of 100 module genes in MEturquoise. **(B)** Enrichment analysis of 25 module genes in MEblue.

References

1. Chen J, Ou L, Liu W, Gao F. Exploring the molecular mechanisms of ferroptosis-related genes in periodontitis: a multi-dataset analysis. *BMC Oral Health*. (2024) 24:611. doi: 10.1186/s12903-024-04342-2

2. Kwon T, Lamster IB, Levin L. Current concepts in the Management of Periodontitis. *Int Dent J*. (2021) 71:462–76. doi: 10.1111/idj.12630

3. Eke PI, Dye BA, Wei L, Slade GD, Thornton-Evans GO, Borgnakke WS, et al. Update on prevalence of periodontitis in adults in the United States: NHANES 2009 to 2012. *J Periodontol*. (2015) 86:611–22. doi: 10.1902/jop.2015.140520

4. Zhang X, Ren L, Yan X, Shan Y, Liu L, Zhou J, et al. Identification of immune-related lncRNAs in periodontitis reveals regulation network of gene-lncRNA-pathway-immunocyte. *Int Immunopharmacol*. (2020) 84:106600. doi: 10.1016/j.intimp.2020.106600

5. Slots J. Periodontitis: facts, fallacies and the future. *Periodontol 2000*. (2017) 75:7–23. doi: 10.1111/prd.12221

6. Qasim SSB, Al-Otaibi D, Al-Jasser R, Gul SS, Zafar MS. An evidence-based update on the molecular mechanisms underlying periodontal diseases. *Int J Mol Sci*. (2020) 21:3829. doi: 10.3390/ijms21113829

7. Sun Q, Zhu E. Molecular mechanism and diagnostic marker investigation of endoplasmic reticulum stress on periodontitis. *BMC Oral Health*. (2023) 23:135. doi: 10.1186/s12903-023-02822-5

8. Curro M, Matarese G, Isola G, Caccamo D, Ventura VP, Cornelius C, et al. Differential expression of transglutaminase genes in patients with chronic periodontitis. *Oral Dis*. (2014) 20:616–23. doi: 10.1111/odi.12180

9. Isola G, Santonocito S, Distefano A, Polizzi A, Vaccaro M, Raciti G, et al. Impact of periodontitis on gingival crevicular fluid miRNAs profiles associated with cardiovascular disease risk. *J Periodontal Res*. (2023) 58:165–74. doi: 10.1111/jre.13078

10. Amano A. Host-parasite interactions in periodontitis: microbial pathogenicity and innate immunity. *Periodontol*. (2010) 54:9–14. doi: 10.1111/j.1600-0757.2010.00376.x

11. Kayal RA. The role of osteoimmunology in periodontal disease. *Biomed Res Int*. (2013) 2013:639368. doi: 10.1155/2013/639368

12. Gorska R, Gregorek H, Kowalski J, Laskus-Perendyk A, Syczewska M, Madalinski K. Relationship between clinical parameters and cytokine profiles in inflamed gingival tissue and serum samples from patients with chronic periodontitis. *J Clin Periodontol*. (2003) 30:1046–52. doi: 10.1046/j.0303-6979.2003.00425.x

13. Gursoy UK, Kononen E, Uitto VJ. Stimulation of epithelial cell matrix metalloproteinase (MMP-2, −9, −13) and interleukin-8 secretion by fusobacteria. *Oral Microbiol Immunol*. (2008) 23:432–4. doi: 10.1111/j.1399-302X.2008.00453.x

14. Gemmell E, Seymour GJ. Immunoregulatory control of Th1/Th2 cytokine profiles in periodontal disease. *Periodontol*. (2000) 2004:21–41. doi: 10.1111/j.0906-6713.2004.003557.x

15. Luo Y, Peng X, Duan D, Liu C, Xu X, Zhou X. Epigenetic regulations in the pathogenesis of periodontitis. *Curr Stem Cell Res Ther*. (2018) 13:144–50. doi: 10.2174/1574888X12666170718161740

16. Kojima A, Kobayashi T, Ito S, Murasawa A, Nakazono K, Yoshie H. Tumor necrosis factor-alpha gene promoter methylation in Japanese adults with chronic periodontitis and rheumatoid arthritis. *J Periodontal Res*. (2016) 51:350–8. doi: 10.1111/jre.12314

17. Grummt I, Pikaard CS. Epigenetic silencing of RNA polymerase I transcription. *Nat Rev Mol Cell Biol*. (2003) 4:641–9. doi: 10.1038/nrm1171

18. Zhang S, Barros SP, Moretti AJ, Yu N, Zhou J, Preisser JS, et al. Epigenetic regulation of TNFA expression in periodontal disease. *J Periodontol*. (2013) 84:1606–16. doi: 10.1902/jop.2013.120294

19. Kobayashi T, Ishida K, Yoshie H. Increased expression of interleukin-6 (IL-6) gene transcript in relation to IL-6 promoter hypomethylation in gingival tissue from patients with chronic periodontitis. *Arch Oral Biol*. (2016) 69:89–94. doi: 10.1016/j.archoralbio.2016.05.018

20. Zhang S, Crivello A, Offenbacher S, Moretti A, Paquette DW, Barros SP. Interferon-gamma promoter hypomethylation and increased expression in chronic periodontitis. *J Clin Periodontol*. (2010) 37:953–61. doi: 10.1111/j.1600-051X.2010.01616.x

21. Schulz S, Immel UD, Just L, Schaller HG, Glaser C, Reichert S. Epigenetic characteristics in inflammatory candidate genes in aggressive periodontitis. *Hum Immunol*. (2016) 77:71–5. doi: 10.1016/j.humimm.2015.10.007

22. Tu Q, Zhao R, Lu N. Evaluation of the diagnostic utility of immune microenvironment-related biomarkers in endometriosis using multidimensional transcriptomic data. *J Assist Reprod Genet*. (2024) 41:3213–23. doi: 10.1007/s10815-024-03261-z

23. Hashmi A, Cahill GL, Zaldana M, Davis G, Cronin BJ, Brandel MG, et al. Can head circumference be used as a proxy for intracranial volume in patients with craniosynostosis? *Ann Plast Surg*. (2019) 82:S295–300. doi: 10.1097/SAP.0000000000001803

24. Zhang H, Zhao L, Li S, Wang J, Feng C, Li T, et al. N6-Methylandenosine-related lncRNAs in tumor microenvironment are potential prognostic biomarkers in Colon Cancer. *Front Oncol*. (2021) 11:697949. doi: 10.3389/fonc.2021.697949

25. Li Y, Zhang X, Jiang G, Min X, Kong Q, Liu L, et al. Downregulation of HSPA12A protects heart against sepsis through suppressing mTOR-mediated inflammatory response in cardiomyocytes. *Int Immunopharmacol*. (2025) 145:113721. doi: 10.1016/j.intimp.2024.113721

26. Zhao M, Zhang Q, Song Z, Lei H, Li J, Peng F, et al. ATP2C2 as a novel immune-related marker that defines the tumor microenvironment in triple-negative breast cancer. *Transl Cancer Res*. (2023) 12:1802–15. doi: 10.21037/tcr-23-83

27. Wang GC, Gan X, Zeng YQ, Chen X, Kang H, Huang SW, et al. The role of NCS1 in immunotherapy and prognosis of human Cancer. *Biomedicine*. (2023) 11:2765. doi: 10.3390/biomedicines11102765

28. Sanada Y, Kumoto T, Suehiro H, Nishimura F, Kato N, Hata Y, et al. RASSF6 expression in adipocytes is down-regulated by interaction with macrophages. *PLoS One*. (2013) 8:e61931. doi: 10.1371/journal.pone.0061931

29. Kopp P, Lammers R, Aepfelbacher M, Woehlke G, Rudel T, Machuy N, et al. The kinesin KIF1C and microtubule plus ends regulate podosome dynamics in macrophages. *Mol Biol Cell*. (2006) 17:2811–23. doi: 10.1091/mbc.e05-11-1010

30. Bai N, Liu M, Li Q. Prognostic and immunologic significance of SH2B2 in colon adenocarcinoma and its relationship to proliferation, migration, and invasion. *Comb Chem High Throughput Screen*. (2024). doi: 10.2174/0113862073346075241118092413

31. Li N, Li Y, Gao H, Li J, Ma X, Liu X, et al. Forkhead-box A3 (FOXA3) represses cancer stemness and partially potentiates chemosensitivity by targeting metastasis-associated in colon cancer 1 (MACC1) signaling pathway in colorectal cancer cells. *Curr Cancer Drug Targets*. (2021) 21:254–67. doi: 10.2174/1568009620666201207150632

32. Xu X, Liu RF, Wan BB, Xing WM, Huang J, Han ZG. Expression of a novel gene FAM43B repressing cell proliferation is regulated by DNA methylation in hepatocellular carcinoma cell lines. *Mol Cell Biochem*. (2011) 354:11–20. doi: 10.1007/s11010-011-0800-y

33. Das S, Mukherjee U, Biswas S, Banerjee S, Karmakar S, Maitra S. Unravelling bisphenol A-induced hepatotoxicity: insights into oxidative stress, inflammation, and energy dysregulation. *Environ Pollut*. (2024) 362:124922. doi: 10.1016/j.envpol.2024.124922

34. Palya M, Chevere JM, Drum M, Fowler S, Nusstein J, Reader A, et al. Pain reduction of ibuprofen sodium dihydrate alone and in combination with acetaminophen

in an untreated endodontic pain model: a randomized, double-blind investigation. *J Endodont*. (2024) 50:881–8. doi: 10.1016/j.joen.2024.04.005

35. Uzel G, Oylumlu E, Durmus L, Ciraci C. Duality of Valproic acid effects on inflammation, oxidative stress and autophagy in human eosinophilic cells. *Int J Mol Sci*. (2023) 24:13446. doi: 10.3390/ijms241713446

36. Makena MR, Ko M, Mekile AX, Senoo N, Dang DK, Warrington J, et al. Secretory pathway ca(2+)-ATPase SPCA2 regulates mitochondrial respiration and DNA damage response through store-independent calcium entry. *Redox Biol*. (2022) 50:102240. doi: 10.1016/j.redox.2022.102240

37. Shi P, Guo Y, Su Y, Zhu M, Fu Y, Chi H, et al. SUMOylation of DDX39A alters binding and export of antiviral transcripts to control innate immunity. *J Immunol*. (2020) 205:168–80. doi: 10.4049/jimmunol.2000053

38. Zhou T, Li S, Xiang D, Liu J, Sun W, Cui X, et al. M6A RNA methylation-mediated HNF3gamma reduction renders hepatocellular carcinoma dedifferentiation and sorafenib resistance. *Signal Transduct Target Ther*. (2020) 5:296. doi: 10.1038/s41392-020-00299-0

39. Kong Q, Li N, Cheng H, Zhang X, Cao X, Qi T, et al. *HSPA12A* is a novel player in nonalcoholic steatohepatitis via promoting nuclear PKM2-mediated M1 macrophage polarization. *Diabetes*. (2019) 68:361–76. doi: 10.2337/db18-0035

40. Pichon X, Moissoglu K, Coleno E, Wang T, Imbert A, Robert MC, et al. The kinesin KIF1C transports APC-dependent mRNAs to cell protrusions. *RNA*. (2021) 27:1528–44. doi: 10.1261/rna.078576.120

41. Marampon F, Antinozzi C, Corinaldesi C, Vannelli GB, Sarchielli E, Migliaccio S, et al. The phosphodiesterase 5 inhibitor tadalafil regulates lipidic homeostasis in human skeletal muscle cell metabolism. *Endocrine*. (2018) 59:602–13. doi: 10.1007/s12020-017-1378-2

42. Zhao J, Huang S, Tan D, Yang K, Chen M, Jia X, et al. PGM1 and ENO1 promote the malignant progression of bladder Cancer via comprehensive analysis of the m6A signature and tumor immune infiltration. *J Oncol*. (2022) 2022:1–15. doi: 10.1155/2022/8581805

43. Younesian S, Shahkarami S, Ghaffari P, Alizadeh S, Mehrasa R, Ghaffari SH. Residual methylation of tumor suppressor gene promoters, RASSF6 and RASSF10, as novel biomarkers for minimal residual disease detection in adult acute lymphoblastic leukemia. *Ann Hematol*. (2019) 98:2719–27. doi: 10.1007/s00277-019-03775-y

44. Zhao L, Yuan L, Li F, Zhang X, Tian H, Ma Z, et al. Whole-genome resequencing of Hu sheep identifies candidate genes associated with agronomic traits. *J Genet Genomics*. (2024) 51:866–76. doi: 10.1016/j.jgg.2024.03.015

45. Yadalam PK, Arumuganainar D, Ronsivalle V, Di Blasio M, Badnjevic A, Marrapodi MM, et al. Prediction of interactomic hub genes in PBMC cells in type 2 diabetes mellitus, dyslipidemia, and periodontitis. *BMC Oral Health*. (2024) 24:385. doi: 10.1186/s12903-024-04041-y

46. Dudek G, Sakowski S, Brzezinska O, Sarnik J, Budlewski T, Dragan G, et al. Machine learning-based prediction of rheumatoid arthritis with development of ACPA autoantibodies in the presence of non-HLA genes polymorphisms. *PLoS One*. (2024) 19:e0300717. doi: 10.1371/journal.pone.0300717

47. Tao W, Concepcion AN, Vianen M, Marijnissen ACA, Lafeber FPGJ, Radstake TRDJ, et al. Multiomics and machine learning accurately predict clinical response to adalimumab and Etanercept therapy in patients with rheumatoid arthritis. *Arthritis Rheumatol*. (2021) 73:212–22. doi: 10.1002/art.41516

48. Zhou W, Zhu Y, Zhang S. Xianling Gubao capsules improve oral health, alveolar bone defects, and bone density in patients with periodontitis. *Am J Transl Res*. (2025) 17:1376–87. doi: 10.62347/GCDD8292

Glossary

WGCNA - Weighted correlation network analysis

SVM - Support Vector Machine

AUC - Area Under the Curve

PGE_2 - prostaglandin E_2

IL-1β - interleukin-1β

TNF-α - tumor necrosis factor-α; IL-8

IFN-γ - interferon-γ

FPKM - Fragments Per Kilobase per Million

DEGs - differentially expressed genes

DMGs - differentially methylated genes

FC - Fold-Change value

DC - Dendritic Cell

CLP - Common Lymphoid Progenitor

CMP - Common Myeloid Progenitor

GMP - Granulocyte-Macrophage Progenitor

HSC - Hematopoietic Stem Cell

MEP - Megakaryocyte-Erythroid Progenitor

MPP - Multipotent Progenitor

MSC - Mesenchymal Stem Cell

NKT - Natural Killer T Cell

CTD - Comparative Toxicogenomics Database

PGM - phosphoglucomutase

RF - random forest

RASSF - Ras-associated structural domain family

GEO - Gene Expression Omnibus database

frontiers | Frontiers in Medicine

TYPE Systematic Review
PUBLISHED 17 October 2025
DOI 10.3389/fmed.2025.1657241

Check for updates

OPEN ACCESS

EDITED BY
HaiHui Huang,
Shaoguan University, China

REVIEWED BY
Changmin Peng,
George Washington University, United States
Eric Chu,
EC Healthcare, Hong Kong SAR, China

*CORRESPONDENCE
Zaigao Liu
✉ 152225206@qq.com

†These authors have contributed equally to this work

RECEIVED 01 July 2025
ACCEPTED 15 September 2025
PUBLISHED 17 October 2025

CITATION
Huang FF, Liu J, Lu M, Wu Y, Zhen S, Cai Y, Liu Z, Qiu M, Xiao W, Huang Y, Liang J, Li M and Liu Z (2025) Neuroimaging evidence for central mechanisms of acupuncture in non-specific low back pain: a systematic review and meta-analysis.
Front. Med. 12:1657241.
doi: 10.3389/fmed.2025.1657241

COPYRIGHT
© 2025 Huang, Liu, Lu, Wu, Zhen, Cai, Liu, Qiu, Xiao, Huang, Liang, Li 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.

Neuroimaging evidence for central mechanisms of acupuncture in non-specific low back pain: a systematic review and meta-analysis

Frank Fan Huang[1†], Jiajun Liu[2†], Manqi Lu[3], Yixun Wu[4], Shanshan Zhen[5], Yihui Cai[1], Zhaoxue Liu[1], Mingwang Qiu[4], Wenwu Xiao[6], Yuxi Huang[7], Junquan Liang[8,9,10,11], Min Li[12] and Zaigao Liu[13]*

[1]Department of Rehabilitation Sciences, The Hong Kong Polytechnic University, Hong Kong, Hong Kong SAR, China, [2]Tianjin Hospital, Tianjin University, Tianjin, China, [3]Department of Rehabilitation Medicine, Shenzhen Hospital, Southern Medical University, Shenzhen, China, [4]The Second School of Clinical Medicine, Guangzhou University of Chinese Medicine, Guangzhou, Guangdong, China, [5]School of Exercise and Health, Shanghai University of Sport, Shanghai, China, [6]Department of Rehabilitation Medicine, Affiliated Renhe Hospital of China Three Gorges University, Yichang, China, [7]The Fourth School of Clinical Medicine, Guangzhou University of Chinese Medicine, Guangzhou, Guangdong, China, [8]Shenzhen Bao'an Chinese Medicine Hospital, The Seventh Clinical Medical School of Guangzhou University of Chinese Medicine, Shenzhen, China, [9]The Brain Cognition and Brain Disease Institute (BCBDI), Shenzhen Institute of Advanced Technology, Shenzhen, China, [10]Chinese Academy of Sciences (CAS), Shenzhen-Hong Kong Institute of Brain Science-Shenzhen Fundamental Research Institutions, Shenzhen, China, [11]University of Chinese Academy of Sciences, Beijing, China, [12]Fifth School of Clinic Medicine, Guangzhou University of Chinese Medicine, Guangzhou, Guangdong, China, [13]The Second Affiliated Hospital, School of Medicine, The Chinese University of Hong Kong, Shenzhen & Longgang District People's Hospital of Shenzhen, Shenzhen, China

Objectives: Non-specific low back pain (NSLBP) is a prevalent disorder with significant global health impacts. This systematic review and meta-analysis assessed acupuncture's clinical effectiveness for NSLBP and explored its brain mechanisms using fMRI.

Methods: A comprehensive search of multiple databases (PubMed, Embase, Cochrane Library, Web of Science, Science Direct, China National Knowledge Infrastructure, Wanfang Data, Chinese Technical Periodicals Database, and Chinese Biomedical Literature Database) was conducted from inception to July 11th, 2024. We included randomized controlled trials (RCTs) or non-RCTs resting-state functional magnetic resonance imaging to observe the effect of acupuncture on NSLBP. GingerALE 3.0.2 was used as the meta-analysis tool, and meta-analysis was performed in the Montreal Neurological Institute coordinate space.

Results: The review synthesized evidence from ten studies involving 358 participants. Subgroup analyses indicated that acupuncture significantly reduced pain scores compared to sham acupuncture in both acute NSLBP (WMD = −1.04, 95% CI: −1.72 to −0.36, $p = 0.003$) and chronic NSLBP (WMD = −0.78, 95% CI: −1.25 to −0.31, $p < 0.001$). Neuroimaging analyses revealed distinct brain activation patterns: acute NSLBP showed positive activation in the right sub-lobar insula, inferior parietal lobule, medial frontal gyrus, and cingulate gyrus, while chronic NSLBP demonstrated positive activation in bilateral sub-lobar insula and negative activation in motor and prefrontal regions.

Conclusion: Acupuncture shows significant efficacy for NSLBP, modulating pain processing through the insula and limbic system. While these results suggest therapeutic potential for both acute and chronic NSLBP, higher-quality research is needed to validate these mechanisms.
Systematic review registration: Prospero registration number: CRD42022342438, URL: https://www.crd.york.ac.uk/PROSPERO/view/CRD42022342438.

KEYWORDS

acupuncture, non-specific low back pain, meta-analysis, functional magnetic resonance imaging, systematic reviews

1 Introduction

Non-specific low back pain (NSLBP), a highly prevalent musculoskeletal disorder in adults, encompasses both nociceptive and neuropathic components that may radiate to the lower extremities, significantly impairing mobility and function (1). The classification of NSLBP falls into acute, subacute, and chronic categories (2, 3). According to the 2021 Global Burden of Disease Study, NSLBP ranks among the top 10 causes of long-term disability in 188 countries (2, 4). The global prevalence of lower back pain is estimated at 18.3%, with higher rates observed among women and in high-income countries (5). Financially, this condition imposes a heavy burden, costing the UK approximately £2.8 billion annually, Australia over $4.8 billion, and the US more than $100 billion (6).

Given its impact, effective treatments for NSLBP are critical for global health. Opioids are frequently prescribed for chronic NSLBP but raise concerns about addiction and risks (7), contributing to a drug abuse crisis and fueling demand for non-opioid alternatives (8). Increasingly, research has pointed to non-pharmacological approaches as safe and effective alternatives for managing NSLBP (9–13), and the effectiveness of acupuncture in pain relief has been demonstrated in numerous studies (14–16). It is also strongly advised to utilize acupuncture for treatment in the American College of Physicians guidelines for treating chronic NSLBP (13).

Regular MRI is used to visualize structural abnormalities such as disc herniations, spinal stenosis, or cancer. Brain imaging studies reveal stage-specific alterations in NSLBP. SPECT imaging and statistical analyses have demonstrated different alterations in brain blood flow among patients with acute and chronic NSLBP (17). In chronic cases, enhanced connectivity within the frontoparietal network (FPN), somatomotor network (SMN), and thalamus (18). This increased connectivity represents neurophysiological changes associated with the chronic phase of the condition. Given these altered connectivity patterns of different phrases of NSLBP, acupuncture has been explored as a potential neuromodulatory intervention. As for mechanism, acupuncture appears to influence several brain networks involved in pain, emotion, and memory, such as the sensorimotor network, the default mode network (DMN), and the limbic system (19). However, the exact neurophysiological mechanisms remain unclear due to acupuncture's engagement of multiple neural circuits (20), highlighting the need for further research to clarify its role across NSLBP phases.

Since the mid-1990s, functional magnetic resonance imaging (fMRI) has been used to observe the human brain's response to acupuncture stimulation (21). As an imaging method, fMRI reveals time-varying changes in brain metabolism, offering researchers precise insights into the anatomical and physiological functions associated with acupuncture. These findings suggest that acupuncture's mechanism is mediated through the central nervous system (19). Therefore, fMRI is a critical tool for investigating how acupuncture exerts its therapeutic effects at the neurophysiological level. Acupuncture's analgesic effects are mediated by neurotransmitters, signaling pathways, and immunological responses, which in turn influence neural activity in specific brain regions (22). Previous study discovered that following acupuncture therapy, neural activation increased in the sensorimotor network, periaqueductal grey, and nucleus accumbens, while the DMN showed decreased activation (23). Moreover, there were common patterns of activation in the sensorimotor cortical network and deactivation in the limbic paralimbic neocortical network after acupuncture stimulation (24). These effects were also observed in participants with NSLBP, where acupuncture improved aberrant brain structure and functional activity, primarily through the pain matrix, DMN, salience network, and descending pain modulatory system (25). In summary, acupuncture's ability to modulate brain networks and neurotransmitter activity contributes to its therapeutic effects on pain.

Several reviews have summarized the mechanisms underlying this treatment using magnetic resonance imaging to explore its effects on NSLBP (26–29). Yet, these analyses did not differentiate NSLBP by duration, limiting understanding of phase-specific analgesic mechanisms. Addressing this gap, our meta-analysis categorizes NSLBP into acute, subacute, and chronic phases to examine pain scales and brain function following acupuncture. By focusing on duration-specific cohorts, our study aims to elucidate neural substrates of acupuncture analgesia, informing clinical decisions and guiding future research directions.

Abbreviations: ACC, anterior cingulate cortex; ALE, activation likelihood estimation; ANSLBP, acute non-specific low back pain; CIs, confidence intervals; CNSLBP, chronic non-specific low back pain; DMN, default mode network; fMRI, functional magnetic resonance imaging; GRADE, Grading of Recommendations, Assessment, Development and Evaluation; MFG, medial frontal gyrus; MNI, Montreal Neurological Institute; NSLBP, nonspecific low back pain; PFC, prefrontal cortex; RCTs, randomized controlled trials; SDM, Seed-based d Mapping; WMD, weighted mean difference; FPN, frontoparietal network; SMN, somatomotor network.

2 Methods

2.1 Data and methods

The protocol of this study was registered at PROSPERO (http://www.crd.york.ac.uk/PROSPERO) (registration number: CRD42022342438).

A systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines (PRISMA guidelines) and neuroimaging guidelines for meta-analyses (30).

2.2 Literature retrieval

A systematic search strategy was conducted in PubMed, Embase, Cochrane Library, Web of Science, Science Direct, Medline, China National Knowledge Infrastructure, Wanfang Data Knowledge Service Platform, Chinese Technical Periodicals Database and Chinese Biomedical Literature Database from inception to July 11th, 2025. Additionally, forward citation tracking were identified by manually searching the included studies. The electronic search procedures are presented in Supplementary materials.

2.3 Inclusion/exclusion criteria

Studies were included based on the following criteria: (1) Randomized controlled trial (RCT) or non-RCT conducted in patients with acute, subacute or chronic NSLBP. Acute back pain is defined as lasting less than 4 weeks, subacute back pain lasts 4 to 12 weeks, and chronic back pain lasts more than 12 weeks (17); (2) fMRI study; (3) acupuncture as the intervention; (4) other therapies including conventional rehabilitation or sham acupuncture as the control group; (5) study setting in clinic, community, hospital, or laboratory; (5) presenting the results in Talairach or Montreal Neurological Institute (MNI) coordinates.

The exclusion criteria were as follows: (1) abstracts, case reports, commentaries, conference papers, cohort studies, cross-sectional studies, descriptive studies, editorials or expert opinions,or letters; (2) animal trial; (3) no extractable data available; (4) not published in English or Chinese.

2.4 Data extraction

For the data extraction, the base information of the author, country, condition, sample size of trial groups and control groups, participant characteristics, duration of NSLBP, outcomes, interventions, methodological quality assessment tool and main conclusions were extracted according to the PRISMA flowchart (31).

Firstly, the clinical outcome measures included assessing pain intensity and functional status. Pain intensity is primarily assessed using the visual analogue scale. Functional status can be assessed through self-reported questionnaires measuring disabilities for functional evaluation (e.g., Roland Disability Questionnaire for Sciatica, World Health Organization Quality of Life in the Brief Edition). Secondly, the outcome measures also included brain imaging. For brain imaging data, the brain-related data including magnetic resonance imaging model, field strength (Tesla), head coil, fMRI acquisition parameters [repetition time (TR): 2000–3,000 ms; echo time (TE): 30–40 ms; voxel size: $2.6 \times 2.6 \times 3.0\ mm^3$ to $3.4 \times 3.4 \times 4.0\ mm^3$], software used for analysis (e.g., SPM, FreeSurfer), coordinate space (MNI or Talairach), smoothing kernel (full-width at half-maximum: 5–8 mm), type I error correction, and functional imaging feature were extracted. Preprocessing steps of fMRI data in included studies consistently included: (1) motion correction; (2) slice-timing correction; (3) normalization to MNI space; (4) spatial smoothing. The coordinates and information for each study were manually extracted by two researchers (F. H. and M. Q. L.) and independently checked for accuracy by the other author (J. J. L.).

2.5 Methodological quality assessment and level of evidence

We employed Risk of Bias 22 and Risk Of Bias In Non-randomized Studies-of Interventions tools to evaluate the risk of bias in the included RCTs and non-RCTs, respectively (32, 33). For RCTs, the assessment focused on several bias sources: bias arising from the randomization process, bias due to deviations from intended interventions, bias due to missing outcome data, bias in outcome measurement, bias in the selection of reported results, and overall risk of bias. Based on these criteria, the risk of bias in RCT studies was categorized as low risk, some concerns, or high risk. In the case of non-RCTs, the assessment considered factors such as bias due to confounding, bias in participant selection, bias in the classification of interventions, bias due to deviations from intended interventions, bias due to missing data, bias in outcome measurement, bias in the selection of reported results, and overall risk of bias. According to these criteria, the risk of bias in non-RCT studies was classified as low, moderate, serious, critical, or no information.

2.6 Data analysis

Stata 12.0 software (Stata Corp, College Station, TX, USA) was used for clinical data meta-analysis. Dichotomous outcomes were reported using risk ratios with corresponding 95% confidence intervals (CIs). Continuous outcomes were presented as weighted mean differences (WMDs) with 95% CIs or standardized mean differences. A fixed-effects model was employed when the I^2 statistic was below 50% Otherwise, a random-effects model was utilized. Subgroup analysis was also conducted. And the level of evidence was used by The Grading of Recommendations, Assessment, Development and Evaluation (GRADE) approach (34).

GingerALE 3.0.2 (http://www.brainmap.org/ale/) is a tool used for neuroimaging meta-analyses, which converts all reported coordinates into MNI space via the icbm2tal transformation. Anatomical structures were identified within the software, with parameters set at $p \leq 0.001$ (cluster-level family-wise error correction = 0.001) (35–38). Mango 4.0.2 (Research Imaging Institute, UTHSCSA) was used for visualization, mapping the three-dimensional ALE results onto the MNI standard template to facilitate precise localization of brain regions.

2.7 Activation likelihood estimation procedure

In ALE analysis, activation hotspots found in existing research were viewed as probability patterns centred on the reported coordinates. For each voxel in a standard space,

activation probabilities were determined to create ALE maps that focus on particular contrasts. To assess the trustworthiness of these ALE maps, null distributions were formed by examining how ALE values were distributed across separate studies (36). This approach was somewhat like performing permutation tests on individual voxels from different experiments. The influence of each study in the analysis was adjusted based on its sample size, and each study is considered to contribute to random effects (35).

2.8 Calculation of frequency about brain regions modulated by acupuncture

To summarize and visualize the frequency of brain regions modulated by acupuncture in acute and chronic LBP, we utilized Excel for data analysis. The frequencies of involvement for various brain regions were calculated and plotted using Excel's graphing capabilities. This allowed us to effectively illustrate the distribution of modulated regions in both acute and chronic LBP.

2.9 Sensitivity analysis

We performed sensitivity analyses to assess the robustness of ALE meta-analysis results based on previous article (37). Studies with a total sample size of <20 were excluded to address potential small sample bias.

3 Results

3.1 Study search results

A total of 1,020 articles were identified through PubMed, Embase, Cochrane Library, Web of Science, Science Direct, China National Knowledge Infrastructure, Wanfang Data Knowledge Service Platform, Chinese Technical Periodicals database and Chinese Biomedical Literature Database. After removing duplicates, trials for which no full-text was available, and screening titles and abstracted, a total of ten studies were included for further evaluation (8, 39–47).

3.2 Characteristics of the included studies

A total of 358 participants were involved in ten articles (Cohen's kappa = 0.85) (8, 39–47). Table 1 shows the characteristics of the included studies. Among these ten articles, three are about acute non-specific low back pain (ANSLBP) (39, 40, 42) and seven are about chronic non-specific low back pain (CNSLBP) (8, 41, 43–47). Among the ten studies included, eight (8, 37, 39, 40, 42, 43, 45, 46) were divided into an acupuncture group and a control group, of which three (39, 40, 42) analyzed the effect of acupuncture on ANSLBP, and the other five (8, 43, 45–47) analyzed the effect of acupuncture on CNSLBP. No subacute NSLBP articles were included. The other two studies (41, 44) had no control group. The selection process is shown in Figure 1.

3.3 Quality assessment of the included trials

In eight RCTs (8, 39, 40, 42, 43, 45–47), one study was rated as "low" overall risks of bias (8), while seven studies were rated as "some concerns" of overall risks of bias due to concerns about the randomization process (39, 40, 42, 43, 45–47). In two non-RCTs (41, 44), all two studies were rated as "low" overall risks of bias (Supplementary Tables 1, 2). According to the GRADE approach, the quality of evidence and the strength of recommendations were rated as "very low" (Supplementary Table 3).

3.4 Meta-analysis results of pain-related scales

Based on the pooled results from five RCTs (Supplementary Figure 1) (8, 43, 45–47), the acupuncture group showed significantly lower pain-related scores of VAS compared with the sham acupuncture group, as illustrated in Figure 2 (5 trials: WMD = −0.78, 95% CI: −1.25 to −0.31, $p < 0.001$), with no heterogeneity ($I^2 = 0\%$, $p = 0.825$) For subgroup analysis, one RCT demonstrated that acupuncture was significantly more effective than sham acupuncture for treating ANSLBP (WMD = −1.04, 95% CI: −1.72 to −0.36, $p = 0.003$) (42). However, four RCTs on CNSLBP showed no significant difference between real and sham acupuncture (Figure 1) (WMD = −0.53, 95% CI: −1.19 to 0.13, $p = 0.113$) (8, 43, 45–47), with no heterogeneity ($I^2 = 0\%$, $p = 0.938$).

3.4.1 Neuroimaging findings after acupuncture for ANSLBP

Three studies utilized acupuncture for the treatment of ANSLBP (39, 40, 42). Following the ALE meta-analysis of these articles, the results identified four clusters of positive activation and seven clusters of negative activation (Figure 3).

Four clusters of positive activation were identified. The first cluster was located in the right cerebrum, specifically in the sub-lobar insula (Brodmann area 13), centered at coordinates $x = 50$, $y = 6$, $z = 12$ (ALE = 0.0022; $p < 0.001$; $Z = 5.20$). The second cluster was found in the right cerebrum, in the inferior parietal lobule (Brodmann area 40), centered at $x = 62$, $y = -26$, $z = 34$ (ALE = 0.0023; $p < 0.001$; $Z = 5.39$). The third cluster was situated in the right cerebrum, in the medial frontal gyrus (MFG) (Brodmann area 6), centered at $x = 12$, $y = 0$, $z = 60$ (ALE = 0.0019; $p < 0.001$; $Z = 4.85$). The final cluster was located in the right cerebrum, in the cingulate gyrus (Brodmann area 31), centered at $x = 18$, $y = -24$, $z = 39$ (ALE = 0.0019; $p < 0.001$; $Z = 4.84$) (Figure 3A; Supplementary Table 4).

Seven clusters of negative activation were identified. The first cluster was located in the left cerebrum, specifically in the sub-lobar insula (Brodmann area 13), centered at coordinates $x = -41$, $y = -13$, $z = 15$ (ALE = 0.0027; $p < 0.001$; $Z = 6.36$). The second cluster was found in the left cerebrum, in the cingulate gyrus (Brodmann area 32), centered at $x = 0$, $y = 33$, $z = 21$ (ALE = 0.0028; $p < 0.001$; $Z = 7.03$). The third cluster was situated in the left cerebrum, in the pulvinar of the thalamus, centered at $x = -4$, $y = -30$, $z = -2$ (ALE = 0.0019; $p < 0.001$; $Z = 4.58$). The fourth cluster was located in the right cerebrum, in the parahippocampal gyrus (Brodmann area 35), centered at $x = 24$, $y = -27$, $z = -18$ (ALE = 0.0019; $p < 0.001$;

TABLE 1 Characteristics of the included studies.

Study	Sample number	Phrase	Location	Design	Treatment duration	Age	Treatment group	Control group	Manipulation modality	Imaging type	MRI parameter	Clinical variables	Adverse events	Outcomes
Ziping (40)	Sum = 15	acute	hospital	RCT	16 min	25.7 ± 2.3	acupuncture	sham acupuncture	MA	RS-fMRI	3 T, MNI	1. Vas score 2. ASS	None	Brain activation
Shi et al. (42)	Sum = 28	acute	NA	RCT	36 min	22–30	acupuncture	sham acupuncture	EA	RS-fMRI	3 T, MNI	1. Vas score	None	ReHo, brain activation
Makary, 2018 (46)	TG = 28 CG = 19	chronic	laboratory	RCT	25 min	38.4 ± 12.7	acupuncture	sham acupuncture	MA	RS-fMRI	3 T, MNI	1. Vas score	None	ROI
Lee, 2019 (43)	TG = 25 CG = 18	chronic	NA	RCT	25 min	38.4 ± 12.7	acupuncture	sham acupuncture	MA	RS-fMRI	3 T, MNI	1. Vas score	None	ROI, FC
Tu, 2019 (45)	TG = 24 CG = 26	chronic	NA	RCT	25 min, 1 or 2 times/week, 2 weeks	26–54	acupuncture	sham acupuncture	MA	TS-fMRI	3 T, MNI	1. Vas score 2. BDI 3. ERS	None	rsFC
Yu, 2020 (8)	TG = 24 CG = 26	chronic	hospital	RCT	25 min, 1 or 2 times/week, 2 weeks	18–60	acupuncture	sham acupuncture	MA	TS-fMRI	3 T, MNI	1. Vas score 2. BDI 3. Bothersomeness scale	None	rsFC
Kim, 2020 (57)	TG = 55 (TG1 = 18, TG2 = 18, TG3 = 19) CG = 23	chronic	hospital	RCT	20 min, 1 or 2 times/week, 4 weeks	41.2 ± 12.0	acupuncture	None	MA	TS-fMRI	3 T, MNI	1. Vas score 2.2PDT	None	ROI
Xiang, 2021 (11)	Sum = 19	chronic	laboratory	Non-RCT	15 min	46.61 ± 7.35	ankle acupuncture	sham acupuncture	MA	RS-fMRI	3 T, MNI	1. Vas score	None	ALFF

BDI, Beck Depression Inventory; TG, treatment group; CG, control group; EA, electroacupuncture; MA, manual acupuncture; ERS, expectations for relief scale; JOA, Japanese Orthopaedic Association scores; NA, not available; Sham, sham acupuncture; VAS, visual analogue scale; Verum, verum acupuncture; 2 PDT, two-point discrimination threshold; ASS, Acupuncture Sensation Scale; rs-FC, resting-state functional connectivity/FC; ALFF, Amplitude of Low-Frequency Fluctuations; ICC, Intrinsic Connectivity Contrast; Reho, Regional Homogeneity; ROI, Regions of Interest.

10.3389/fmed.2025.1657241

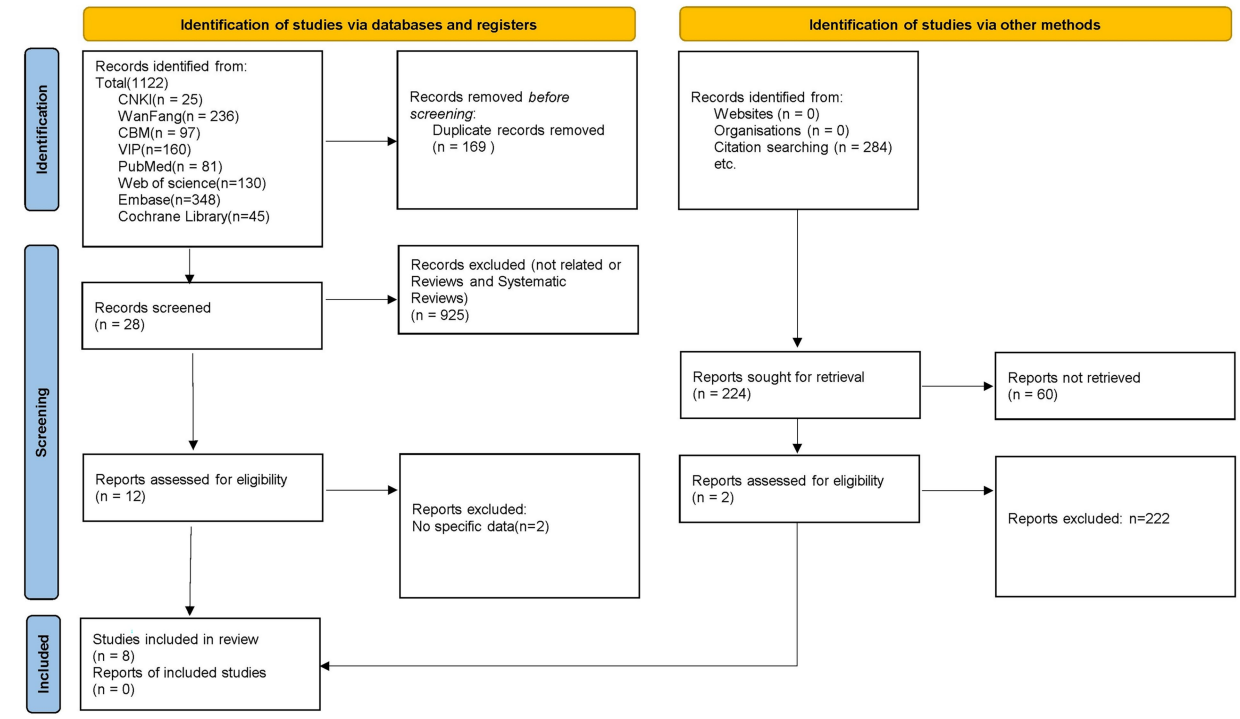


FIGURE 1
Flowchart of the study selection process.

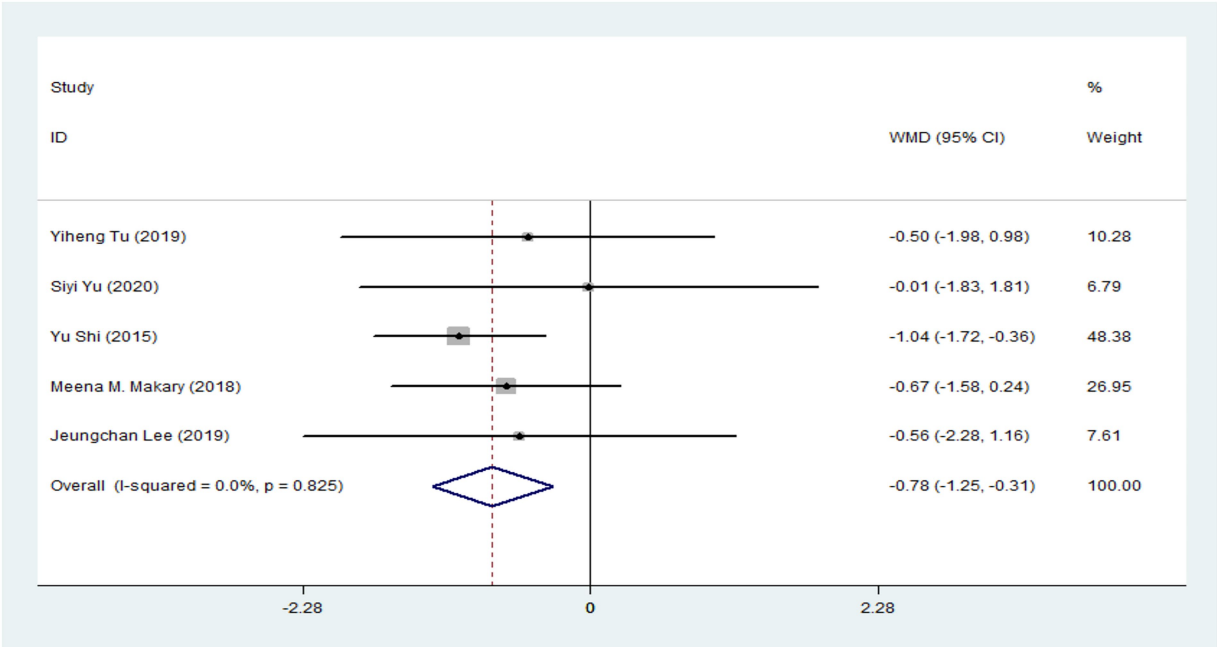


FIGURE 2
Forest plot of the sum total.

$Z = 4.61$). The fifth cluster was found in the right cerebrum, in the MFG (Brodmann area 8), centered at $x = 14$, $y = 33$, $z = 44$ (ALE = 0.0019; $p < 0.001$; $Z = 4.57$). The sixth cluster was located in the right cerebrum, in the angular gyrus (Brodmann area 39), centered at $x = 54$, $y = -60$, $z = 39$ (ALE = 0.0019; $p < 0.001$; $Z = 4.61$). The final cluster was situated in the left cerebrum, in the superior frontal gyrus (Brodmann area 6), centered at $x = -14$, $y = 34$, $z = 52$ (ALE = 0.0019; $p < 0.001$; $Z = 4.57$) (Figure 3B; Supplementary Table 5).

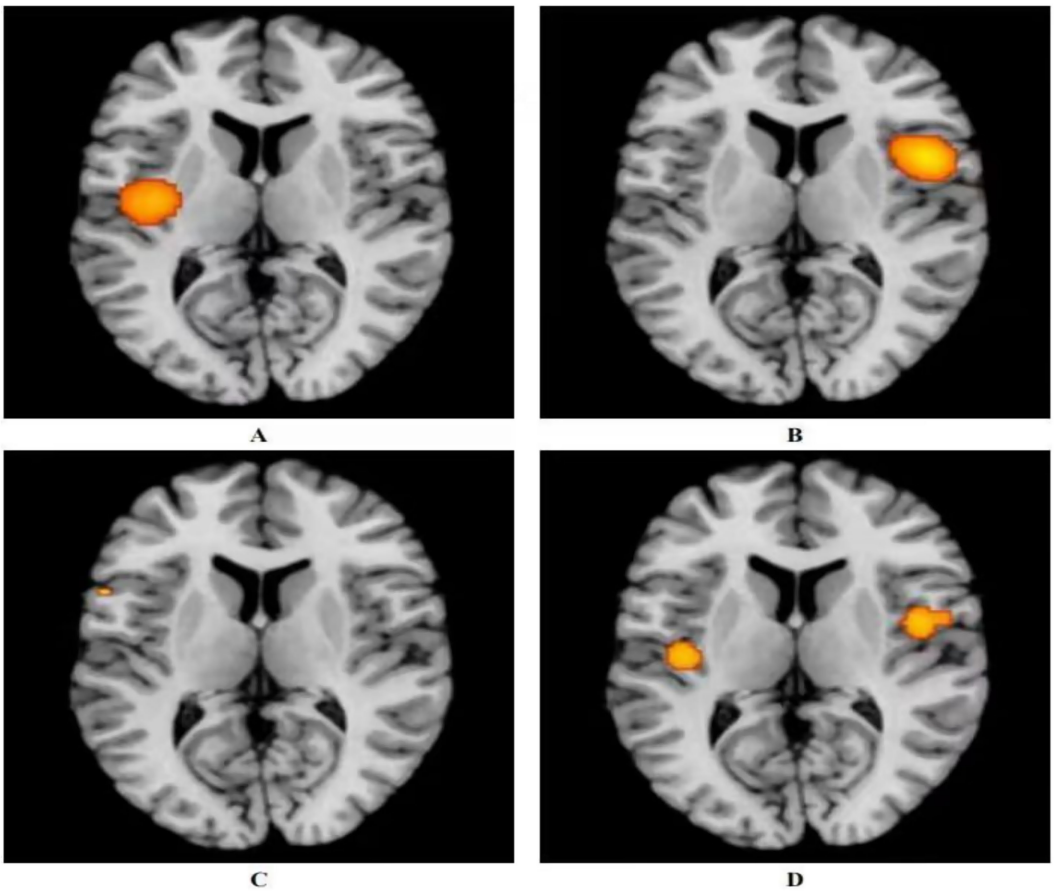


FIGURE 3
Activation of fMRI signals in cortical and subcortical structures in the acupuncture group. **(A)** Positive activation of brain regions after acupuncture with ANSLBP; **(B)** Negative activation of brain regions after acupuncture with ANSLBP; **(C)** Positive activation of brain regions after acupuncture with CNSLBP; **(D)** Negative activation of brain regions after acupuncture with CNSLBP.

3.4.2 Neuroimaging findings after acupuncture for CNSLBP

Seven studies utilized acupuncture as a treatment for CNSLBP (8, 41, 43–47). Analyzing the related articles revealed two clusters of positive activation and two clusters of negative activation (Figure 3).

The two positive activation clusters identified were as follows: one was located in the right cerebrum, specifically in the sub-lobar insula (Brodmann area 13), centered at coordinates $x = 46$, $y = -2$, $z = 2$ (ALE = 0.0019; $p < 0.001$; $Z = 4.65$). The other was found in the left cerebrum, also in the sub-lobar insula (Brodmann area 13), centered at $x = -42$, $y = -16$, $z = 2$ (ALE = 0.0015; $p < 0.001$; $Z = 3.93$) (Figure 3C; Supplementary Table 6).

The two negative activation clusters were located as follows: one was situated in the left cerebrum, in the precentral gyrus (Brodmann area 44), centered at $x = -56$, $y = 12$, $z = 6$ (ALE = 0.00095; $p < 0.001$; $Z = 3.41$). The other was located in the right cerebrum, in the middle frontal gyrus (Brodmann area 8), centered at $x = 33$, $y = 40$, $z = 42$ (ALE = 0.00095; $p < 0.001$; $Z = 3.41$) (Figure 3D; Supplementary Table 7).

3.4.3 Neuroimaging results after the control group's treatment on ANSLBP

The results showed that the control group's treatment for ANSLBP activated over eight clusters, predominantly located in the sub-lobar regions of the left cerebrum, including the insula, lentiform nucleus, thalamus, caudate, and hippocampus. Only the caudate in the right cerebrum showed activation. It was clear that most of the activated clusters were in the sub-lobar area. Additionally, the ALE values ranged from 0.0009 to 0.001, with the insula in the left cerebrum displaying the highest activation, specifically centered at $x = -40$, $y = 6$, $z = 18$ (ALE = 0.0014, $p < 0.001$, $Z = 4.05$) (Figure 4A). In summary, the results indicate that significant brain activation occurs primarily in the sub-lobar region, with the left insula showing the highest activation, which may be central to the neurophysiological response to ANSLBP treatment.

3.4.4 Neuroimaging results after the control group's treatment on CNSLBP

Long-term control group's regulates wider areas, such as the limbic lobe (Brodmann area 40; peak MNI coordinates: −50, −28, 52; peak SDM-Z (Seed-based d Mapping): 3.597; $p < 0.001$), the parietal

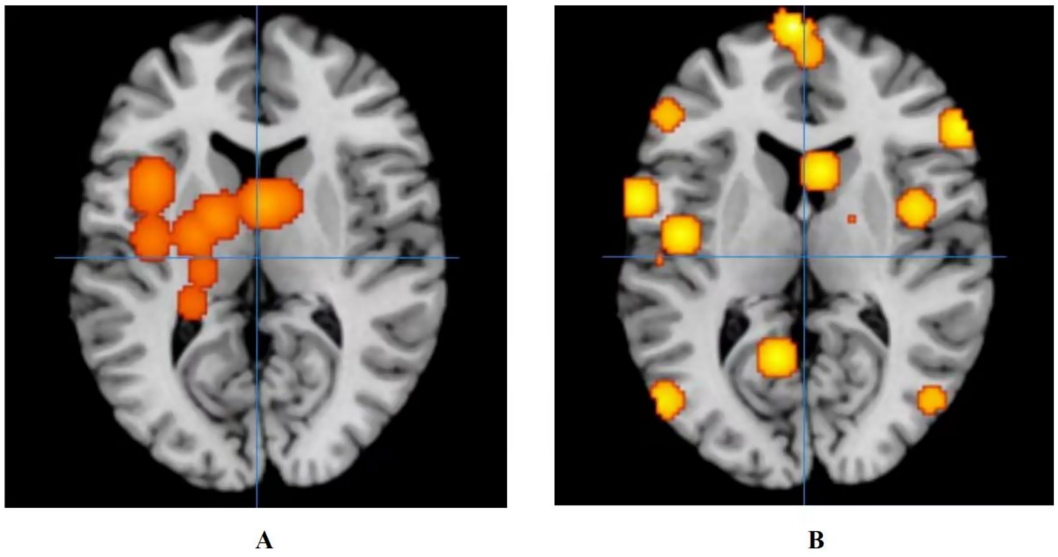


FIGURE 4
Activation of fMRI signals in cortical and subcortical structures in the control group. **(A)** Activation of brain regions in patients with ANSLBP after receiving treatment in control group; **(B)** Activation of brain regions in patients with CNSLBP after receiving treatment in control group.

lobe (Brodmann area 36; peak MNI coordinates: −50, −34, −16; peak SDM-Z: 3.597; $p < 0.001$), the frontal lobe (Brodmann area 6; peak MNI coordinates: 44, 6, 30; peak SDM-Z: 3.597; $p < 0.001$), the temporal lobe (Brodmann area 22; peak MNI coordinates: −66, −42, 18; peak SDM-Z: 3.597; $p < 0.001$), the sub-lobar (Caudate Body; peak MNI coordinates: 18, −4, 16; peak SDM-Z: 3.597; $p < 0.001$), and the occipital lobe (Brodmann area 37; peak MNI coordinates: −52, −72, 4; peak SDM-Z: 3.597; $p < 0.001$) (Figure 4B).

3.5 Sensitivity analysis

Two studies with small sample sizes were excluded: Ziping (40; total $n = 15$) and Xiang et al. (11) (total $n = 19$). After exclusion, only one study on ANSLBP remained, while all CNSLBP studies met the sample size criterion ($n \geq 20$). Consistent with the initial analysis, the key neuroimaging findings remained unchanged. Acupuncture treatment on ANSLBP showed positive activation in the right sub-lobar insula, inferior parietal lobule, medial frontal gyrus, and cingulate gyrus; CNSLBP demonstrated positive activation in the bilateral sub-lobar insula and negative activation in the motor and prefrontal regions. Due to low heterogeneity ($I^2 = 0\%$) and consistent study inclusion, no re-calculation of ALE statistics was required, confirming the stability of the primary results.

3.6 Frequency of brain regions modulated by acupuncture in acute and chronic LBP

By combining these findings, we summarized and visualized the frequency of brain regions modulated by acupuncture in acute and chronic LBP (Supplementary Figure 2). As shown in the Supplementary Figure 2, in ANSLBP, the most frequently involved regions were the insula and lentiform nucleus (approximately 75%), followed by the thalamus, caudate, and hippocampus (around 25%). In CNSLBP, the most frequently involved regions were the middle frontal gyrus (>50%) and precentral gyrus (approximately 45%), followed by the parahippocampal gyrus, anterior cingulate cortex, superior parietal lobule, and inferior parietal lobule (about 20–30%).

3.7 Adverse events

As shown in Table 1, none of the RCTs or non-RCTs reported adverse events.

4 Discussion

This study compares brain changes following acupuncture treatment of acute and chronic NSLBP using pain-related scales and resting-state fMRI. The research results indicated that acupuncture has demonstrable clinical efficacy for treating NSLBP. Through a meta-analysis of all eligible articles, three studies used acupuncture for ANSLBP (39, 40, 42), with brain activation mainly in the bilateral limbic lobe and right inferior lobe. In the seven studies of acupuncture treatment of CNSLBP (8, 41, 43–47), we identified four clusters of activation, including the sub-lobar insula, precentral gyrus on the left side, and the sub-lobar insula and middle frontal gyrus on the right side.

4.1 Pain-related outcomes analysis of acupuncture for NSLBP

Acupuncture is currently recognized as an effective treatment for spinal-related diseases. In recent years, RCTs have demonstrated their role in treating degenerative diseases, chronic pain, and acute pain (48,

49). Chen et al. found that patients with chronic low back pain exhibit widespread alterations in brain regions related to pain perception and modulation, including the left inferior temporal gyrus, bilateral postcentral gyrus, superior and middle frontal gyri, thalamus, and occipital cortex. Notably, acupuncture appears to modulate functional activity in several of these pathological areas (50). Specifically, increased cerebral blood flow has been observed in the right postcentral gyrus and superior parietal lobule (regions implicated in somatosensory processing and sensorimotor integration), while in the bilateral occipital cortex and posterior cingulate gyrus is reduced (51).

In addition to targeted brain modulation by acupuncture, it is useful to compare its effects with those of other non-acupuncture treatments for NSLBP. Acupuncture has been shown to restore altered DMN connectivity, particularly in the dorsolateral and medial prefrontal cortices, anterior cingulate, and precuneus, with these changes correlating with pain relief (52). Similarly, physical or manual therapies, such as spinal manipulative therapy (SMT), modulate DMN regions including the right parahippocampal gyrus, posterior cingulate cortex, and precuneus, indicating altered intrinsic connectivity related to pain processing (53). Cognitive Behavioral Therapy (CBT) engages cognitive control and emotional regulation networks, with magnetoencephalography studies showing normalization of activity in the right inferior frontal gyrus and dorsolateral prefrontal cortex, correlating with pain reduction (54). Structural MRI further reveals increased gray matter in the dorsolateral prefrontal and posterior parietal cortices after CBT, associated with decreased catastrophic thinking (55, 56).

However, among the articles we included, only three addressed acute pain. Clinically, acupuncture is sometimes used for acute low back pain in emergency settings of traditional Chinese medicine clinics, making research challenging due to low follow-up rates. Conversely, patient compliance is higher for chronic back pain, resulting in more reliable therapeutic outcomes in the included articles. Although our meta-analysis did not demonstrate positive results without grouping, this might be due to the efficacy of the control group (i.e., open-label studies, lack of blinding). Such variability in control group selection may affect meta-analysis results. Nonetheless, the treatment effectiveness of acupuncture for chronic NSLBP remains significantly different when compared with healthy controls in our study, confirming its clinical relevance and ongoing research importance in traditional Chinese medicine.

4.2 Neuroimaging analysis of acupuncture for NSLBP

According to our findings, both acute and CNSLBP activate the right insula following acupuncture, a region crucial for integrating sensory processing and cognitive regulatory systems (45). Activation of the insula observed in the acupuncture group was accompanied by significant reductions in VAS scores, suggesting that modulation of this key pain-processing region may underlie acupuncture's clinically meaningful analgesic effects. Research suggests that the anterior insula plays a key role in the salience network, responsible for identifying and filtering salient stimuli, particularly during exposure to unpleasant stimuli (57). Acupuncture has been shown to reduce cross-network functional connectivity between the insula and the DMN, and this reduction correlates with the degree of clinical pain alleviation (47). These findings suggest that right insula activation is critical to acupuncture's analgesic effects for both acute and chronic NSLBP.

Beyond the insula, the limbic lobe, located at the cerebral cortex's periphery, also plays a significant role in pain processing (58). Regions such as the amygdala, orbitofrontal cortex, hippocampus, and cingulate cortex form part of this network (59). The anterior cingulate cortex (ACC), in particular, is involved in emotion and behavior regulation (60). Acupuncture somatosensory afference can transmit tactile information from the spinal cord to the thalamus, periaqueductal grey, and reticular formation, subsequently affecting the ACC, insula, and sensory cortices (61). Activation in the ACC, especially in its dorsal sub-region, has been linked to acute pain stimulation, suggesting that the ACC's activation in this study may correspond to acupuncture's pain-relieving effects. (dACC, BA 24) (62). Therefore, the insula, ACC, and other limbic structures appear to mediate acupuncture's analgesic effects in both acute and chronic NSLBP.

We also summarized and visualized the frequency of brain regions modulated by acupuncture in acute and chronic LBP. In ANSLBP, the insula and lentiform nucleu, followed by the thalamus, caudate, and hippocampus areas are associated with pain perception, emotional processing, and pain-related memory (63, 64). In CNSLBP, the middle frontal gyrus, precentral gyrus, followed by the parahippocampal gyrus, anterior cingulate cortex, superior parietal lobule, and inferior parietal lobule are more closely related to motor planning (65), execution (66), emotional regulation, attention control, and the persistence of chronic pain. Therefore, acute low back pain is more associated with nociceptive processing and emotion/memory circuits (insula–basal ganglia–limbic system), whereas chronic low back pain is more related to higher-order cognitive and motor control networks (frontal–parietal).

4.3 Analysis of the current neuroimaging results in the control group

The activated brain regions in the acupuncture group were primarily located in the bilateral limbic lobe and right inferior lobe, while the control group for ANSLBP primarily exhibited changes in the limbic system, basal ganglia, and thalamus (39, 40, 42). In contrast, the control group for CNSLBP showed a wider range of activation, including the frontal, temporal, sub-lobar, and occipital lobes (8, 41, 43–46). The lack of blinding in the control group, combined with open-label placebos, likely amplified this reward effect and further alleviated pain (67). These regions, particularly in the somatosensory cortex and pain conduction system, are crucial components of the central nervous system that regulate pain (68–70). Interestingly, despite the lack of a correlation analysis between ANSLBP and CNSLBP after sham acupuncture, there appears to be a similar modulation pattern in the limbic system across both conditions. This observation suggests that the limbic system may play a significant role in alleviating NSLBP. Comprising cortical and subcortical structures such as the prefrontal cortex (PFC), cingulate gyrus, hippocampus, and amygdala, the limbic system integrates sensory input from the environment to regulate emotional, autonomic, motor, and cognitive responses essential for survival (71–74). Previous research

highlights the involvement of the reward system in acupuncture's effects (75–77), particularly the PFC's role in self-regulation and pain relief (78, 79). Taken together, the sensory stimuli received by the control group may convey positive reinforcement through the limbic system, particularly the PFC, contributing to pain reduction.

Our findings and previous literature indicate that sham acupuncture often induces neural activations in brain regions associated with attention, expectation, and pain modulation, reflecting placebo-related and nonspecific neural responses rather than acupuncture-specific effects (25, 80). This overlap complicates the interpretation of neuroimaging results and underscores the necessity of cautious attribution of brain activity solely to acupuncture. Future studies should further delineate these mechanisms to improve the specificity of acupuncture-related neurobiological findings.

4.4 Advantages and limitations

Neuroimaging results on the effects of acupuncture on NSLBP have been elusive, particularly due to the varied causes of the condition and differences in pain types (81). The duration of NSLBP may also significantly affect experimental outcomes, adding complexity to studies. Besides, Variability in acupuncture protocols (e.g., needle retention: 15–36 min; point selection) may confound neuroimaging effects. Future trials should adhere to the standards for reporting interventions in clinical trials of acupuncture guidelines. Additionally, the definition of chronic pain remains unclear (82, 83), which may lead to inaccuracies in clinical diagnosis and complicate research on brain function changes associated with chronic pain. One limitation of earlier studies is their failure to differentiate between acute and chronic NSLBP (84). Furthermore, pooling data from studies with different designs in meta-analyses can introduce heterogeneity and bias (85). In addition, the limited number of included studies in certain subgroup analyses (e.g., only three ANSLBP studies) restricts the statistical power of our findings, which should be considered when interpreting the results. Accordingly, further research with larger sample sizes is needed to yield more robust evidence. Besides, most included studies were rated as "very low" quality according to the GRADE approach, which weakens the strength of our conclusions. Future fMRI research on acupuncture should focus on methodological enhancements, such as rigorous randomization, appropriate blinding, and adequate sample size calculation, to improve evidence reliability. Despite the low quality of the study design, all MRI scans were conducted using 3 T machines, guaranteeing reliable imaging findings. Our study, however, addresses these limitations by distinguishing between acute and chronic NSLBP, allowing for a clearer comparison of acupuncture's effects on brain function changes. This distinction helps resolve inconsistencies in prior research. Our pooled results offer a comprehensive overview of the post-acupuncture effects on clinical outcomes and brain activation in patients with NSLBP, providing valuable insights for both clinicians and researchers (85).

5 Conclusion

Acupuncture has shown considerable clinical efficacy in alleviating pain for patients with NSLBP, with key brain regions such as the sub-lobar insula and medial frontal gyrus playing a crucial role in the analgesic mechanism for both acute and chronic conditions. In our study on acupuncture treatment for ANSLBP, we identified four clusters of positive activation (right sub-lobar insula, inferior parietal lobule, MFG, and cingulate gyrus) and seven clusters of negative activation (left sub-lobar insula, cingulate gyrus, pulvinar of the thalamus, superior frontal gyrus, right parahippocampal gyrus, MFG, and angular gyrus). In contrast, during our research on CNSLBP, we discovered two clusters of positive activation (right and left sub-lobar insula) and two clusters of negative activation (left precentral gyrus and right MFG). Subgroup analyses revealed different neuroimaging outcomes based on duration. Despite these findings, the quality of evidence and strength of recommendations were rated "very low." by the GRADE approach, highlighting the need for methodological improvements in fMRI studies on acupuncture for NSLBP.

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

FH: Writing – review & editing, Software, Writing – original draft, Investigation, Conceptualization. JiL: Project administration, Writing – original draft, Conceptualization, Writing – review & editing. MLu: Data curation, Writing – review & editing, Investigation. YW: Data curation, Methodology, Investigation, Writing – original draft, Formal analysis. SZ: Software, Writing – review & editing, Methodology. YC: Writing – review & editing, Resources. ZhL: Formal analysis, Writing – review & editing, Validation. MQ: Writing – review & editing, Visualization. WX: Writing – review & editing, Visualization. YH: Writing – review & editing, Validation. JuL: Methodology, Writing – review & editing. MLi: Funding acquisition, Writing – review & editing, Supervision. ZaL: Funding acquisition, Writing – review & editing, Validation.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the National Natural Science Foundation of China (NO. 82305372); Shenzhen Natural Science Foundation General Program for Basic Research (NO. JCYJ20230807115516032); 2024 District Public Hospital High-Quality Development Project General Program (NO. YNXM2024045).

Acknowledgments

The authors would like to thank the faculty librarian for developing the search strategy with the team.

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.

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

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

Publisher's note

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/fmed.2025.1657241/full#supplementary-material

References

1. Knezevic NN, Candido KD, Vlaeyen JWS, Van Zundert J, Cohen SP. Low back pain. *Lancet*. (2021) 398:78–92. doi: 10.1016/S0140-6736(21)00733-9

2. Theo V, Ryan MB, Brad B, Amelia BV, Stan B, Ian B, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the global burden of disease study 2013. *Lancet*. (2015) 386:743–800. doi: 10.1016/S0140-6736(15)60692-4

3. Jones GT, Macfarlane GJ. Epidemiology of low back pain in children and adolescents. *Arch Dis Child*. (2005) 90:312–6. doi: 10.1136/adc.2004.056812

4. Manuela LF, Katie de L, Lydia MH, Jaimie DS, Garland TC, Marita C, et al. Global, regional, and national burden of low back pain, 1990-2020, its attributable risk factors, and projections to 2050: a systematic analysis of the global burden of disease study 2021. *Lancet Rheumatol*. (2023) 5:e316-e329. doi: 10.1016/S2665-9913(23)00098-X

5. Maher C, Underwood M, Buchbinder R. Non-specific low back pain. *Lancet*. (2017) 389:736–47. doi: 10.1016/s0140-6736(16)30970-9

6. Hong J, Reed C, Novick D, Happich M. Costs associated with treatment of chronic low back pain: an analysis of the UK general practice research database. *Spine*. (2013) 38:75–82. doi: 10.1097/BRS.0b013e318276450f

7. Erin EK, Amy G, Sean N, Agnes CJ, Beth D, Elizabeth SG, et al. Effect of opioid vs nonopioid medications on pain-related function in patients with chronic back pain or hip or knee osteoarthritis pain: the SPACE randomized clinical trial. *JAMA*. (2018) 319:872–82. doi: 10.1001/jama.2018.0899

8. Yu S, Ortiz A, Gollub RL, Wilson G, Gerber J, Park J, et al. Acupuncture treatment modulates the connectivity of key regions of the descending pain modulation and reward Systems in Patients with chronic low Back pain. *J Clin Med*. (2020) 9:1719. doi: 10.3390/jcm9061719

9. Furlan AD, van Tulder MW, Cherkin DC, Tsukayama H, Lao L, Koes BW, et al. Acupuncture and dry-needling for low back pain. *Cochrane Database Syst Rev*. (2005):CD001351. doi: 10.1002/14651858.CD001351.pub2

10. Berman BM, Langevin HM, Witt CM, Dubner R. Acupuncture for chronic low back pain. *N Engl J Med*. (2010) 363:454–61. doi: 10.1056/NEJMct0806114

11. Xiang A, Chen M, Qin C, Rong J, Wang C, Shen X, et al. Frequency-specific blood oxygen level dependent oscillations associated with pain relief from ankle acupuncture in patients with chronic low Back pain. *Front Neurosci*. (2021) 15:786490. doi: 10.3389/fnins.2021.786490

12. Wang Y, Xu J, Zhang Q, Zhang Q, Yang Y, Wei W, et al. Immediate analgesic effect of acupuncture in patients with primary dysmenorrhea: a fMRI study. *Front Neurosci*. (2021) 15:647667. doi: 10.3389/fnins.2021.647667

13. Qaseem A, Wilt TJ, McLean RM, Forciea MA, Denberg TD, Barry MJ, et al. Noninvasive treatments for acute, subacute, and chronic low back pain: a clinical practice guideline from the American College of Physicians. *Ann Intern Med*. (2017) 166:514–30. doi: 10.7326/M16-2367

14. Luo W, Zhang Y, Yan Z, Liu X, Hou X, Chen W, et al. The instant effects of continuous transcutaneous auricular Vagus nerve stimulation at Acupoints on the functional connectivity of amygdala in migraine without Aura: a preliminary study. *Neural Plast*. (2020) 2020:1. doi: 10.1155/2020/8870589

15. Zhang R, Lao L, Ren K, Berman BM. Mechanisms of acupuncture-electroacupuncture on persistent pain. *Anesthesiology*. (2014) 120:482–503. doi: 10.1097/aln.0000000000000101

16. Chou R, Deyo R, Friedly J, Skelly A, Hashimoto R, Weimer M, et al. Nonpharmacologic therapies for low Back pain: a systematic review for an American College of Physicians Clinical Practice Guideline. *Ann Intern Med*. (2017) 166:493–505. doi: 10.7326/M16-2459

17. Nakamura Y, Nojiri K, Yoshihara H, Takahata T, Honda-Takahashi K, Kubo S, et al. Significant differences of brain blood flow in patients with chronic low back pain and acute low back pain detected by brain SPECT. *J Orthop Sci*. (2014) 19:384–9. doi: 10.1007/s00776-014-0534-2

18. Zhu K, Chang J, Zhang S, Li Y, Zuo J, Ni H, et al. The enhanced connectivity between the frontoparietal, somatomotor network and thalamus as the most significant network changes of chronic low back pain. *NeuroImage*. (2024) 290:120558. doi: 10.1016/j.neuroimage.2024.120558

19. Zhang J, Zhang Y, Hu L, Huang X, Liu Y, Li J, et al. Global trends and performances of magnetic resonance imaging studies on acupuncture: a bibliometric analysis. *Front Neurosci*. (2020) 14:620555. doi: 10.3389/fnins.2020.620555

20. Ji Z, Ning Z. Discussion on the mechanism of acupuncture analgesia. *Chinese Acupuncture*. (2007) 27:72–5. doi: 10.13703/j.0255-2930.2007.01.032

21. Cho ZH, Chung SC, Jones JP, Park JB, Park HJ, Lee HJ, et al. New findings of the correlation between acupoints and corresponding brain cortices using functional MRI. *Proc Natl Acad Sci USA*. (1998) 95:2670–3. doi: 10.1073/pnas.95.5.2670

22. Huang L, Xu G, He J, Tian H, Zhou Z, Huang F, et al. Bibliometric analysis of functional magnetic resonance imaging studies on acupuncture analgesia over the past 20 years. *J Pain Res*. (2021) 14:3773–89. doi: 10.2147/JPR.S340961

23. Makary MM, Lee J, Lee J-H, Lee E, Shin J-Y, Napadow V, et al. Dissociation of somatosensory needling and needling credibility of the acupuncture effect on low back pain: FMRI study. *Integr Med Res*. (2015) 4:1. doi: 10.1016/j.imr.2015.04.137

24. Chae Y, Chang DS, Lee SH, Jung WM, Lee IS, Jackson S, et al. Inserting needles into the body: a meta-analysis of brain activity associated with acupuncture needle stimulation. *J Pain*. (2013) 14:215–22. doi: 10.1016/j.jpain.2012.11.011

25. Wen Q, Ma P, Dong X, Sun R, Lan L, Yin T, et al. Neuroimaging studies of acupuncture on low Back pain: a systematic review. *Front Neurosci*. (2021) 15:730322. doi: 10.3389/fnins.2021.730322

26. Cai RL, Shen GM, Wang H, Guan YY. Brain functional connectivity network studies of acupuncture: a systematic review on resting-state fMRI. *J Integr Med*. (2018) 16:26–33. doi: 10.1016/j.joim.2017.12.002

27. Huang W, Pach D, Napadow V, Park K, Long X, Neumann J, et al. Characterizing acupuncture stimuli using brain imaging with FMRI--a systematic review and meta-analysis of the literature. *PLoS One*. (2012) 7:e32960. doi: 10.1371/journal.pone.0032960

28. He T, Zhu W, Du SQ, Yang JW, Li F, Yang BF, et al. Neural mechanisms of acupuncture as revealed by fMRI studies. *Auton Neurosci*. (2015) 190:1–9. doi: 10.1016/j.autneu.2015.03.006

29. Scheffold BE, Hsieh CL, Litscher G. Neuroimaging and Neuromonitoring effects of electro and manual acupuncture on the central nervous system: a literature review and analysis. *Evid Based Complement Alternat Med*. (2015) 2015:641742. doi: 10.1155/2015/641742

30. Müller VI, Cieslik EC, Laird AR, Fox PT, Radua J, Mataix-Cols D, et al. Ten simple rules for neuroimaging meta-analysis. *Neurosci Biobehav Rev*. (2018) 84:151–61. doi: 10.1016/j.neubiorev.2017.11.012

31. 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

32. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. (2019) 366:l4898. doi: 10.1136/bmj.l4898

33. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. (2016) 355:i4919. doi: 10.1136/bmj.i4919

34. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). *J Clin Epidemiol*. (2011) 64:407–15. doi: 10.1016/j.jclinepi.2010.07.017

35. Eickhoff SB, Bzdok D, Laird AR, Kurth F, Fox PT. Activation likelihood estimation meta-analysis revisited. *NeuroImage*. (2012) 59:2349–61. doi: 10.1016/j.neuroimage.2011.09.017

36. Eickhoff SB, Laird AR, Grefkes C, Wang LE, Zilles K, Fox PT. Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty. *Hum Brain Mapp*. (2009) 30:2907–26. doi: 10.1002/hbm.20718

37. Kim D, Chae Y, Park HJ, Lee IS. Effects of chronic pain treatment on altered functional and metabolic activities in the brain: a systematic review and Meta-analysis of functional neuroimaging studies. *Front Neurosci*. (2021) 15:684926. doi: 10.3389/fnins.2021.684926

38. Turkeltaub PE, Eickhoff SB, Laird AR, Fox M, Wiener M, Fox P. Minimizing within-experiment and within-group effects in activation likelihood estimation meta-analyses. *Hum Brain Mapp*. (2012) 33:1–13. doi: 10.1002/hbm.21186

39. Ziping L, Wen W, Shanshan Z, Shigui G, Jianming Y. Pain matrix response to acupuncture stimuli in individuals with acute low back pain: an FMRI study. *Chin J Pain Med*. (2013) 19:201–5. doi: 10.3969/j.issn.1006-9852.2013.04.003

40. Ziping L. *Acupuncture modulates acute low Back pain in the brain networks: An fMRI study*. Phd Southern Medical University (2013).

41. Runhui J, Xinzhu W, Xiaofei C, Fuwen D, Sheng Z, Qiqing C, et al. The impact of acupuncture treatment on brain functional connectivity networks in patients with low back pain. *J Clin Radiol*. (2021) 40:646–50.

42. Shi Y, Liu Z, Zhang S, Li Q, Guo S, Yang J, et al. Brain network response to acupuncture stimuli in experimental acute low Back pain: an fMRI study. *Evid Based Complement Alternat Med*. (2015) 2015:210120. doi: 10.1155/2015/210120

43. Lee J, Eun S, Kim J, Lee JH, Park K. Differential influence of acupuncture somatosensory and cognitive/affective components on functional brain connectivity and pain reduction during low Back pain state. *Front Neurosci*. (2019) 13:1062. doi: 10.3389/fnins.2019.01062

44. Xiang A, Yu Y, Jia X, Ma H, Liu H, Zhang Y, et al. The low-frequency BOLD signal oscillation response in the insular associated to immediate analgesia of ankle acupuncture in patients with chronic low back pain. *J Pain Res*. (2019) 12:841–50. doi: 10.2147/jpr.S189390

45. Tu Y, Ortiz A, Gollub RL, Cao J, Gerber J, Lang C, et al. Multivariate resting-state functional connectivity predicts responses to real and sham acupuncture treatment in chronic low back pain. *Neuroimage Clin*. (2019) 23:101885. doi: 10.1016/j.nicl.2019.101885

46. Makary MM, Lee J, Lee E, Eun S, Kim J, Jahng GH, et al. Phantom acupuncture induces placebo credibility and vicarious sensations: a parallel fMRI study of low Back pain patients. *Sci Rep*. (2018) 8:930. doi: 10.1038/s41598-017-18870-1

47. Kim H, Mawla I, Lee J, Gerber J, Walker K, Kim J, et al. Reduced tactile acuity in chronic low back pain is linked with structural neuroplasticity in primary somatosensory cortex and is modulated by acupuncture therapy. *NeuroImage*. (2020) 217:116899. doi: 10.1016/j.neuroimage.2020.116899

48. Zhu L, Sun Y, Kang J, Liang J, Su T, Fu W, et al. Effect of acupuncture on neurogenic claudication among patients with degenerative lumbar spinal stenosis: a randomized clinical trial. *Ann Intern Med*. (2024) 177:1048–57. doi: 10.7326/m23-2749

49. Lu L, Zhang Y, Tang X, Ge S, Wen H, Zeng J, et al. Evidence on acupuncture therapies is underused in clinical practice and health policy. *BMJ*. (2022) 376:e067475. doi: 10.1136/bmj-2021-067475

50. Chen XY, Chen N, Lai P, Sun YQ, Yu J, Xin M, et al. Multimodal abnormalities of brain function in chronic low back pain: a systematic review and meta-analysis of neuroimaging studies. *Front Neurosci*. (2025) 19:153288. doi: 10.3389/fnins.2025.1535288

51. Sacca V, Maleki N, Reddy S, Hodges S, Kong J. Assessing the modulatory effects of tDCS and acupuncture on cerebral blood flow in chronic low Back pain using arterial spin labeling perfusion imaging. *Brain Sci*. (2025) 15:261. doi: 10.3390/brainsci15030261

52. Li J, Zhang JH, Yi T, Tang WJ, Wang SW, Dong JC. Acupuncture treatment of chronic low back pain reverses an abnormal brain default mode network in correlation with clinical pain relief. *Acupunct Med*. (2014) 32:102–8. doi: 10.1136/acupmed-2013-010423

53. Tan WL, Wang W, Yang YC, Chen YL, Kang YJ, Huang YW, et al. Spinal manipulative therapy alters brain activity in patients with chronic low Back pain: a longitudinal brain fMRI study. *Front Integr Neurosci*. (2020) 14:534595. doi: 10.3389/fnint.2020.534595

54. Yoshino A, Maekawa T, Kato M, Chan HL, Otsuru N, Yamawaki S. Changes in resting-state brain activity after cognitive behavioral therapy for chronic pain: a magnetoencephalography study. *J Pain*. (2024) 25:104523. doi: 10.1016/j.jpain.2024.104523

55. Bao S, Qiao M, Lu Y, Jiang Y. Neuroimaging mechanism of cognitive behavioral therapy in pain management. *Pain Res Manage*. (2022) 2022:6266619. doi: 10.1155/2022/6266619

56. Seminowicz DA, Shpaner M, Keaser ML, Krauthamer GM, Mantegna J, Dumas JA, et al. Cognitive-behavioral therapy increases prefrontal cortex gray matter in patients with chronic pain. *J Pain*. (2013) 14:1573–84. doi: 10.1016/j.jpain.2013.07.020

57. Uddin LQ. Salience processing and insular cortical function and dysfunction. *Nat Rev Neurosci*. (2015) 16:55–61. doi: 10.1038/nrn3857

58. Broca P. Comparative anatomy of the cerebral convolutions: the great limbic lobe and the limbic fissure in the mammalian series. *J Comp Neurol*. (2015) 523:2501–54. doi: 10.1002/cne.23856

59. Vogt BA. Cingulate cortex in the three limbic subsystems. *Handb Clin Neurol*. (2019) 166:39–51. doi: 10.1016/B978-0-444-64196-0.00003-0

60. Rolls ET. The cingulate cortex and limbic systems for emotion, action, and memory. *Brain Struct Funct*. (2019) 224:3001–18. doi: 10.1007/s00429-019-01945-2

61. Almeida TF, Roizenblatt S, Tufik S. Afferent pain pathways: a neuroanatomical review. *Brain Res*. (2004) 1000:40–56. doi: 10.1016/j.brainres.2003.10.073

62. Davis KD, Taylor SJ, Crawley AP, Wood ML, Mikulis DJ. Functional MRI of pain- and attention-related activations in the human cingulate cortex. *J Neurophysiol*. (1997) 77:3370–80. doi: 10.1152/jn.1997.77.6.3370

63. Huang WC, Lyu D, Stieger JR, Gotlib IH, Buch V, Wagner AD, et al. Direct interactions between the human insula and hippocampus during memory encoding. *Nat Neurosci*. (2025) 28:1763–71. doi: 10.1038/s41593-025-02005-1

64. Zhang RH, Deng HF, Xiao X. The insular cortex: an interface between sensation, emotion and cognition. *Neurosci Bull*. (2024) 40:1763–73. doi: 10.1007/s12264-024-01211-4

65. Pepe S, von Gal A, Fabiani G, Piccardi L. Common brain areas in spatial navigation and visuo-spatial planning: a meta-analysis. *Neurosci Biobehav Rev*. (2025) 176:106295. doi: 10.1016/j.neubiorev.2025.106295

66. Jing YH, Lin T, Li WQ, Wu C, Li X, Ding Q, et al. Comparison of activation patterns in Mirror neurons and the swallowing network during action observation and execution: a task-based fMRI study. *Front Neurosci*. (2020) 14:867. doi: 10.3389/fnins.2020.00867

67. Charlesworth JEG, Petkovic G, Kelley JM, Hunter M, Onakpoya I, Roberts N, et al. Effects of placebos without deception compared with no treatment: a systematic review and meta-analysis. *J Evid Based Med*. (2017) 10:97–107. doi: 10.1111/jebm.12251

68. Akitsuki Y, Decety J. Social context and perceived agency affects empathy for pain: an event-related fMRI investigation. *NeuroImage*. (2009) 47:722–34. doi: 10.1016/j.neuroimage.2009.04.091

69. Bornhövd K, Quante M, Glauche V, Bromm B, Weiller C, Büchel C. Painful stimuli evoke different stimulus-response functions in the amygdala, prefrontal, insula and somatosensory cortex: a single-trial fMRI study. *Brain*. (2002) 125:1326–36. doi: 10.1093/brain/awf137

70. Ploner M, Schmitz F, Freund HJ, Schnitzler A. Parallel activation of primary and secondary somatosensory cortices in human pain processing. *J Neurophysiol*. (1999) 81:3100–4. doi: 10.1152/jn.1999.81.6.3100

71. Swanson LW, Petrovich GD. What is the amygdala? *Trends Neurosci*. (1998) 21:323–31. doi: 10.1016/s0166-2236(98)01265-x

72. Heimer L. A new anatomical framework for neuropsychiatric disorders and drug abuse. *Am J Psychiatry*. (2003) 160:1726–39. doi: 10.1176/appi.ajp.160.10.1726

73. Price JL, Drevets WC. Neurocircuitry of mood disorders. *Neuropsychopharmacology*. (2010) 35:192–216. doi: 10.1038/npp.2009.104

74. McLachlan RS. A brief review of the anatomy and physiology of the limbic system. *Can J Neurol Sci*. (2009) 36:S84–7.

75. Lee IS, Wallraven C, Kong J, Chang DS, Lee H, Park HJ, et al. When pain is not only pain: inserting needles into the body evokes distinct reward-related brain responses in the context of a treatment. *Physiol Behav*. (2015) 140:148–55. doi: 10.1016/j.physbeh.2014.12.030

76. Wang Z, Wang X, Liu J, Chen J, Liu X, Nie G, et al. Acupuncture treatment modulates the corticostriatal reward circuitry in major depressive disorder. *J Psychiatr Res*. (2017) 84:18–26. doi: 10.1016/j.jpsychires.2016.09.014

77. Kong J, Wang Z, Leiser J, Minicucci D, Edwards R, Kirsch I, et al. Enhancing treatment of osteoarthritis knee pain by boosting expectancy: a functional neuroimaging study. *Neuroimage Clin*. (2018) 18:325–34. doi: 10.1016/j.nicl.2018.01.021

78. Hashmi JA, Baria AT, Baliki MN, Huang L, Schnitzer TJ, Apkarian VA. Brain networks predicting placebo analgesia in a clinical trial for chronic back pain. *Pain*. (2012) 153:2393–402. doi: 10.1016/j.pain.2012.08.008

79. Borsook D, Becerra LR. Breaking down the barriers: fMRI applications in pain, analgesia and analgesics. *Mol Pain*. (2006) 2:30. doi: 10.1186/1744-8069-2-30

80. Dhond RP, Kettner N, Napadow V. Do the neural correlates of acupuncture and placebo effects differ? *Pain*. (2007) 128:8–12. doi: 10.1016/j.pain.2007.01.001

81. Bennett MI, Kaasa S, Barke A, Korwisi B, Rief W, Treede RD. The IASP classification of chronic pain for ICD-11: chronic cancer-related pain. *Pain*. (2019) 160:38–44. doi: 10.1097/j.pain.0000000000001363

82. Kang Y, Trewern L, Jackman J, McCartney D, Soni A. (2023). Chronic pain: definitions and diagnosis. BMJ. 381:e076036. doi: 10.1136/bmj-2023-076036

83. Raffaeli WA-O, Tenti M, Corraro A, Malafoglia V, Ilari S, Balzani EA-O, et al. Chronic pain: what does it mean? A review on the use of the term chronic pain in clinical practice. *J Pain Res*. (2021) 14:827–35. doi: 10.2147/JPR.S303186

84. Yu Z, Wang RR, Wei W, Liu LY, Wen CB, Yu SG, et al. A coordinate-based meta-analysis of acupuncture for chronic pain: evidence from fMRI studies. *Front Neurosci*. (2022) 16:1049887. doi: 10.3389/fnins.2022.1049887

85. Chang JR, Fu SN, Li X, Li SX, Wang X, Zhou Z, et al. The differential effects of sleep deprivation on pain perception in individuals with or without chronic pain: a systematic review and meta-analysis. *Sleep Med Rev*. (2022) 66:101695. doi: 10.1016/j.smrv.2022.101695

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 12 November 2025
DOI 10.3389/fmed.2025.1698050

Check for updates

OPEN ACCESS

EDITED BY
HaiHui Huang,
Shaoguan University, China

REVIEWED BY
Xuan Song,
Harvard University, United States
Tuo Han,
The Second Affiliated Hospital of Xi'an Jiaotong University, China

*CORRESPONDENCE
Zhengya Zhu
✉ xyfyzzy281@126.com

RECEIVED 03 September 2025
ACCEPTED 27 October 2025
PUBLISHED 12 November 2025

CITATION
Zhang Z, Chen Y, Wang Q, Li Z, Dai B, Dong C and Zhu Z (2025) Identification and validation of novel risk genes for intervertebral disc disorder by integrating large-scale multi-omics analyses and experimental studies.
Front. Med. 12:1698050.
doi: 10.3389/fmed.2025.1698050

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

Identification and validation of novel risk genes for intervertebral disc disorder by integrating large-scale multi-omics analyses and experimental studies

Zhe Zhang[1,2], Yu Chen[3], Qianjin Wang[1], Zheng Li[4], Bingyang Dai[5,6], Cheng Dong[5,6] and Zhengya Zhu[2]*

[1]Department of Orthopaedics and Traumatology,Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, China, [2]Department of Orthopaedics, The Affiliated Hospital of Xuzhou Medical University, Xuzhou, Jiangsu, China, [3]Division of Spine Surgery, Department of Orthopaedic Surgery, Nanjing Drum Tower Hospital, Affiliated Hospital of Medical School, Nanjing University, Nanjing, China, [4]Department of Orthopaedics, The Affiliated Hospital, Southwest Medical University, Luzhou, Sichuan, China, [5]Department of Biomedical Engineering, Faculty of Engineering, The Hong Kong Polytechnic University, Hong Kong, Hong Kong SAR, China, [6]The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen, China

Introduction: Although genome-wide association studies (GWAS) have identified multiple genetic loci linked to intervertebral disc disorder (IDD), their functional characterization remains largely unelucidated. We aim to leverage an integrative analytical pipeline to identify novel IDD risk genes from genetic associations and experimentally validate their functional roles.

Methods: We integrated transcriptome-wide association studies (TWAS), proteome-wide association studies (PWAS), expression and protein quantitative trait loci (eQTL and pQTL) colocalization analyses to identify potential causal genes for IDD. Enrichment analysis, expression profiling, protein-protein interaction (PPI) network construction, and druggability evaluation were also performed for the prioritized causal candidates. Subsequently, human intervertebral disc (IVD) tissues spanning degeneration grades and an *in vivo* mouse IDD model were employed to functionally characterize candidate risk genes.

Results: Integrative analysis of TWAS and PWAS with colocalization studies identified 104 genes and 10 proteins exhibiting causal associations with IDD. The identified genes/proteins were enriched in extracellular matrix organization, cellular senescence and collagen formation. Crucially, TMEM190, CILP2, and FOXO3 were demonstrated consistent evidence across TWAS, two independent PWAS datasets, and corresponding colocalization analyses, with CILP2 emerging as a potentially druggable target. Differential expression analysis revealed significant upregulated TMEM190 and CILP2, along with downregulated FOXO3 during IVD degeneration. These results were subsequently confirmed at protein levels in clinical specimens. Mouse model experiments further established that down-regulation of CILP2 alleviated IDD progression.

Discussion: Collectively, this work provides an updated compendium of putative IDD risk genes and delineates pathogenic roles for TMEM190, CILP2, and FOXO3, providing a broad hint for further research on novel mechanism and therapeutic targets for IDD.

KEYWORDS

intervertebral disc disorder, transcriptome-wide association study, proteome-wide association study, genome-wide association study, validation study

Introduction

Intervertebral disc disorder (IDD) represents a primary etiology of low back pain (1). Both genetic predispositions and environmental risk factors involved in its pathogenesis (2–5). Currently, the treatment of IDD primarily relies on symptomatic management with NSAIDs and surgical interventions for more severe cases. However, symptomatic treatments fail to address underlying disease mechanisms, while surgery entails significant costs, potential complications, and surgical morbidity (6). Therefore, identifying causative genes and developing targeted therapeutic strategies is imperative for IDD management.

Recent GWAS have identified multiple loci associated with IDD, predominantly within non-coding genomic regions (7, 8). These regions exhibit complex regulatory mechanisms and linkage disequilibrium, complicating the identification of underlying causal genes. TWAS coupled with eQTL colocalization address this limitation by linking non-coding disease-associated variants to transcriptional changes. In a TWAS study, genetic predictors of gene expression, specifically cis-eQTLs regulating nearby genes, are identified in reference populations, such as the Genotype-Tissue Expression (GTEx) project. These genetic predictors subsequently impute transcriptomic profiles in GWAS cohorts to identify gene expression levels and disease traits (9). eQTL colocalization analyses determine whether shared causal variants gene expression and disease risk share the same causal variants underlie both gene expression and disease risk, strengthening causal inference for candidate genes (10). While IDD-specific TWAS study remains scarce due to the limited large-scale human transcriptomic datasets of disc tissues, GTEx demonstrates substantial eQTLs conservation across tissues (11, 12). Thus, regulatory variants identified in non-disc tissues may still modulate disc biology and IDD susceptibility.

Complementary to TWAS, PWAS utilizes pQTL data to identify protein-level associations with diseases, providing enhanced mechanistic insight. Recently, large-scale human plasma proteome datasets, including those from the ARIC study and Iceland Biobank, have enabled robust pQTL derivation, facilitating practical PWAS implementation (13). Plasma proteins, which serve as key druggable targets and biomarkers for complex traits, can reflect systemic pathological changes associated with IDD. While PWAS has been applied to other diseases (14–16), its application to IDD remains unexplored. Future integration of PWAS with TWAS and eQTL/pQTL colocalization will enable the identification of disease-causing genes with higher precision, while minimizing confounding effects from horizontal pleiotropy. This multimodal approach will elucidate IDD molecular mechanisms and accelerate development of targeted therapeutics.

This study aimed to identify and validate potential causal genes associated with IDD by integrating multi-omics analyses with experimental approaches. We performed both TWAS and PWAS to uncover novel genes implicated in the pathogenesis of IDD (Figure 1). Colocalization analyses were performed to establish potential causal relationships between these genes and IDD risk. The expression patterns of the prioritized genes in human degenerated intervertebral discs were also assessed. Furthermore, enrichment analysis was conducted to identify key pathways and biological terms associated with IDD. Additionally, we explored protein-protein interactions among the candidates and evaluated their druggability. Finally, experimental validation of top-prioritized genes (TMEM190, CILP2, FOXO3) was conducted using clinical specimens and animal model.

Materials and methods

IDD GWAS summary data sources

The GWAS data were obtained from meta-analyses of data from several large cohorts, including deCODE Genetics from Iceland, the Danish Blood Donor Study, the Copenhagen Hospital Biobank, and the UK Biobank (7). Participants provided blood or buccal samples with informed consent, permitting the use of their samples and data in deCODE Genetics and the UK Biobank dataset were included. Each dataset was assumed to share a common odds ratio (OR), while allowing for different population frequencies of alleles and genotypes. Variants with imputation information scores below 0.8 were excluded from the analyses. The GWAS summary data used in our analysis come from the worldcome y data used in These analyses included 58,854 IDD cases and 922,958 controls, with the participants being of European descent. A total of 53.5 million sequence variants were included in the GWAS analysis.

Transcriptomic data from multiple human tissues

The eQTL data were obtained from GTEx Version 8 (49 tissues) (11). GTEx provides extensive data on the relationship between genetic variation and gene expression, sourced from 838 postmortem donors, and 15,201 RNA-sequencing samples were included, primarily of European descent.

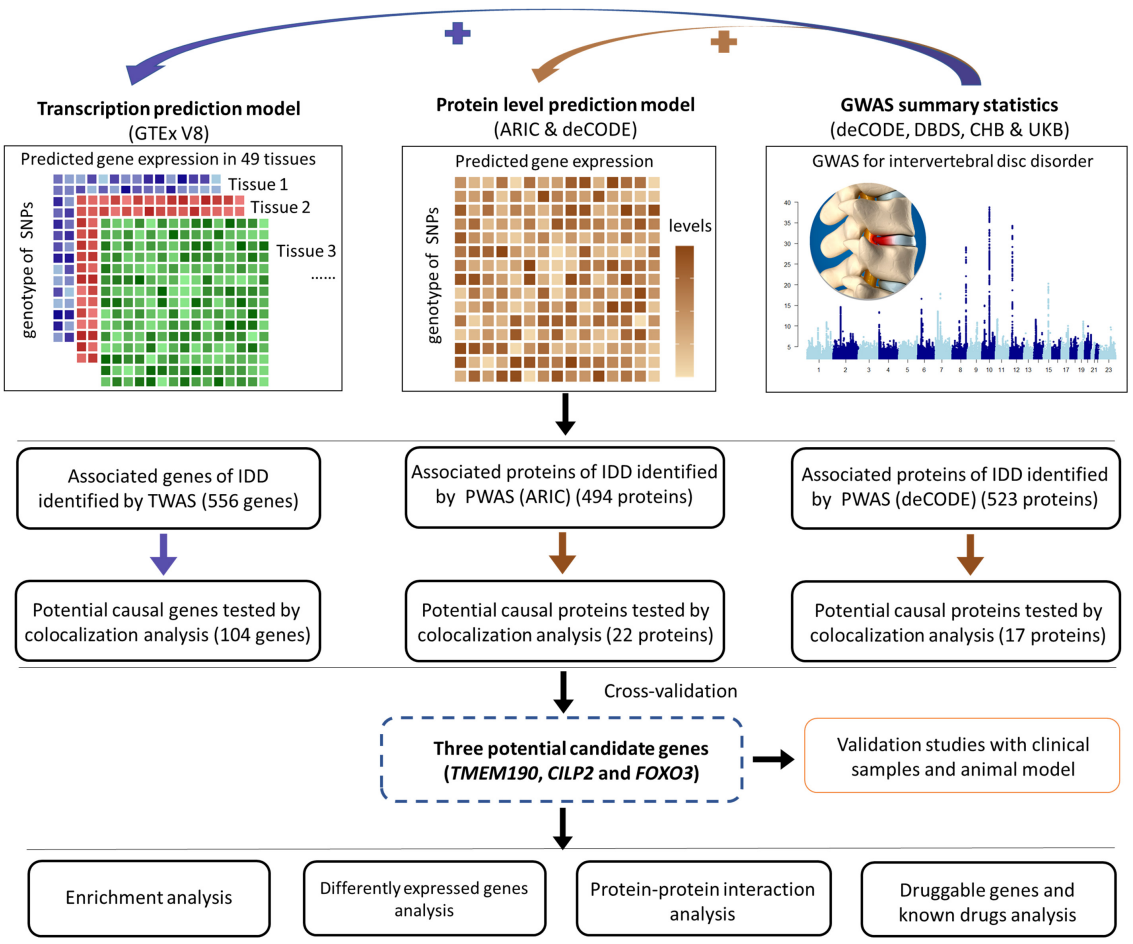


FIGURE 1
Overview of the study. This schematic illustrates the multistep approach employed to identify potential genes associated with IDD. First, TWAS and eQTL analyses were conducted to identify potential risk genes. Second, two independent PWAS and pQTL analyses were performed to identify potential causal proteins of IDD. Third, enrichment analyses were conducted to elucidate the functions of these potential causal genes/proteins. Fourth, data from the GEO were analyzed to identify differentially expressed potential causal genes. Additionally, PPI analyses were performed to explore interactions among the identified genes. Furthermore, the druggability of the potential causal genes and available drugs that target these genes were explored. Finally, validation studies were performed with clinical samples and an animal model.

Human protein abundance references in discovery proteome-wide association studies

The pQTL datasets incorporated in this study were derived from two large-scale investigations: the ARIC study, which includes data on 4,423 proteins from 7,213 individuals (13), and deCODE Genetics, which encompasses 4,428 proteins from 35,559 individuals (17), both primarily of European descent.

Human intervertebral disc degeneration microarray datasets

Human disc tissue expression microarray datasets were obtained retrospectively from the GEO (GSE56081: n = 10, with five degenerative and five normal samples).

Transcriptome-wide association studies

We performed TWAS analysis by integrating genome-wide summary statistics from an IDD GWAS with eQTL data from GTEx Version 8 across 49 tissue types as descripted before (18). To ensure consistency between datasets, we harmonized the IDD GWAS single nucleotide polymorphisms (SNPs) with the GTEx reference data, aligning SNP reference alleles, effect alleles, and associated metadata. Single-tissue TWAS was conducted for all tissues via SPrediXcan, followed by cross-tissue analysis via S-MultiXcan. S-MultiX can, which is based on a multitissue integration approach, allows for the combination of gene expression data across tissues, enhancing statistical power and enabling the identification of candidate susceptibility genes. We utilized default parameters for the software, with the exception of adjusting the "—cutoff_condition_number" parameter to 30. Only protein-coding genes were considered in the analysis, and significance was determined via a false discovery rate (FDR) threshold of $p < 0.05$.

FastENLOC colocalization

FastENLOC colocalization tool was used to strengthen the causal inferences drawn from our TWAS findings (19). Briefly, we computed posterior inclusion probability (PIP) values from IDD GWAS data via the torus tool, which quantifies the likelihood of each SNP's association with IDD. These PIPs were then input into fastENLOC, which performs colocalization analysis by integrating GWAS PIP values with precomputed GTEx multitissue eQTL annotations. Colocalization was performed for each tissue, producing gene-level colocalization probabilities (GLCPs), indicating the likelihood that a given variant influences both IDD GWAS and gene expression in each tissue. The results across all the tissues were then merged, and for each gene, the maximum GLCP value was retained to identify the tissue with the strongest colocalization signal. Genes with a maximum GLCP $>$ 0.5 were considered to have significant evidence of colocalization.

Proteome-wide association studies

BLISS method was used for PWAS analysis (20). Traditional PWAS approaches rely on individual-level proteomic data, which restricts the use of extensive summary-level pQTL datasets available in public repositories. The BLISS method enables the utilization of large-scale summary-level datasets for more efficient proteomic association analysis by constructing protein imputation models directly from summary-level pQTL data. In this study, we performed PWAS analyses via summary-level pQTL data from two large-scale cohorts: the ARIC study and deCODE Genetics. These datasets include over 4,000 proteins, facilitating a comprehensive analysis of protein–trait associations in the context of IDD. For discovery purposes, proteins with a nominal $p < 0.05$ were considered significant.

Bayesian colocalization analysis

We also performed Bayesian colocalization analyses via the coloc R package to investigate whether the identified associations between plasma proteins and IDD share the same causal variants rather than being affected by linkage disequilibrium (21). The Bayesian colocalization method evaluated evidence for five distinct hypotheses at each locus: (1) no association with either trait, (2) association with trait 1 only, (3) association with trait 2 only, (4) both traits are associated, but each has distinct causal variants, and (5) both traits share a common causal variant (22). Posterior probabilities for each hypothesis (H0, H1, H2, H3, and H4) were calculated as part of the analysis. Initial prior probabilities were assigned as follows: a SNP exclusively associated with trait 1 (p1) had a probability of $1 \times 10-4$, a SNP exclusively associated with trait 2 (p2) had a probability of $1 \times 10-4$, and a SNP shared by both traits (p12) had a probability of $1 \times 10-5$ (23). A posterior probability of H4 (PPH4) $>$ 0.5 was considered evidence of a shared causal variant between the two traits.

Enrichment analysis of significant findings

To further investigate the biological role of the identified genes, we performed enrichment analysis via Gene Ontology (GO) categories [encompassing biological processes (BPs), molecular functions (MFs), and cellular components (CCs)], Kyoto Encyclopedia Genes and Genomes (KEGGs) pathways, and Reactome pathways (24, 25). The analysis was performed via the clusterProfiler and Reactome PA R packages. Significant genes or proteins were defined as those with a $p < 0.05$. The background set for the enrichment analysis consisted of all genes or proteins tested in the study, representing the total gene/protein pool from which the significant findings were derived. The ggplot2 R package was used for visualization.

Annotation of prioritized genes/proteins

The genes/proteins prioritized through TWAS, PWAS, and colocalization were further annotated by evaluating their expression levels in degenerative disc tissues and constructing gene coexpression networks. First, we obtained human disc tissue expression microarray datasets from the GEO (GSE56081: $n = 10$, with five degenerative and five normal samples).[1] After normalization of the expression matrix, differential expression analysis was performed via the lmFit() and eBayes() functions from the limma package (26). Gene coexpression networks were subsequently constructed to further explore the relationships among the prioritized risk genes, including TMEM190, CILP2, and FOXO3 (15). Briefly, the gene expression matrix of IVD was used to perform correlation analysis between each risk gene and all other genes. The genes were then ranked on the basis of their correlation indices. These correlation coefficients were then used for gene set enrichment analysis (GSEA) with pathway data from Reactome, which was performed via the clusterProfiler package and visualized via the Ridgeplot R package. This approach highlighted significant biological functions and pathways associated with each prioritized gene. Significant enrichment was determined on the basis of an adjusted $p < 0.05$, a normalized enrichment score ($|NES|$) > 1, and an FDR < 0.25.

PPI analysis

To investigate potential causal genes implicated in IDD, we employed the STRING database to perform extensive network analysis. The 88 proteins associated with IDD in both TWAS and PWAS were analyzed (Supplementary Data 1). We only reserved connections with an interaction score greater than 0.4.

Analysis of druggable genes and known drugs

To explore the druggability of potential causal genes of IDD, we conducted druggable gene and known drug analyses. A previous

1 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc

study categorized druggable genes into three tiers (27). We categorized the genes identified via TWAS and eQTL colocalization analyses, as well as genes identified via PWAS (ARIC and deCODE) and pQTL colocalization analyses, and further searched for updated information on the drugs targeting the identified putative causal proteins in the Open Target Platform,[2] which is a comprehensive tool that promotes drug target identification via the integration of multiple databases.

Human IVD tissue collection

IVD tissue samples were collected from 6 patients undergoing spinal fusion surgery with the ethics approval of the Affiliated Hospital of Xuzhou Medical University (XYFY2023-KL337-01). The IDD cases were classified according to Pfirrmann's method (28). Samples falling within the I–II classification were labeled as controls, while falling within the III–V classification were designated as severe IDD samples (29). Patient information is provided in Supplementary Table 1. Informed consent was obtained from all patients.

Western blot analysis

Total protein was extracted from IVD using a complete cell lysis buffer and quantified with the BCA protein assay kit (Beyotime, China). Protein samples were separated via sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and transferred onto 0.2 μm PVDF membranes (Sigma-Aldrich, United States). The membranes were blocked with a 5% skim milk solution at room temperature and then incubated overnight at 4°C with primary antibodies specific to TMEM190 (1:500; Invitrogen, PA5-70986), CILP2 (1:500; Proteintech, 11813-1-AP), and FOXO3 (1:1,000; Biotime Biotechnology, AF609-1). Following three washes with Tris-buffered saline containing Tween 20 (TBST), the membranes were treated with secondary antibodies at room temperature. Immunoblotting was detected using the UVP ChemiDoc-It Imaging System (UVP, CA, United States) with an enhanced chemiluminescence detection kit (Thermo Fisher; 34,580) applied to the membranes. β-actin served as the loading control, and each blot was analyzed for integrated density using ImageJ software.

Animal experiments

All experiments were reviewed and approved by the committee of the Institutional Animal Care and Use Committee at Nanjing Drum Tower Hospital (approval number: DWSY-25005637). The methods were described as before (30). Briefly, after anesthesia, 31-G needle was poked into the 8-week-old male C57BL/6 IVD at 90° vertically, rotated 360° and held for 1 min. Sham operation was also performed. These operations were performed on coccygeal IVD. Immediately after the puncture, 2 ul shRNA targeted *Cilp2* gene or negative control encapsulated with the GV112 vector (Shanghai Genechem Co., Ltd.) were injected into the IVD. Four weeks after acupuncture and shRNA injection, the caudal IVD of mice were examined by MRI. Degeneration grade of mice IVD was calculated as described before (28). IVD tissues were then collected for the test of Cilp2 protein levels and histological staining.

Histological staining and analysis

The harvested IVD tissue was immersed in a 4% paraformaldehyde solution (Solarbio, China) for 48 h to preserve its morphology. It was then decalcified in a 10% EDTA solution (pH 7.2–7.4) for 2 weeks, with daily changes of the solution. The tissue underwent dehydration through a series of graded ethanol baths and was subsequently cleared using an environmentally friendly clearing agent (Solarbio, China). After clearing, the tissue was embedded in paraffin wax and sectioned into 5 μm-thick slices using a microtome. Hematoxylin and Eosin (H & E) staining were applied according to the instructions (Solarbio, China) to observe the histological morphology of the IVD. Finally, histological scoring of the H & E samples was conducted following methodologies described before (31). Simply, the stained results were evaluated from two perspectives, including annulus fibrosus and nucleus pulposus. Each category was assigned a score ranging from 0 to 3, yielding a cumulative score between 0 and 6. Higher scoring levels indicated greater degrees of degeneration.

Statistical analyses

A chi-square test was conducted to assess the difference in the number of colocalizing genes between IDD-associated and non-associated genes identified by the TWAS. Furthermore, two-tailed *t*-test or one-way ANOVA test were employed to assess: differential expression of TMEM190, CILP2, and FOXO3 in degenerative compared to non-degenerative IVD using GEO dataset, protein levels of these genes in clinical and animal specimens, and degeneration grade, as well as H & E scores in a murine IDD model. Statistical significance ($p < 0.05$) is denoted by asterisks (*) in figures.

Results

Identification of genes associated with IDD by TWAS

As shown in the pipeline (Figure 1), we initiated our analysis by performing a cross-tissue TWAS based on data from reference gene expression predictions from GTEx and the largest intervertebral disc disorder GWAS conducted to date. TWAS analysis of 17,342 protein-coding genes identified 556 significantly IDD-associated genes (FDR-adjusted $p < 0.05$). The top 10 genes most strongly correlated with IDD were *CHST3*, *SOX5*, *SPOCK2*, *SMAD3*, *FGFR3*, *C6orf106*, *GFPT1*, *TWIST1*, *ASCC1*, *IGFBP3* (Figure 2 and Supplementary Data 2). Among the 556 genes, 20 genes, including SMAD3, mapped to nearest genes at the previously reported GWAS susceptibility loci. Our analysis also suggested the novel associations of the remaining 536 genes with IDD risk.

2 https://platform.opentargets.org/

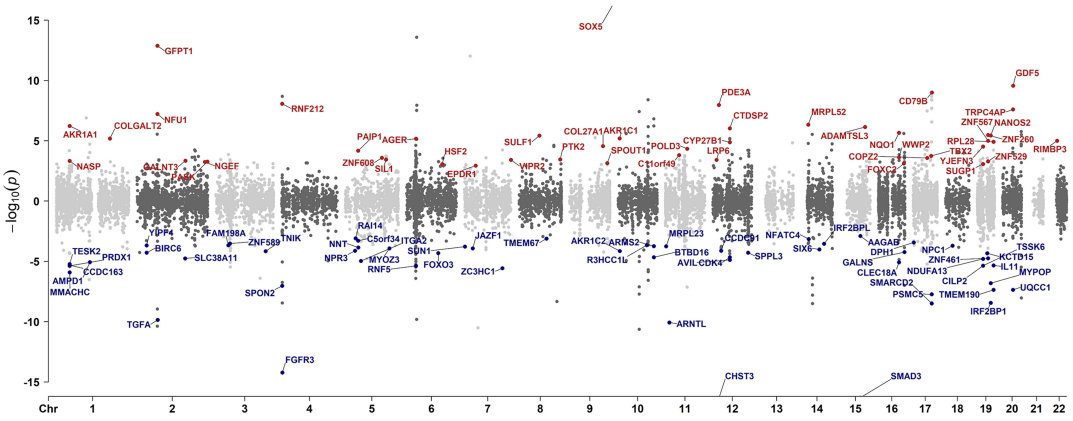


FIGURE 2
Manhattan plot illustrating TWAS gene associations. Each dot represents a gene plotted according to its genomic position (x-axis) and the significance of the association, measured as the $-\log_{10}$(FDR-adjusted p-value) (y-axis). Highlighted points with corresponding gene labels indicate genes meeting stringent colocalization criteria: FDR-adjusted $p < 0.05$ and colocalization max-GLCP > 0.5. The color of the highlighted points indicates the directionality of the genetic effect: red represents positive z-mean values (z-mean > 0) and blue for negative z-mean values (z-mean < 0).

We then conducted enrichment analysis of the significant findings using the KEGG, Reactome, and GO databases. Notably, the top enriched pathways of TWAS included extracellular matrix organization, cellular senescence, and skeletal system and connective tissue development, all of which established mechanisms in IDD pathogenesis. Other prominent enriched terms included calcineurin activates NFAT, glycosphingolipid biosynthesis, aspartate and asparagine metabolism, cartilage development, response to transforming growth factor-beta, chondrocyte differentiation, regulation of lipid kinase activity, and signal transduction pathways (Figures 3A, B and Supplementary Figure 1).

Colocalization between IDD risk loci and eQTLs

To determine whether TWAS-identified associations with IDD are driven by shared causal variants, we performed eQTL colocalization analysis using fastENLOC across 49 GTEx tissues for all protein-coding genes. This analysis identified 146 genes with strong colocalization evidence (max-GLCP > 0.5), among which 104 genes were TWAS- prioritized (Figure 2 and Supplementary Data 3, 4). TWAS-prioritized genes showed significant enrichment for colocalization signals compared to a matched background set ($\chi^2 = 2195.6$, fold-enrichment = 91.7, $p < 0.001$) (Supplementary Table 2), highlighting the specificity and robustness of our findings.

Identification of plasma proteins associated with IDD by PWAS

To identify proteins potentially associated with IDD for further validation, we conducted two independent PWAS by integrating GWAS summary statistics with human plasma proteomic data from the ARIC consortium and deCODE Genetics. The ARIC-based PWAS identified 494 significant associations, and the deCODE-based PWAS yielded 523 associations (Figures 4A, B; Supplementary Data 5, 6) ($p < 0.05$). Among these, 153 proteins were consistently associated with IDD across both datasets (Supplementary Data 7). Among them, six proteins, TMEM190, CILP2, FOXO3, SPON2, GALNT3, and NUF1, were additionally supported by TWAS and eQTL colocalization analyses, further reinforcing their relevance to IDD.

For the PWAS results from the ARIC cohort, the most significant pathways were collagen formation, extracellular matrix organization, and glycosphingolipid/sphingolipid metabolism (Figures 3C, D and Supplementary Figure 2). Similarly, the deCODE PWAS analysis revealed glycosphingolipid/sphingolipid metabolism, regulation of actin cytoskeleton, extra-nuclear estrogen signaling, and signaling pathways associated with GPER1, NOTCH1, PI3K-Akt, and Hedgehog (Figures 3E, F and Supplementary Figure 3). The high degree of consistency between the TWAS and PWAS results across both datasets confirms the robustness of these findings.

Colocalization between IDD risk loci and pQTLs

To provide causal evidence for IDD-associated proteins, we performed pQTL colocalization analyses. In the ARIC dataset, 26 proteins demonstrated strong colocalization signals with IDD risk loci (PPH4 > 0.5; Supplementary Data 8). Among these, 22 proteins were also significantly associated with IDD in the PWAS (Figure 4A; Supplementary Data 9). In the deCODE dataset, pQTL colocalization identified 24 significant proteins (PPH4 > 0.5; Supplementary Data 10), of which 16 proteins showed consistent PWAS associations with IDD (Figure 4B; Supplementary Data 11). Totally, 10 proteins were identified as causal proteins via ARIC based and deCODE based PWAS and their respective colocalization analyses (Supplementary Data 12). Among these,

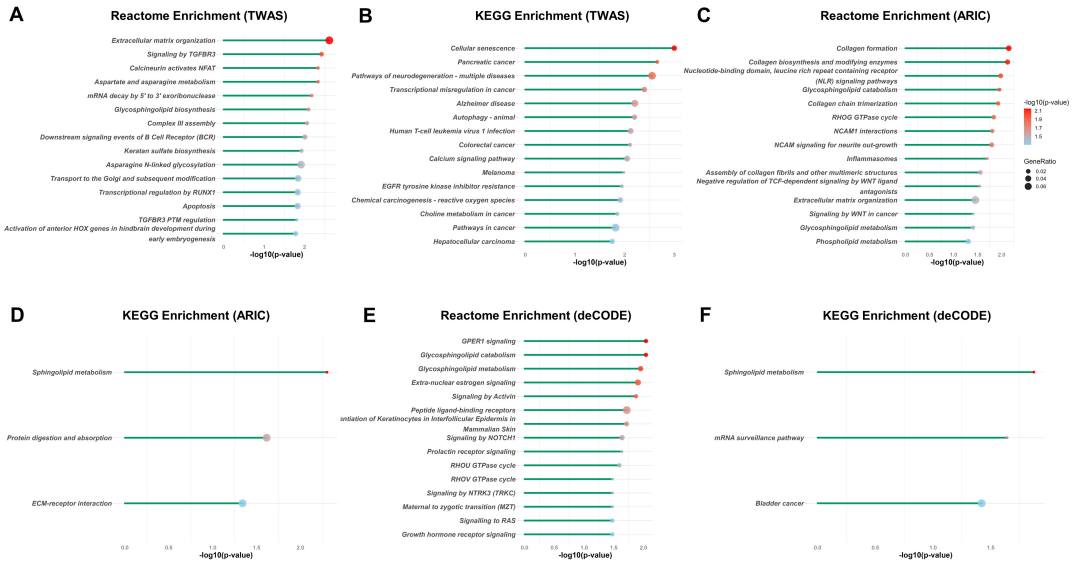


FIGURE 3
Reactome and KEGG enrichment analyses of the potential associated genes of IDD identified with TWAS and PWAS. Reactome and KEGG enrichment analysis of the potential associated genes of IDD identified with TWAS **(A,B)**, with PWAS using ARIC dataset **(C,D)**, and with PWAS using deCODE dataset **(E,F)**. Each line represents a pathway with significance defined by an FDR-adjusted $p < 0.05$. The color intensity represents statistical significance. The dot size corresponds to the gene ratio, which is defined as the number of genes of a pathway to the total number of genes analyzed.

TMEM190, CILP2, and FOXO3 were also supported by TWAS and eQTL colocalization analysis.

Collectively, TMEM190, CILP2, and FOXO3 emerged as proteins with strong causal evidence for IDD, supported across multiple omics layers including TWAS, two independent PWAS datasets, and both eQTL and pQTL colocalizations (Table 1).

Evaluation of the expression levels of genes/proteins identified by TWAS/PWAS

To identify differentially expressed genes (DEGs) in degenerative intervertebral discs, we analyzed mRNA expression profiles from human disc tissues using the microarray dataset GSE56081. Analysis of the GSE56081 dataset encompassing 13,170 genes captured 537 (96.6%), 395 (80.0%), and 404 (77.2%) of IDD-associated genes/proteins identified by TWAS, ARIC based, and deCODE based PWAS, respectively. Among these genes, 2,877 were significantly upregulated and 3,140 were downregulated in degenerated discs (Figure 5A; Supplementary Data 13). We observed 189 genes overlapped with TWAS-prioritized candidates (Supplementary Data 14) and 53 overlapped with proteins identified in both PWAS analyses (Supplementary Data 15).

In particular, all three genes (*TMEM190*, *CILP2*, and *FOXO3*) priorized by multiple-omics analyses were found to be significantly differentially expressed in degenerative intervertebral discs ($p < 0.05$) (Figures 5A, B and Table 1). Specifically, *TMEM190* and *CILP2* were upregulated in IDD samples, whereas *FOXO3* was downregulated compared to the control group. Notably, *CILP2* showed the most pronounced change, exhibiting a 1.5-fold increase in expression in degenerated discs relative to controls.

Functional annotation of *TMEM190*, *CILP2*, and *FOXO3*

The three potential causal genes were analyzed within the framework of GSEA to investigate their functions by coexpression analysis (Figure 5C). The GSEA results revealed that all of the three genes involved in the expression and translation of olfactory receptors, sensory perception and SRP-dependent cotranslational protein targeting to the membrane pathways.

TABLE 1 Summary of three potential causal genes of IDD indicated by TWAS, PWAS and colocalization analyses.

Gene	Chr	Discovery of TWAS		Validation of PWAS (ARIC)		Validation of PWAS (deCODE)		Expression
		Max-GLCP	FDR *p*	PPH4	*P*-value	PPH4	*P*-value	
TMEM190	19	0.93	1.83×10^{-5}	0.99	3.00×10^{-5}	0.99	3.74×10^{-5}	Up-regulated
CILP2	19	0.69	8.62×10^{-4}	0.92	9.00×10^{-4}	0.70	1.75×10^{-5}	Up-regulated
FOXO3	6	0.87	0.006	0.79	0.002	0.69	0.016	Down-regulated

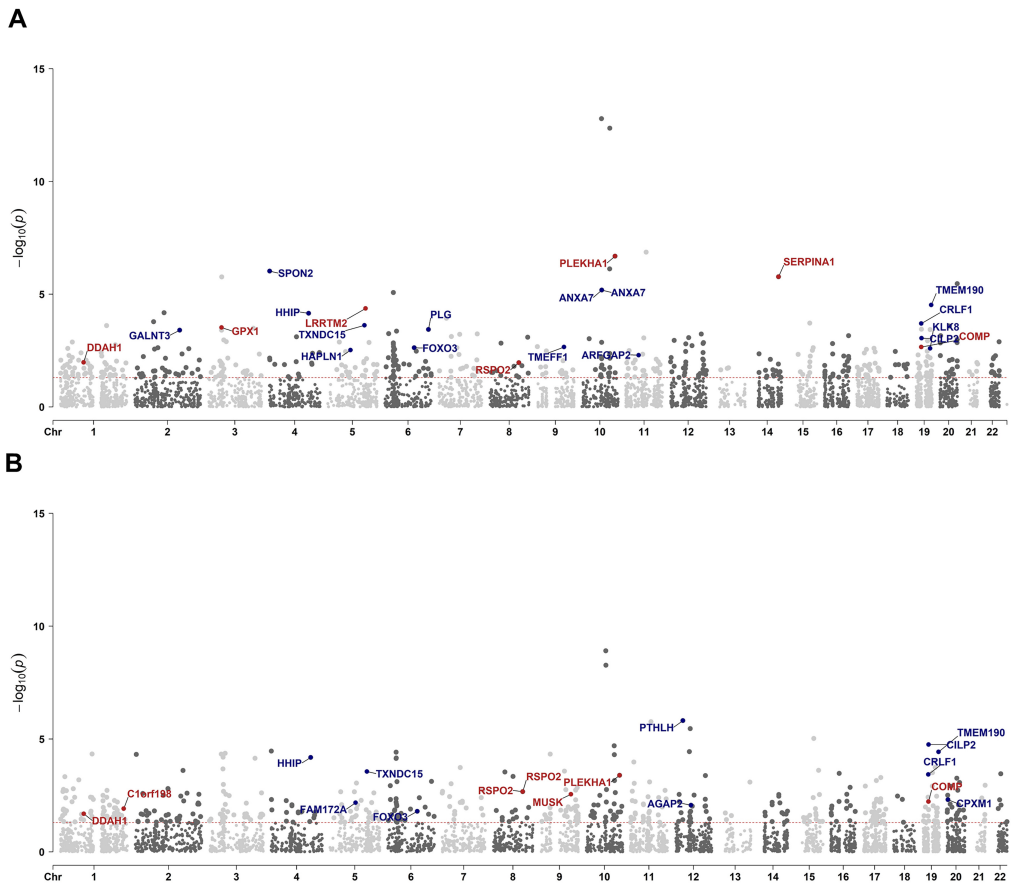


FIGURE 4
Manhattan plot illustrating PWAS protein associations on the basis of ARIC and deCODE data. Manhattan plot for the ARIC based **(A)** and deCODE based **(B)** PWAS of IDD. Each dot represents a protein plotted according to its genomic position (x-axis), and the significance of the association was measured as the $-\log_{10}(p\text{-value})$ (y-axis). Highlighted points and their protein labels indicate proteins meeting stringent colocalization criteria: $p < 0.05$ and colocalization PPH4 > 0.5. The color of the highlighted points indicates the directionality of the genetic effect: red for positive beta values (beta > 0) and blue for negative beta values (beta < 0).

To elucidate the interactions among the candidate genes associated with IDD (TMEM190, CILP2, and FOXO3), we performed PPI analysis involving 88 proteins associated with IDD identified in TWAS and PWAS (Supplementary Data 1). There were 27 genes whose connections had interaction scores greater than 0.4. Notably, TMEM190 did not interact with either CILP2 or FOXO3. Although no direct interactions were observed between CILP2 and FOXO3, several core proteins—SMAD3, COMP, IGFBP3, IGF1R, COL10A1, RUNX3, and PTK2—were identified as mediators of the interaction between CILP2 and FOXO3 (Figure 6).

Druggability of the identified genes and proteins

Among the genes identified through TWAS and eQTL analyses, 33 protein-coding genes were classified within the druggable genome: including 16 in tier 1, 8 in tier 2, and 9 in tier 3 (Supplementary Data 16). By searching the Open Target Platform, we identified several approved or investigational drugs targeting risk genes indicated by TWAS and eQTL analyses, including FGFR3, TGFA, CD79B, PDE3A, NQO1, AGER, ITGA2, CDK4, COL27A1, and PTK2 (Supplementary Data 17). Additionally, of the proteins identified through PWAS and pQTL analyses via ARIC or deCODE, 14 protein-coding genes were classified within the druggable genome: including 4 in tier 1 and 10 in tier 3 (Supplementary Data 18). We further identified several approved or investigational drugs targeting risk genes of IDD indicated by PWAS and pQTL analyses, namely PLG and PTHLH (Supplementary Data 19). Among the three potential causal genes, only CILP2 was druggable, while TMEM190 and FOXO3 were not in the druggable genome.

Validation studies for potential causal genes of IDD with clinical samples and animal model

To explore the roles of *TMEM190*, *CILP2*, and *FOXO3* in IDD, we assessed their expression in human IVD specimens from mild degeneration (Grades I and II) and severe degeneration (Grades III, IV, and V). Western blot results showed increased expression of *TMEM190* and *CILP2* with concurrent decreased expression

of *FOXO3* in severely degenerated IVD tissues (Figures 7A, B), aligning with findings in DEGs analysis.

Given the combined evidence supporting CILP2, and its classification as a druggable target, we further investigated its role in IDD using a needle-induced IVD degeneration mouse model. As shown in Figures 8A, B, the protein level of CILP2 in the IVD was up-regulated after puncture treatment. However, this level decreased significantly following shRNA transfection. MRI examinations revealed that the grade score of the IVD was significantly increased after puncture treatment, while it decreased markedly with CILP2 knockdown (KD) (**Figures 8C,D**). H & E staining showed the degenerated progression was alleviated following the down-regulation of CILP2 (Figures 8E, F). These results suggest that down-regulation of CILP2 reduces the susceptibility to intervertebral disc degeneration progression in the mouse model of IDD.

Discussion

To the best of our knowledge, this study is the first to employ multidimensional multi-omics data, including high-throughput genomics, whole-body transcriptomics, plasma proteomics and intervertebral disc transcriptomics, to investigate potential risk genes for IDD. Our integrative approach presented 104 TWAS-identified genes and 10 PWAS-identified proteins with IDD based on converging evidence supported by eQTL/pQTL colocalization analyses. These genes/proteins were enriched for key regulators of disc pathology, such as glycosphingolipid/sphingolipid metabolism. Three potential causal genes, *TMEM190*, *CILP2*, and *FOXO3*, were consistently supported by TWAS, two independent PWAS and colocalization analyses. These three genes were dysregulated in degenerated human discs, with *CILP2* further classified as druggable. We also validated the role of these causal genes, *TMEM190*, *CILP2* and *FOXO3* with clinical samples, as well as the role of CILP2 with animal model in IDD.

Glycosphingolipid/sphingolipid metabolism consistently emerged as a key pathway across all enrichment analyses of the identified associations. Both the TWAS and the PWAS results from the ARIC and deCODE cohorts strongly highlight this pathway as a critical factor in the pathogenesis of IDD. Sphingolipids, including ceramide and sphingosine-1-phosphate, constitute a major class of lipids found in all eukaryotic cells. Sphingolipids regulate a wide range of biological processes, including inflammation, mitochondrial function and apoptosis (32–34). The metabolic processes involved in sphingolipid biosynthesis and regulation were significantly enriched, underscoring their potential role in IDD. This pathway's involvement in inflammation and apoptosis suggests that targeting sphingolipids synthesis could serve as a promising therapeutic strategy to alleviate disc disorders and associated pain.

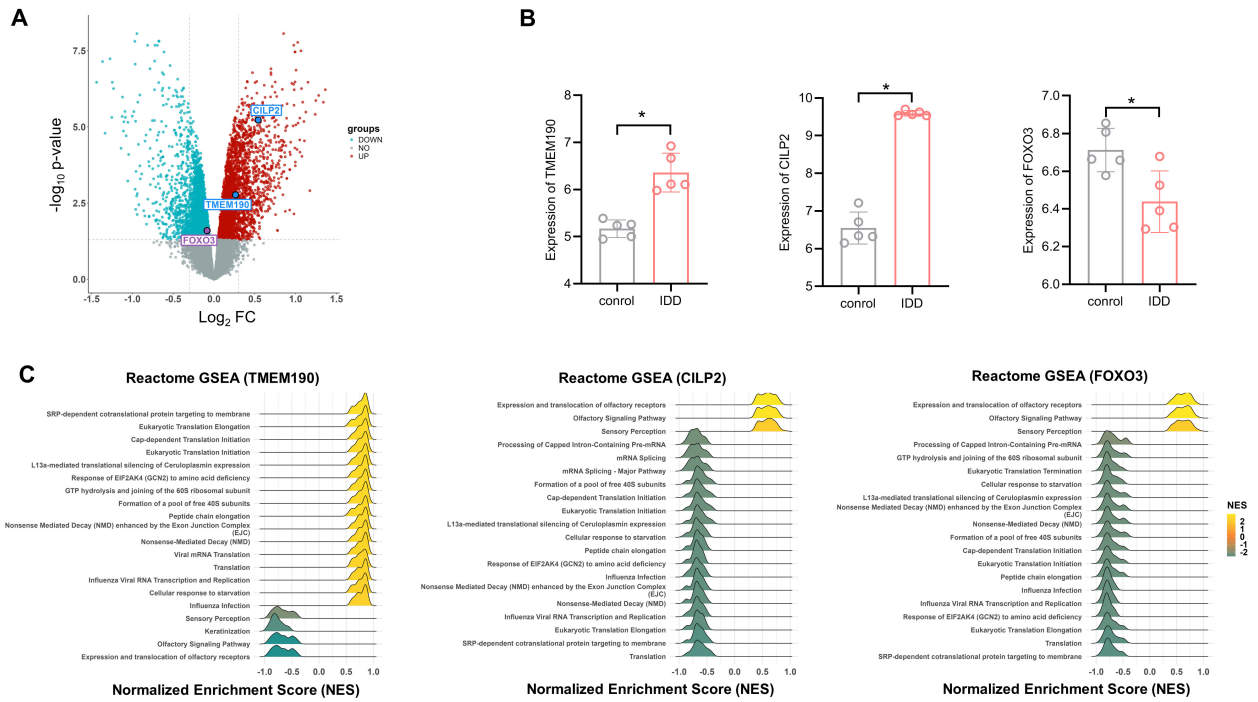


FIGURE 5
DEGs and enrichment analysis of the three potential causal genes of IDD. **(A)** Volcano plot of the DEG analysis. Each dot represents a gene plotted according to the significance of the association measured as the $-\log_{10}$ (FDR-adjusted p-value) (y-axis). The colors of the points are as follows: red for upregulated genes, blue for downregulated genes, and gray for non-differentially expressed genes (FDR-adjusted $p \geq 0.05$). **(B)** Expression levels of potential causal genes for IDD in degenerative vs. non-degenerative groups, based on microassy data from clinical samples $n = 5$. Two-tailed t-test. *$p < 0.05$. **(C)** Significantly enriched pathways for the three potential causal genes (*TMEM190*, *CILP2* and *FOXO3*) as determined by GSEA. Each line represents a pathway with significance defined by an FDR-adjusted $p < 0.05$. Yellow indicates upregulation (NES > 0), while green indicates downregulation (NES < 0).

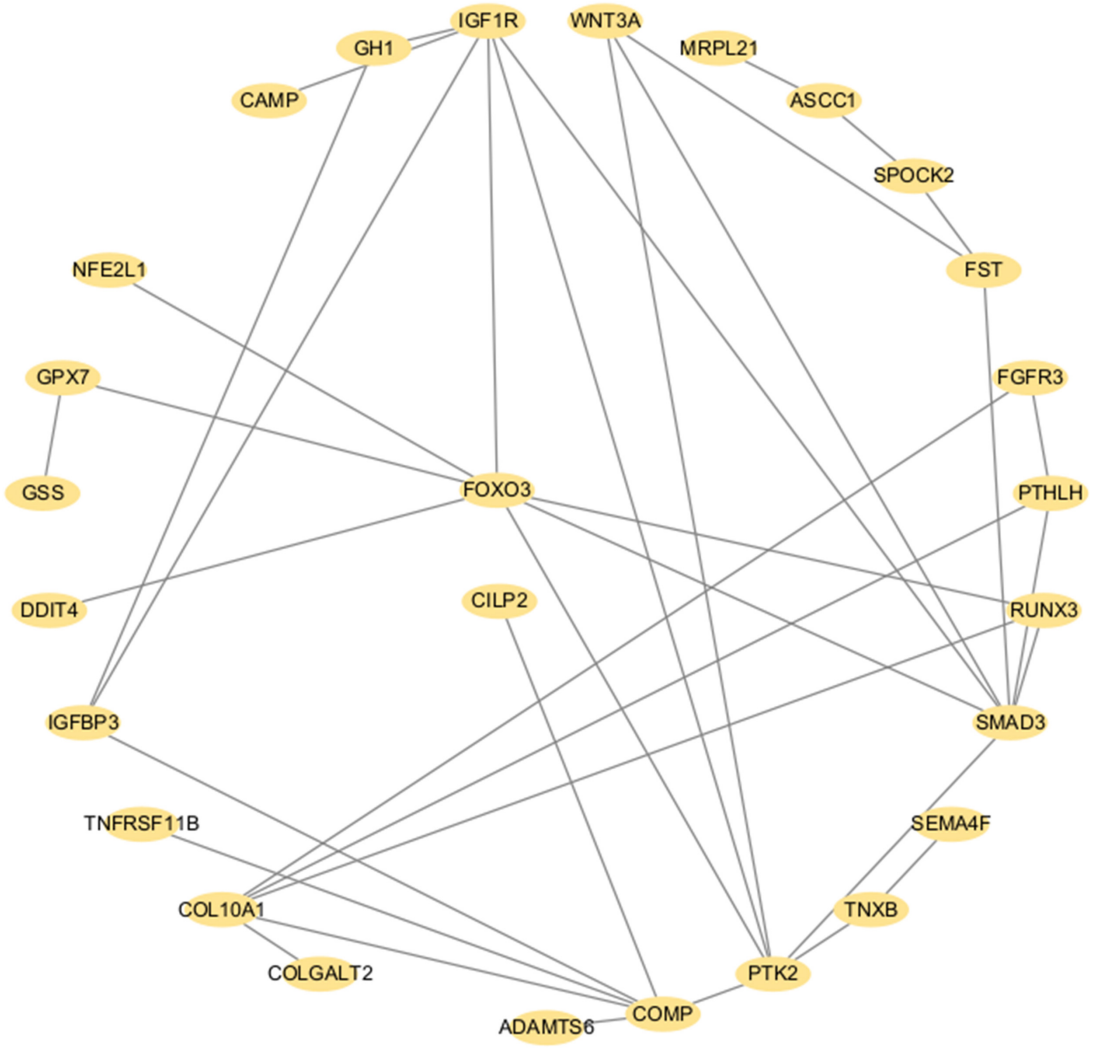


FIGURE 6
PPI network of the significant proteins associated with IDD identified via TWAS and PWAS analyses. Each dot represents a protein. Lines denote physical PPIs.

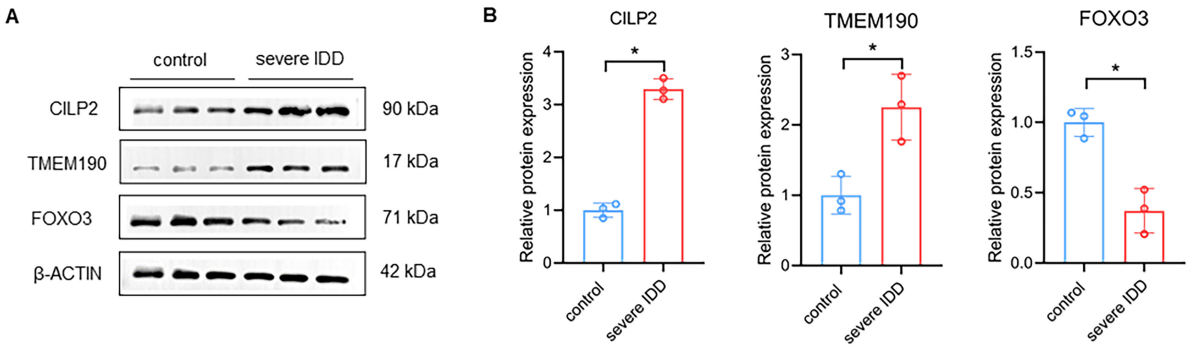


FIGURE 7
Validation of risk genes in clinical IVD tissues. **(A,B)** Western-blot analysis of TMEM190, CILP2, and FOXO3 in control and severe IDD patients. $n = 3$. Data represent the mean (SD). Two tailed t-test. *$p < 0.05$.

Among the identified genes (*TMEM190*, *CILP2*, and *FOXO3*), *FOXO3* has been previously investigated in the context of IDD. FOXO3, a member of the forkhead box O transcription factor family, is known to regulate critical cellular processes, including the cell cycle, apoptosis, and metabolism, and is implicated in age-related diseases (35, 36). FOXO3 has been linked to IDD in

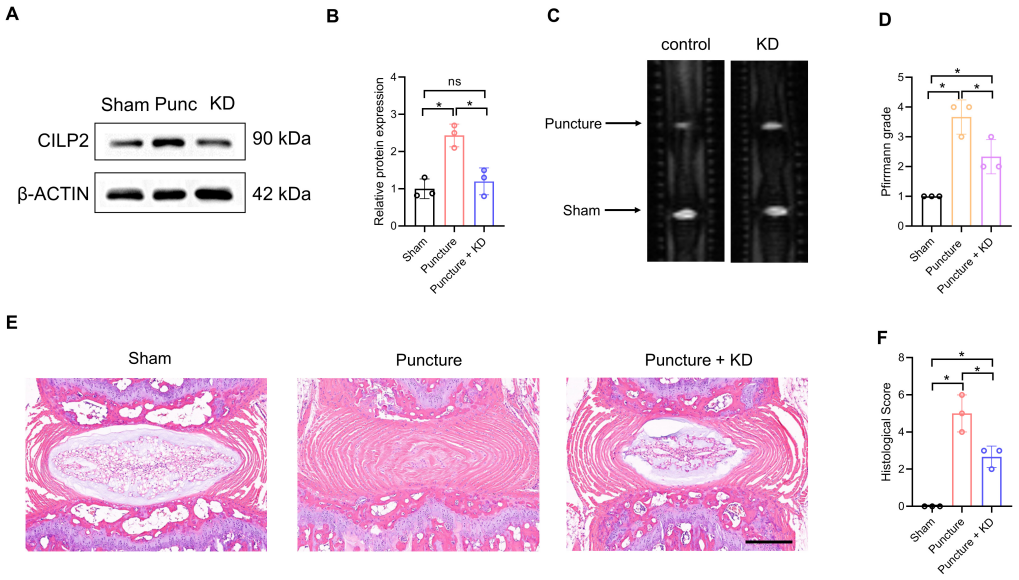


FIGURE 8
Down-regulation of CILP2 alleviated IDD progression in mouse model of IDD. **(A,B)** Western-blot analysis of CILP2 in Sham, puncture (Punc) and knocking down (KD) IVD tissues in mouse model of IDD. $n = 3$. **(C,D)** Magnetic resonance imaging (MRI) and Pfirrmann grades of IVD in mice treated as in **(A)**. $n = 3$. **(E,F)** Hematoxylin and Eosin (H&E) staining and histological score assessment of IVD in mice treated as in **(A)**. $n = 3$. Scale bar = 200 μm. Data represent the mean (SD). One-way ANOVA. $*p < 0.05$.

numerous studies, where it functions as a mediator regulating the role of specific genes in the disease, such as YTHDF2 and P300 (37, 38). Furthermore, FOXO3 is involved in the molecular mechanisms of potential therapeutic agents for IDD, such as stem cell-derived exosomes and procyanidin C1, primarily by regulating oxidative stress (39, 40), which reinforces the potential of FOXO3 as therapeutic interventions for this condition. The current study found that FOXO3 was down-regulated in severely degenerated disc tissues, which aligns with previous findings (38, 41). Our study provides additional evidence for dysregulated FOXO3 expression in IDD. However, the contribution of FOXO3 to IDD still requires further exploration.

The roles of the other two identified genes, *TMEM190* and *CILP2*, in IDD are less well characterized. TMEM190 is located on chromosome 7 and contains five exons, which encode a small single-pass transmembrane protein (42). Small single-pass transmembrane proteins may be associated with mitochondrial oxidative phosphorylation (43), which has been linked to IDD. Additionally, TMEM190 may contribute to chondrocyte dedifferentiation (44). Given that cartilage endplate remodeling and altered chondrocyte subsets play key roles in IDD progression (45, 46), TMEM190 may involve in IDD pathogenesis. CILP2, a member of the cartilage intermediate layer protein family, encodes a matricellular protein predominantly expressed in cartilage cells but also in various other tissues (47). Quantitative proteomic analysis and immunohistochemistry have demonstrated increased *CILP2* levels in degenerated human intervertebral discs (48, 49). In current study, we found CILP2 was up-regulated in severe IDD tissues, which provides additional evidence that CILP2 play a role in IDD. Importantly, the inhibition of Cilp2 has been shown to improve mitochondrial dysfunction in sarcopenia via the WNT signaling pathway (47). Given the established roles of mitochondrial dysfunction and WNT signaling in IDD (50, 51), CILP2 is likely to play an important role in this condition. In the current study, we found that down-regulation of CILP2 alleviated IDD progression in mouse model of IDD. Our results provide more direct evidence for the role of CILP2 in the progression of IDD. Our multi-omics investigation and validation study with clinical samples and animal model offer evidence supporting the role of CILP2 as a disease-causing gene and therapeutic target in IDD. However, the functional mechanism of CILP2 in IDD was not explored in current study. It has been reported that CILP2 affect sarcopenia and hypertrophic scar by antagonizing Wnt signaling pathway (47), and reducing the ubiquitination of ACLY (52), respectively. Further research is warranted to elucidate the precise role of CILP2 in modulating IDD progression through these candidate signaling pathways.

While no direct interactions among the three genes (*TMEM190*, *CILP2* and *FOXO3*) were reported, GSEA revealed their collective involvement in the olfactory signaling pathway, and sensory perception. Notably, olfactory stem cells have been shown to exhibit a chondrogenic phenotype, promoting IVD regeneration in a rat model of disc injury (53), which indicates these three genes collectively contributed to the pathology of IDD. In addition, pain is a significant symptom of IDD. It has been reported that anti-sensory nerve transmission significantly suppresses inflammatory pain markers (54). The involvement of these three genes in sensory perception indicates that they share a similar pathway for the contribution of pain to IDD. To further explore the correlations, we performed a PPI analysis using data from TWAS and PWAS. Although no direct interactions were found among the three genes, several mediators of interaction between CILP2 and FOXO3 were identified, including COMP and IGFBP3, which have been linked to IDD progression (55, 56). These findings suggest that CILP2 and FOXO3 may collaboratively influence IDD through these mediators.

Our study has several strengths. Primarily, it integrates both genomic and proteomic data to provide comprehensive insights into the complex biological systems underlying IDD. Additionally, the validation of potential causal genes through two independent PWAS analyses strengthens the reliability of our findings. Furthermore, the datasets utilized, comprising extensive human transcriptomes, proteomes, and IDD GWAS data, are among the largest and most complete to date, enhancing the robustness of the results. Finally, we validated the risk genes derived from public datasets using clinical samples and animal model, which enhances the reliability of our findings.

Some limitations must be acknowledged. First, the PWAS for IDD utilized human blood proteome data; however, plasma proteins serve as systemic biomarkers and may not fully capture disc-specific changes, potentially introducing bias. Future studies should examine the proteomes specific to human intervertebral discs. Secondly, the identification of susceptible genes from a European database, coupled with clinical validation using specimens from the Chinese population, introduces population heterogeneity that may limit the generalizability of the findings. Future cross-ethnic validation studies should be conducted to assess the robustness of these findings across diverse populations and ensure their applicability in broader clinical contexts. Additionally, the mechanisms by which identified risk genes and the relevance of their enriched pathways contribute to IDD remain unclear, and additional studies are needed to further evaluate their potential as therapeutic targets. Besides, only the mouse tail disc puncture model was employed, which is an acute injury model and may not adequately replicate the chronic, progressive nature of human IDD. Also, the current transcriptomic samples predominantly represent European populations, whereas the proteomic samples are from American populations, and expanding the diversity of these datasets may yield more accurate estimations and broader applicability. Finally, gene-environment interactions and assortative mating could influence genetic effects and contribute to variance in the analysis. Unfortunately, due to the limitations inherent in the current dataset and the scope of the study design, it is not feasible to adjust for these factors in this particular analysis. Nevertheless, the strength of our study lies in its innovative integration of multi-omics data, which positions it as one of the first efforts to identify and validate novel genetic risk factors for IDD in such a comprehensive manner.

Conclusion

In summary, we present an expanded resource of putatively causal genes associated with IDD, and highlight three novel potential causal genes (*TMEM190*, *CILP2*, and *FOXO3*). These findings provided a broad hint for further research on the potential mechanisms underlying IDD pathogenesis and highlight novel therapeutic targets for future investigations.

Data availability statement

All the data used in this study are publicly available without the need for special access. The specific sources of the data are as follows: TWAS Process: https://github.com/xqwen/fastenloc. IDD GWAS: https://pmc.ncbi.nlm.nih.gov/articles/PMC8810832. TWAS Prediction Model: https://predictdb.org/post/2021/07/21/gtex-v8-models-on-eqtl-and-sqtl/. fastENLOC Colocalization: Pre-computed GTEx multi-tissue eQTL annotations with hg38 position ID from https://github.com/xqwen/fastenloc. BLISS Software: Data calculations were performed using BLISS, which can be accessed at https://github.com/gcb-hub/BLISS. GEO database (GSE56081).

Ethics statement

The studies involving humans were approved by the Committee of the Affiliated Hospital of Xuzhou Medical University. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin. The animal study was approved by the committee of the Institutional Animal Care and Use Committee at Nanjing Drum Tower Hospital. The study was conducted in accordance with the local legislation and institutional requirements.

Author contributions

ZZha: Writing – original draft. YC: Writing – review & editing. QW: Writing – review & editing. ZL: Writing – review & editing. BD: Writing – review & editing. CD: Writing – review & editing. ZZhu: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. This work was supported by the Talent Introduction Project of the Affiliated Hospital of Xuzhou Medical University (2024203033), the Science and Technology Project of the Affiliated Hospital of Xuzhou Medical University (2024ZL32), and the National Nature Science Foundation of China (82202700).

Acknowledgments

We thank all the participants and investigators of GWAS, transcriptomics and proteomics studies 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.

Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

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

Publisher's note

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/fmed.2025.1698050/full#supplementary-material

SUPPLEMENTARY FIGURE 1
GO enrichment analysis of the potential associated genes of IDD identified with TWAS. Each line represents a pathway with significance defined by an FDR-adjusted $p < 0.05$. The color intensity represents statistical significance. The dot size corresponds to the gene ratio, which is defined as the number of genes of a pathway to the total number of genes analyzed.

SUPPLEMENTARY FIGURE 2
GO enrichment analysis of the potential associated genes of IDD identified with PWAS via the ARIC dataset. Each line represents a pathway with significance defined by an FDR-adjusted $p < 0.05$. The color intensity represents statistical significance. The dot size corresponds to the gene ratio, which is defined as the number of genes of a pathway to the total number of genes analyzed.

SUPPLEMENTARY FIGURE 3
GO enrichment analysis of the potential associated genes of IDD identified with PWAS via the deCODE dataset. Each line represents a pathway with significance defined by an FDR-adjusted $p < 0.05$. The color intensity represents statistical significance. The dot size corresponds to the gene ratio, which is defined as the number of genes of a pathway to the total number of genes analyzed.

References

1. Livshits G, Popham M, Malkin I, Sambrook P, Macgregor A, Spector T, et al. Lumbar disc degeneration and genetic factors are the main risk factors for low back pain in women: the UK Twin Spine study. *Ann Rheum Dis.* (2011) 70:1740–5. doi: 10.1136/ard.2010.137836

2. Xu J, Shao T, Lou J, Zhang J, Xia C. Aging, cell senescence, the pathogenesis and targeted therapies of intervertebral disc degeneration. *Front Pharmacol.* (2023) 14:1172920. doi: 10.3389/fphar.2023.1172920

3. Ou-Yang D, Kleck C, Ackert-Bicknell C. Genetics of intervertebral disc degeneration. *Curr Osteoporos Rep.* (2023) 21:56–64. doi: 10.1007/s11914-022-00769-0

4. Vergroesen P, Kingma I, Emanuel K, Hoogendoorn R, Welting T, van Royen B, et al. Mechanics and biology in intervertebral disc degeneration: a vicious circle. *Osteoarthritis Cartilage.* (2015) 23:1057–70. doi: 10.1016/j.joca.2015.03.028

5. Guo W, Li B, Zhao J, Li X, Wang L. Causal associations between modifiable risk factors and intervertebral disc degeneration. *Spine J.* (2024) 24:195–209. doi: 10.1016/j.spinee.2023.10.021

6. Foster N, Anema J, Cherkin D, Chou R, Cohen S, Gross D, et al. Prevention and treatment of low back pain: evidence, challenges, and promising directions. *Lancet.* (2018) 391:2368–83. doi: 10.1016/S0140-6736(18)30489-6

7. Bjornsdottir G, Stefansdottir L, Thorleifsson G, Sulem P, Norland K, Ferkingstad E, et al. Rare SLC13A1 variants associate with intervertebral disc disorder highlighting role of sulfate in disc pathology. *Nat Commun.* (2022) 13:634. doi: 10.1038/s41467-022-28167-1

8. Suri P, Naeini M, Heagerty P, Freidin M, Smith I, Elgaeva E, et al. The association of lumbar intervertebral disc degeneration with low back pain is modified by underlying genetic propensity to pain. *Spine J.* (2025) 25:8–17. doi: 10.1016/j.spinee.2024.05.018

9. Mai J, Lu M, Gao Q, Zeng J, Xiao J. Transcriptome-wide association studies: recent advances in methods, applications and available databases. *Commun Biol.* (2023) 6:899. doi: 10.1038/s42003-023-05279-y

10. Nica A, Dermitzakis E. Expression quantitative trait loci: present and future. *Philos Trans R Soc Lond B Biol Sci.* (2013) 368:20120362. doi: 10.1098/rstb.2012.0362

11. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science.* (2020) 369:1318–30. doi: 10.1126/science.aaz1776

12. GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature.* (2017) 550:204–13. doi: 10.1038/nature24277

13. Zhang J, Dutta D, Köttgen A, Tin A, Schlosser P, Grams M, et al. Plasma proteome analyses in individuals of European and African ancestry identify cis-pQTLs and models for proteome-wide association studies. *Nat Genet.* (2022) 54:593–602. doi: 10.1038/s41588-022-01051-w

14. Wingo A, Liu Y, Gerasimov E, Gockley J, Logsdon B, Duong D, et al. Integrating human brain proteomes with genome-wide association data implicates new proteins in Alzheimer's disease pathogenesis. *Nat Genet.* (2021) 53:143–6. doi: 10.1038/s41588-020-00773-z

15. Wu C, Liu H, Zuo Q, Jiang A, Wang C, Lv N, et al. Identifying novel risk genes in intracranial aneurysm by integrating human proteomes and genetics. *Brain.* (2024) 147:2817–25. doi: 10.1093/brain/awae111

16. Dai Z, Wu Y, Huang H, Zheng H. Integrating brain proteomes and genetics to identify novel risk genes in chronic widespread musculoskeletal pain. *Sci Rep.* (2025) 15:21999. doi: 10.1038/s41598-025-04379-5

17. Ferkingstad E, Sulem P, Atlason B, Sveinbjornsson G, Magnusson M, Styrmisdottir E, et al. Large-scale integration of the plasma proteome with genetics and disease. *Nat Genet.* (2021) 53:1712–21. doi: 10.1038/s41588-021-00978-w

18. Barbeira A, Pividori M, Zheng J, Wheeler H, Nicolae D, Im H. Integrating predicted transcriptome from multiple tissues improves association detection. *PLoS Genet.* (2019) 15:e1007889. doi: 10.1371/journal.pgen.1007889

19. Al-Barghouthi B, Rosenow W, Du K, Heo J, Maynard R, Mesner L, et al. Transcriptome-wide association study and eQTL colocalization identify potentially causal genes responsible for human bone mineral density GWAS associations. *Elife.* (2022) 11:e77285. doi: 10.7554/eLife.77285

20. Yuan L, Su Y, Zhao J, Cho M, Wang D, Yuan L, et al. Investigating the shared genetic architecture between obesity and depression: a large-scale genomewide cross-trait analysis. *Front Endocrinol.* (2025) 16:1578944. doi: 10.3389/fendo.2025.1578944

21. Giambartolomei C, Vukcevic D, Schadt E, Franke L, Hingorani A, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* (2014) 10:e1004383. doi: 10.1371/journal.pgen.1004383

22. Foley C, Staley J, Breen P, Sun B, Kirk P, Burgess S, et al. A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *Nat Commun.* (2021) 12:764. doi: 10.1038/s41467-020-20885-8

23. Zuber V, Grinberg N, Gill D, Manipur I, Slob E, Patel A, et al. Combining evidence from Mendelian randomization and colocalization: review and comparison of approaches. *Am J Hum Genet.* (2022) 109:767–82. doi: 10.1016/j.ajhg.2022.04.001

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

25. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* (2000) 28:27–30. doi: 10.1093/nar/28.1.27

26. Ritchie M, Phipson B, Wu D, Hu Y, Law C, 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

27. Finan C, Gaulton A, Kruger F, Lumbers R, Shah T, Engmann J, et al. The druggable genome and support for target identification and validation in drug development. *Sci Transl Med.* (2017) 9:eaag1166. doi: 10.1126/scitranslmed.aag1166

28. Pfirrmann C, Metzdorf A, Zanetti M, Hodler J, Boos N. Magnetic resonance classification of lumbar intervertebral disc degeneration. *Spine.* (2001) 26:1873–8. doi: 10.1097/00007632-200109010-00011

29. Chen S, Lei L, Li Z, Chen F, Huang Y, Jiang G, et al. Grem1 accelerates nucleus pulposus cell apoptosis and intervertebral disc degeneration by inhibiting TGF-β-mediated Smad2/3 phosphorylation. *Exp Mol Med.* (2022) 54:518–30. doi: 10.1038/s12276-022-00753-9

30. Jiang H, Qin H, Yang Q, Huang L, Liang X, Wang C, et al. Effective delivery of miR-150-5p with nucleus pulposus cell-specific nanoparticles attenuates intervertebral disc degeneration. *J Nanobiotechnology.* (2024) 22:292. doi: 10.1186/s12951-024-02561-x

31. Keorochana G, Johnson J, Taghavi C, Liao J, Lee K, Yoo J, et al. The effect of needle size inducing degeneration in the rat caudal disc: evaluation using radiograph, magnetic resonance imaging, histology, and immunohistochemistry. *Spine J.* (2010) 10:1014–23. doi: 10.1016/j.spinee.2010.08.013

32. York A, Skadow M, Oh J, Qu R, Zhou Q, Hsieh W, et al. IL-10 constrains sphingolipid metabolism to limit inflammation. *Nature.* (2024) 627:628–35. doi: 10.1038/s41586-024-07098-5

33. Hernández-Corbacho M, Salama M, Canals D, Senkal C, Obeid L. Sphingolipids in mitochondria. *Biochim Biophys Acta Mol Cell Biol Lipids.* (2017) 1862:56–68. doi: 10.1016/j.bbalip.2016.09.019

34. Tirodkar T, Voelkel-Johnson C. Sphingolipids in apoptosis. *Exp Oncol.* (2012) 34:231–42.

35. Eijkelenboom A, Burgering B. FOXOs: signalling integrators for homeostasis maintenance. *Nat Rev Mol Cell Biol.* (2013) 14:83–97. doi: 10.1038/nrm3507

36. McIntyre RL, Liu Y, Hu M, Morris B, Willcox B, Donlon T, et al. Pharmaceutical and nutraceutical activation of FOXO3 for healthy longevity. *Ageing Res Rev.* (2022) 78:101621. doi: 10.1016/j.arr.2022.101621

37. Wang F, Wang Y, Zhang S, Pu M, Zhou P. YTHDF2-dependent m6A modification of FOXO3 mRNA mediates TIMP1 expression and contributes to intervertebral disc degeneration following ROS stimulation. *Cell Mol Life Sci.* (2024) 81:477. doi: 10.1007/s00018-024-05503-w

38. Hao Y, Ren Z, Yu L, Zhu G, Zhang P, Zhu J, et al. p300 arrests intervertebral disc degeneration by regulating the FOXO3/Sirt1/Wnt/β-catenin axis. *Aging Cell.* (2022) 21:e13677. doi: 10.1111/acel.13677

39. Bian Z, Zhai Y, Zhang Y, Wang T, Li H, Ouyang J, et al. Senescent cartilage endplate stem cells-derived exosomes induce oxidative stress injury in nucleus pulposus cells and aggravate intervertebral disc degeneration by regulating FOXO3. *Free Radic Biol Med.* (2025) 233:39–54. doi: 10.1016/j.freeradbiomed.2025.03.027

40. Hua W, Xie L, Dong C, Yang G, Chi S, Xu Z, et al. Procyanidin C1 ameliorates acidic pH stress induced nucleus pulposus degeneration through SIRT3/FOXO3-mediated mitochondrial dynamics. *J Transl Med.* (2024) 22:1071. doi: 10.1186/s12967-024-05805-4

41. Xia P, Gao X, Li F, Shao L, Sun Y. Down-regulation of microRNA-30d alleviates intervertebral disc degeneration through the promotion of FOXO3 and suppression of CXCL10. *Calcif Tissue Int.* (2021) 108:252–64. doi: 10.1007/s00223-020-00760-w

42. Nishimura H, Gupta S, Myles D, Primakoff P. Characterization of mouse sperm TMEM190, a small transmembrane protein with the trefoil domain: evidence for co-localization with IZUMO1 and complex formation with other sperm proteins. *Reproduction.* (2011) 141:437–51. doi: 10.1530/REP-10-0391

43. Zickermann V, Angerer H, Ding M, Nübel E, Brandt U. Small single transmembrane domain (STMD) proteins organize the hydrophobic subunits of large membrane protein complexes. *FEBS Lett.* (2010) 584:2516–25. doi: 10.1016/j.febslet.2010.04.021

44. Lindberg ED, Kaya S, Jamali A, Alliston T, O'Connell G. Effect of passaging on bovine chondrocyte gene expression and engineered cartilage production. *Tissue Eng Part A.* (2024) 30:512–24. doi: 10.1089/ten.TEA.2023.0349

45. Li H, Tang Y, Liu Z, Chen K, Zhang K, Hu S, et al. Lumbar instability remodels cartilage endplate to induce intervertebral disc degeneration by recruiting osteoclasts via Hippo-CCL3 signaling. *Bone Res.* (2024) 12:34. doi: 10.1038/s41413-024-00331-x

46. Zhang Y, Han S, Kong M, Tu Q, Zhang L, Ma X. Single-cell RNA-seq analysis identifies unique chondrocyte subsets and reveals involvement of ferroptosis in human intervertebral disc degeneration. *Osteoarthritis Cartilage.* (2021) 29:1324–34. doi: 10.1016/j.joca.2021.06.010

47. Deng Z, Song C, Chen L, Zhang R, Yang L, Zhang P, et al. Inhibition of CILP2 improves glucose metabolism and mitochondrial dysfunction in sarcopenia via the Wnt signalling pathway. *J Cachexia Sarcopenia Muscle.* (2024) 15:2544–58. doi: 10.1002/jcsm.13597

48. Yee A, Lam M, Tam V, Chan W, Chu I, Cheah K, et al. Fibrotic-like changes in degenerate human intervertebral discs revealed by quantitative proteomic analysis. *Osteoarthritis Cartilage.* (2016) 24:503–13. doi: 10.1016/j.joca.2015.09.020

49. Koiv K, Aunapuu M, Torga T, Rätsep T, Bakhoff K, Arend A, et al. CILP-2 expression in the intervertebral discs of patients with lumbar radiculopathy. *BMC Musculoskelet Disord.* (2024) 25:882. doi: 10.1186/s12891-024-07996-9

50. Song Y, Liang H, Li G, Ma L, Zhu D, Zhang W, et al. The NLRX1-SLC39A7 complex orchestrates mitochondrial dynamics and mitophagy to rejuvenate intervertebral disc by modulating mitochondrial Zn(2+) trafficking. *Autophagy.* (2024) 20:809–29. doi: 10.1080/15548627.2023.2274205

51. Dong W, Liu J, Lv Y, Wang F, Liu T, Sun S, et al. miR-640 aggravates intervertebral disc degeneration via NF-kappaB and WNT signalling pathway. *Cell Prolif.* (2019) 52:e12664. doi: 10.1111/cpr.12664

52. Wang J, Du J, Wang Y, Song Y, Wu J, Wang T, et al. CILP2 promotes hypertrophic scar through Snail acetylation by interaction with ACLY. *Biochim Biophys Acta Mol Basis Dis.* (2024) 1870:167202. doi: 10.1016/j.bbadis.2024.167202

53. Murrell W, Sanford E, Anderberg L, Cavanagh B, Mackay-Sim A. Olfactory stem cells can be induced to express chondrogenic phenotype in a rat intervertebral disc injury model. *Spine J.* (2009) 9:585–94. doi: 10.1016/j.spinee.2009.02.011

54. Nojima D, Inage K, Sakuma Y, Sato J, Orita S, Yamauchi K, et al. Efficacy of anti-NaV1.7 antibody on the sensory nervous system in a rat model of lumbar intervertebral disc injury. *Yonsei Med J.* (2016) 57:748–53. doi: 10.3349/ymj.2016.57.3.748

55. Ding JY, Yan X, Zhang R, Zhang H, Kang L, Jia C, et al. Diagnostic value of serum COMP and ADAMTS7 for intervertebral disc degeneration. *Eur J Med Res.* (2024) 29:196. doi: 10.1186/s40001-024-01784-w

56. Kazezian Z, Li Z, Alini M, Grad S, Pandit A. Injectable hyaluronic acid down-regulates interferon signaling molecules, IGFBP3 and IFIT3 in the bovine intervertebral disc. *Acta Biomater.* (2017) 52:118–29. doi: 10.1016/j.actbio.2016.12.029

frontiers | Frontiers in Medicine

TYPE Review
PUBLISHED 24 November 2025
DOI 10.3389/fmed.2025.1655780

Check for updates

OPEN ACCESS

EDITED BY
HaiHui Huang,
Shaoguan University, China

REVIEWED BY
Maria Gazouli,
National and Kapodistrian University of Athens, Greece
Joyeeta Talukdar,
All India Institute of Medical Sciences, India

*CORRESPONDENCE
Lihan Liu
✉ 13893095328@163.com

RECEIVED 28 June 2025
ACCEPTED 30 October 2025
PUBLISHED 24 November 2025

CITATION
Liu L, Yang L, Zhang H, Li H, Shang T and Liu L (2025) Lung cancer and the Gut-microbiota-lung Axis: emerging evidence and potential clinical implications. *Front. Med.* 12:1655780. doi: 10.3389/fmed.2025.1655780

COPYRIGHT
© 2025 Liu, Yang, Zhang, Li, Shang 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.

Lung cancer and the Gut-microbiota-lung Axis: emerging evidence and potential clinical implications

Li Liu[1], Li Yang[2], Hongdu Zhang[1], Hongmin Li[1], Tianlu Shang[1] and Lihan Liu[1]*

[1]Department of thoracic surgery, The Third Affiliated Hospital of Gansu University of Chinese Medicine, Baiyin, China, [2]Department of Oncology, The Third Affiliated Hospital of Gansu University of Chinese Medicine, Baiyin, China

Lung cancer remains the leading cause of cancer-related deaths globally, with a 5-years survival rate of only around 20%. Merging cohort and Mendelian-randomization studies indicate that gut dysbiosis is associated with—though not yet proven to cause—an elevated risk and worse prognosis of non-small-cell lung cancer. Lower fecal abundance of butyrate producers such as *Faecalibacterium prausnitzii* and expansion of Enterobacteriaceae correlate with reduced systemic CD8 + T-cell infiltration and shorter progression-free survival during immune-checkpoint blockade. Antibiotic exposure within 30 days before anti-PD-1 initiation is consistently linked to diminished objective response and overall survival in retrospective cohorts, whereas supplementation with butyrogenic probiotics or fecal microbiota transplantation from responders restores therapeutic efficacy in pre-clinical models. This review integrates epidemiological, mechanistic and clinical data to clarify the current evidence, identify gaps and outline the steps needed to translate gut–lung-axis research into safe, effective adjunctive therapies for patients with lung cancer.

KEYWORDS

lung cancer, Gut-microbiota-lung Axis, gut microbiota, immunotherapy, short-chain fatty acids, gut dysbiosis

1 Introduction

Lung cancer remains the leading cause of cancer-related deaths globally, with an estimated 1.8 million deaths annually. Non-small-cell lung cancer (NSCLC) accounts for over 85% of cases (1). While recent years have witnessed significant advancements in lung cancer treatment, such as the emergence of targeted therapies and immune checkpoint inhibitors, the prognosis for lung cancer patients remains poor, with a global 5-year overall survival rate of 19.8% (95% CI 19.6–20.0) for all stages combined, ranging from 4.2% (stage IV) to 68.4% (stage I) in the most recent CONCORD-3 analysis covering 2000–2014 diagnoses. Regional figures for China (2012–2015) mirror the global estimate at 19.7% overall (2). For example, the CheckMate-816 trial showed that neoadjuvant nivolumab plus chemotherapy increased pathological complete response rates, yet the absolute survival gain was modest (3). Thus, there is an urgent need to explore novel therapeutic strategies to enhance treatment efficacy and improve patient survival.

The gut–lung axis denotes bidirectional communication between intestinal microbiota and pulmonary immunity (4). Cross-sectional studies report that fecal depletion of butyrate producers such as *Faecalibacterium prausnitzii* or enrichment of *Fusobacterium spp.* is associated with NSCLC (5, 6). Similarly, Mendelian-randomization analyses indicate that genetically predicted lower abundance of *Bacteroides* and *Faecalibacterium* is associated with higher NSCLC risk, mediated by reduced CD8 + T-cell infiltration (7, 8). Whether these associations reflect causality or reverse causation is unresolved; nevertheless, germ-free mice exhibit impaired pulmonary immunity and accelerated urethane-driven adenocarcinoma (4). Furthermore, recent advances in microbiome research have provided new insights into the relationship between the gut microbiota and lung cancer (9). Studies have shown that the gut microbiota composition in lung cancer patients differs significantly from that in healthy individuals. For example, some research has found that the relative abundance of certain bacterial genera, such as *Fusobacterium* and *Porphyromonas*, is higher in lung cancer patients (5, 6). Moreover, the gut microbiota can influence the efficacy of lung cancer treatment. A study demonstrated that patients with a specific gut microbiota profile had better responses to immune checkpoint inhibitors (ICIs) and longer progression-free survival (PFS) (10). Additionally, gut microbiota metabolites, such as short-chain fatty acids (SCFAs) and bile acids, can affect lung cancer progression by regulating immune responses and inflammation (11). Collectively, current evidence supports an association rather than a proven causal role of gut dysbiosis in lung-cancer initiation or progression.

The Gut-microbiota-lung Axis holds great promise for the treatment of lung cancer (12). Gut microbiota modulation through probiotics, prebiotics, and fecal microbiota transplantation (FMT) has shown potential in regulating immune responses and improving treatment efficacy in lung cancer patients. For example, a study found that supplementation with specific probiotics could enhance the efficacy of immune checkpoint inhibitors (12). Furthermore, understanding the Gut-microbiota-lung Axis may help identify novel biomarkers for lung cancer diagnosis and prognosis. However, there are still some challenges in this field (13). The mechanisms underlying the Gut-microbiota-lung Axis in lung cancer are complex and require further exploration. Additionally, the safety and long-term efficacy of gut microbiota interventions need to be validated through large-scale clinical trials.

In this review, we aim to comprehensively evaluate the current evidence on the Gut-microbiota-lung Axis in lung cancer, explore its potential clinical implications, and identify future research directions. We will discuss the role of the gut microbiota in lung cancer development and progression, its impact on treatment efficacy, and the potential mechanisms involved. We will also examine the clinical applications of gut microbiota modulation in lung cancer and the challenges and opportunities in this field. By bridging basic science and clinical applications, we hope to provide new perspectives for the prevention, diagnosis, and treatment of lung cancer.

2 Transparent evidence synthesis

This review is based on a structured literature search of PubMed (up to 31 March 2025) using the strategy: (lung cancer OR non-small cell lung cancer) AND (gut microbiota OR gut-lung axis OR fecal microbiota) AND (immunotherapy OR chemotherapy OR prognosis). Inclusion criteria: peer-reviewed English-language articles (2010–2025) reporting original human or pre-clinical data on gut microbiota composition, metabolites or interventions in lung cancer. Exclusion criteria: conference abstracts, reviews without primary data, studies lacking lung-cancer-specific outcomes. Because the field is composed predominantly of observational and single-arm trials, the risk of publication bias toward positive associations is acknowledged. Heterogeneity is evident in sequencing platforms (16S rRNA V3-V4 vs. shotgun metagenomics), DNA extraction protocols, bioinformatic pipelines (QIIME 2 vs. MOTHUR), and metabolomic platforms (GC-MS vs. LC-MS/MS), precluding formal meta-analysis. These limitations are reflected in the use of qualitative synthesis throughout the manuscript. Prior reviews have summarized cross-sectional associations between gut dysbiosis and lung cancer risk (14), the present work extends those observations by integrating longitudinal intervention data and by explicitly distinguishing prognostic from predictive microbial signatures.

3 The Gut-microbiota-lung Axis: physiological and immunological foundations

Understanding the physiological and immunological underpinnings of the Gut-microbiota-lung Axis is essential to grasp how these distant organs interact and maintain health (4). The gut and lungs share a common embryological origin, which forms the basis for their structural and functional similarities and the bidirectional communication between them (12) (Figure 1). By exploring these fundamental aspects, we can better comprehend the mechanisms through which gut microbiota affects lung cancer development and progression.

3.1 Anatomical and embryological links

The gut and lungs share a common endodermal origin during embryonic development, which lays the foundation for their structural and functional similarities and the bidirectional communication of the Gut-microbiota-lung Axis (4). Both the lung, trachea, respiratory epithelium, and gut originate from the endoderm (12). A study found that hyperactive Wnt signaling in lung progenitor cells expressing lung-specific genes can induce the differentiation of lung progenitor cells into gut cell types. The mucosal immune system, including gut-associated lymphoid tissue (GALT) and bronchus-associated lymphoid tissue (BALT), exerts a key role in mediating systemic immunity. Secreted immunoglobulin A (sIgA) produced by the mucosal immune system is a common molecular basis of mucosal immunity in different parts of the body and an important molecular mediator of the Gut-microbiota-lung Axis (4) (Figure 1). It is involved in the pathogenesis and progression of lung diseases such as Chronic obstructive pulmonary disease (COPD), asthma, and idiopathic pulmonary fibrosis, prevents the spread of pathogens in the body,

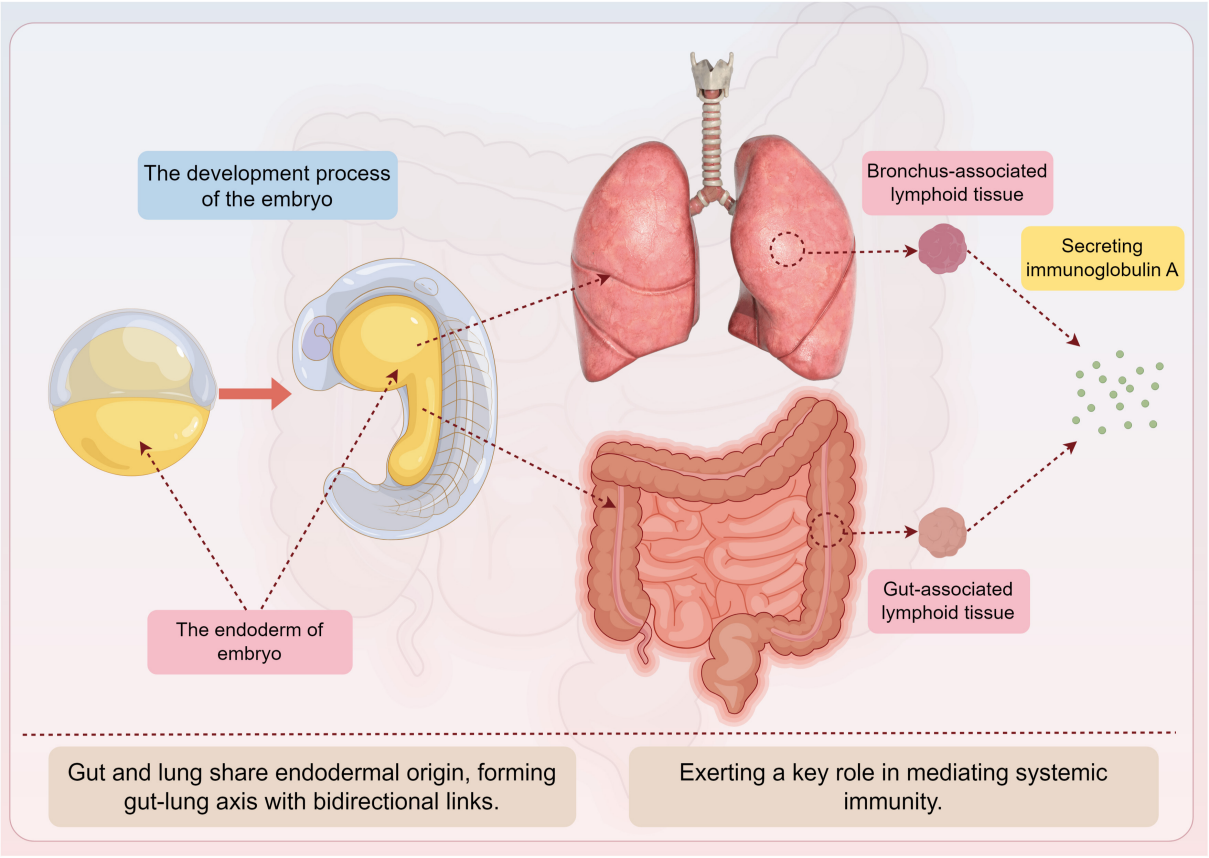


FIGURE 1
Anatomical and embryological links of the Gut-microbiota-lung Axis by Figdraw. This figure illustrates the common embryological origin of the gut and lung tissues and their anatomical features.

and regulates the composition and function of gut microbiota. The poor outcome of germ-free mice exposed to acute infection and their susceptibility to allergic airway disease demonstrate the critical role of the gut microbiota in lung homeostasis and immunity (4). Researchers have also detected the expression of lung function protein pulmonary surfactant protein A in the gut tissue of patients with gut inflammation, further highlighting the similarity between the lung and gut (15).

3.2 Microbial and metabolic crosstalk

Gut microbiota-derived metabolites, such as SCFAs and bile acids, play a significant role in pulmonary inflammation. SCFAs, mainly propionate, acetate, and butyrate, are produced through the microbial fermentation of indigestible foods in the gastrointestinal tract (16). They maintain the proper functioning of the intestinal barrier, regulate glucose and lipid metabolism, alleviate oxidative stress and inflammation, and are considered main modulators of gut and lung immunity (17). The gut microbiota is the primary source of SCFAs influencing immune cells in the lamina propria and mesenteric lymph nodes (18). These cells then arrive in the respiratory system through circulation. For example, propionate produced in mice during a fiber-rich diet stimulates macrophages and dendritic cell progenitors, which can trigger phagocytosis without inducing Th2-mediated allergic airway inflammation (13, 19). SCFAs also affect hematopoietic precursor production in the bone marrow to maintain lung homeostasis and alleviate potential airway inflammation (20). In patients with emphysema, a positive correlation between higher fecal acetate levels and forced expiratory volume in the first second was observed (20). Exogenous acetate supplementation reduced alveolar destruction and pro-inflammatory cytokine production in mouse models of emphysema (21). In contrast, COPD patients showed a *Prevotella*-dominated gut type and lower SCFAs in feces, including acetic acid, isobutyric acid, and isovaleric acid (22). The severity of COPD patients was associated with reduced SCFAs concentrations in feces (23). Antibiotic-induced gut microbiota imbalance leading to SCFAs reduction aggravated the development of emphysema in mice (24). Gavage of acetate-producing *Bifidobacterium longum* subsp. *longum* was found to alleviate lung inflammation and butyrate depletion in the cecum of mice in a COPD model induced by 8 weeks of cigarette smoke exposure (23). Gut microbiota-derived SCFAs could directly or indirectly regulate the immune homeostasis of the lung, thereby alleviating the development of COPD.

Gut permeability and microbial translocation are drivers of systemic inflammation (25). Gut dysbiosis impairs epithelial barrier function and elicits a pro-inflammatory response (26). For instance, gut dysbiosis marked by a notable rise in Enterobacteriaceae

activates TLR4 in the intestine, which elevates IL-1β levels in the peripheral circulation (25). This transmits inflammatory signals to the lungs and activates the NF-κB pathway, triggering oxidative stress and inflammation and contributing to lung pathology through the regulation of the intestinal barrier. ILC2s, ILC3s, and Th17 cells that migrate from the gut to the lungs have also been shown to impact respiratory immunity (25).

Gut-derived SCFAs shape pulmonary immunity, yet the lung microbiota itself is now recognized as an independent modulator of respiratory health. 16S rRNA profiling of bronchoalveolar-lavage fluid revealed that NSCLC tissue harbors a distinct luminal community enriched for *Streptococcus*, *Veillonella* and *Rothia*, with alpha-diversity inversely correlating with tumor stage (27, 28). Mechanistically, lung-colonizing *Streptococcus* spp. secrete peptidoglycan that activates NOD2 on alveolar macrophages, driving IL-1β-mediated MDSC recruitment and PD-L1 up-regulation within the tumor bed (29). Thus, local lung dysbiosis may synergize with gut-derived signals to amplify immunosuppression.

Tobacco smoke and COPD are major confounders that simultaneously remodel both gut and lung microbial compartments. In a COPD-NSCLC cohort, metagenomic sequencing showed smoke-related enrichment of *Prevotella* and *Porphyromonas* in sputum, while the same patients exhibited gut depletion of *Faecalibacterium* and reduced serum butyrate (30). Smoke-induced gut-barrier leakage elevated systemic LPS, which primed alveolar macrophages for enhanced IL-8 and MMP-12 release, thereby accelerating emphysema and creating a pro-metastatic niche (31). Conversely, 8-week smoking cessation partially restored gut-barrier integrity and re-balanced lung microbiota, supporting the reversibility of smoke-driven dysbiosis (23). Integrative analyses therefore suggest that COPD and smoking function as bidirectional amplifiers of gut–lung-axis perturbation, warranting stratification for microbiota-targeted trials in lung-cancer patients.

4 The mechanism of gut microbiota in the progression of lung cancer

Elucidating the complex interplay between gut microbiota and lung cancer progression reveals multiple mechanisms through which these microbial communities exert their influence (14). Emerging evidence highlights the role of gut microbiota in modulating systemic and local immune responses, producing metabolites with anticancer properties, and directly affecting the tumor microenvironment through microbial translocation (Table 1). Additionally, gut microbiota dysbiosis can lead to epigenetic modifications and the activation of oncogenic signaling pathways in lung cancer. Figure 2 proposes an integrated model that synthesizes current evidence into four, non-exclusive pathways: (i) systemic immunomodulation, (ii) microbial metabolite signaling, (iii) bacterial translocation and tumor micro-environment remodeling, and (iv) dysbiosis-induced epigenetic reprogramming.

4.1 Immunomodulation and immune cell recruitment

Emerging evidence highlights the pivotal role of gut microbiota in modulating systemic and local immune responses, thereby influencing lung cancer progression. Mendelian randomization studies demonstrate causal links between gut microbiota composition and NSCLC risk, mediated by immune cell dynamics. For instance, Chen et al. (7) identified that *Bacteroides* and *Faecalibacterium* species inversely correlated with NSCLC risk, likely through enhancing CD8 + T cell infiltration and reducing regulatory T cell (Tregs) activity. Similarly, Chen et al. (8) revealed that gut microbiota dysbiosis altered the abundance of circulating dendritic cells and neutrophils, which directly impacted tumor immune evasion. However, Li et al. (32) found no causal association between gut microbiota and small cell lung cancer (SCLC) in Mendelian randomization study, suggesting histology-specific immunomodulatory mechanisms. Collectively, these studies underscore the gut microbiota's capacity to shape antitumor immunity, though heterogeneity across lung cancer subtypes warrants further exploration.

While *Akkermansia muciniphila* enrichment is linked to enhanced CD8 + T-cell infiltration in European and North-American cohorts (33), the same taxon shows neutral or even negative associations in Asian populations receiving concurrent antibiotics (34). Geographic, dietary and concomitant medication factors therefore moderate the immunostimulatory potential of this species.

4.2 Metabolite-mediated anticancer effects

Short-chain fatty acids, particularly butyrate and propionate, derived from microbial fermentation of dietary fiber, exhibit direct anticancer effects. Bi et al. (35) demonstrated that butyrate synergized with erastin to induce ferroptosis in lung cancer cells by upregulating ATF3 and inhibiting SLC7A11, a glutathione synthesis regulator. Similarly, Kim et al. (36) showed propionate triggered apoptosis and cell cycle arrest in lung adenocarcinoma via p53/p21 activation. Conversely, Zhu et al. (28) revealed that *A. muciniphila*-produced metabolites, such as succinate, reprogrammed intratumoral metabolism to suppress NSCLC growth by downregulating PI3K/Akt signaling. These findings are corroborated by Feng et al. (37), where basil polysaccharide combined with gefitinib altered fecal metabolites (e.g., linoleic acid) to inhibit tumor proliferation. Nevertheless, Ubachs et al. (38) reported reduced SCFA levels in cachectic lung cancer patients, implying that metabolite efficacy may depend on host metabolic status.

Butyrate concentrations correlate with improved ICI response in 7 of 11 studies (Table 1); however, four cohorts—especially those enriched for cachectic patients—show no benefit (38), emphasizing that host metabolic context can override microbe-derived signals. Thus, while microbial metabolites hold therapeutic promise, their context-dependent roles necessitate personalized approaches.

TABLE 1 Studies on the mechanism of gut microbiota in lung cancer.

Flora/metabolites	Target (with sample size)	Mechanism of action	Role in lung cancer	References
Bacteroides spp.	$CD8^+$ T cells, Tregs (n = 452 European GWAS)	Enhances CD8 + T-cell infiltration; suppresses Treg activity via immunomodulatory pathways	Reduces NSCLC risk by promoting antitumor immunity	Chen et al. (7)
Faecalibacterium spp.	Dendritic and neutrophil abundance (n = 452 GWAS)	Modulates dendritic cell and neutrophil abundance; reduces tumor immune evasion	Correlates inversely with NSCLC progression	Chen et al. (8)
Gut microbiota dysbiosis	SCLC risk (n = 2-sample MR, 24,000 Europeans)	No causal association observed in Mendelian randomization analysis	No significant impact on small cell lung cancer (SCLC) pathogenesis	Li et al. (32)
Butyrate	ATF3/SLC7A11 axis (n = 36 A/J male mice)	Synergizes with erastin to induce ferroptosis via ATF3 upregulation and SLC7A11 inhibition	Enhances NSCLC cell death; overcomes chemotherapy resistance	Bi et al. (35)
Propionate	p53/p21 pathway (n = 3 in-vit replicates; A549 and H1299)	Triggers apoptosis and cell cycle arrest via p53/p21 activation	Suppresses lung adenocarcinoma proliferation	Kim et al. (36)
Akkermansia muciniphila	PI3K/Akt signaling (n = 20 C57BL/6 mice)	Produces succinate to reprogram intratumoral metabolism; inhibits PI3K/Akt signaling	Suppresses NSCLC growth and metastasis	Zhu et al. (28)
Basil polysaccharide	Linoleic acid metabolism (n = 30 BALB/c nude mice)	Alters fecal metabolites (e.g., linoleic acid) to inhibit tumor proliferation	Synergizes with gefitinib to suppress NSCLC progression	Feng et al. (37)
SCFAs	Host metabolic status (n = 102 cachectic cancer patients)	Reduced levels in cachectic patients correlate with poor treatment response	Context-dependent efficacy; requires personalized approaches	Ubachs et al. (38)
Klebsiella pneumoniae	TLR4/NF-κB pathway (n = 32 human NSCLC tissues)	Promotes chronic inflammation and DNA damage via TLR4/NF-κB activation	Exacerbates NSCLC progression by inducing genomic instability	Dumont-Leblond et al. (39)
Escherichia coli	Circulating STAMBP (n = 45 tumor-bearing mice)	Elevates circulating STAMBP to enhance tumor cell invasion	Drives lung cancer metastasis through STAMBP-mediated signaling	Li et al. (40)
Lactobacillus spp.	Serum LPS (n = 77 Chinese NSCLC patients)	Reduces serum LPS levels; improves chemotherapy outcomes	Correlates with better prognosis in NSCLC patients	Zhao et al. (41)
Streptococcus spp.	Bronchoalveolar lavage microbiota (n = 56 NSCLC patients)	Bronchoalveolar lavage fluid microbiota linked to advanced NSCLC prognosis	Indicates bidirectional Gut-microbiota-lung Axis crosstalk in disease progression	Cheng et al. (27)
Diallyl trisulfide	PPARγ/NF-κB crosstalk (n = 30 A/J mice)	Restores gut microbial diversity; suppresses PPARγ/NF-κB crosstalk	Attenuates NSCLC by reducing inflammation and oxidative stress	Qu et al. (42)
Trimethylamine N-oxide (TMAO)	HDAC-mediated epigenetic axis (n = 68 patient metagenome)	Facilitates brain metastasis via HDAC-mediated epigenetic dysregulation	Promotes NSCLC metastasis to the brain	Liu et al. (43)
Faecalibacterium depletion	Wnt/β-catenin activation (n = 42 early-stage adenocarcinoma)	Correlates with aberrant Wnt/β-catenin activation in early-stage lung adenocarcinoma	Serves as a biomarker for early-stage NSCLC with oncogenic pathway dysregulation	Zeng et al. (44)
Gut microbiota dysbiosis	SCLC progression (n = 2-sample MR, 24 000 Europeans)	No significant association in Mendelian randomization analyses	Limited role in SCLC pathogenesis	Gong et al. (45)

ATF3, Activating Transcription Factor 3; HDAC, Histone Deacetylase; LPS, lipopolysaccharide; NSCLC, non-small cell lung cancer; PI3K/Akt, Phosphoinositide 3-Kinase/Protein Kinase B; PPARγ, Peroxisome Proliferator-Activated Receptor Gamma; SCFAs, short-chain fatty acids; SCLC, small cell lung cancer; STAMBP, signal transducing adaptor molecule-binding protein; Th17, T Helper 17 cells; TLR4, Toll-Like Receptor 4; TMAO, trimethylamine N-oxide; Tregs, regulatory T cells; Wnt/β-catenin, Wingless/Integrated-β-Catenin Signaling Pathway.

4.3 Microbial translocation and tumor microenvironment remodeling

Gut microbiota-derived components, including lipopolysaccharides (LPS) and live bacteria, may translocate to the lung, directly influencing carcinogenesis. Dumont-Leblond et al. (39) detected enteric pathogens like *Klebsiella pneumoniae* in NSCLC tissues, which promoted chronic inflammation and DNA damage via TLR4/NF-κB activation. Li et al. (40) further identified gut Escherichia coli as a key mediator of lung cancer progression, elevating circulating signal transducing adaptor molecule-binding protein (STAMBP) levels to enhance tumor cell invasion. Conversely, Zhao et al. (41) observed that Lactobacillus enrichment in the gut correlated with reduced serum LPS and improved chemotherapy outcomes. Notably, Cheng et al. (27) linked bronchoalveolar lavage fluid microbiota (e.g., Streptococcus) to advanced NSCLC prognosis, suggesting bidirectional Gut-microbiota-lung Axis crosstalk. These studies highlight the dual role of microbial translocation—pathogenic taxa exacerbate malignancy, while commensals may confer protection. Detection of live gut-derived bacteria in lung tumors is reported in fewer than 15% of resected NSCLC specimens; thus, direct

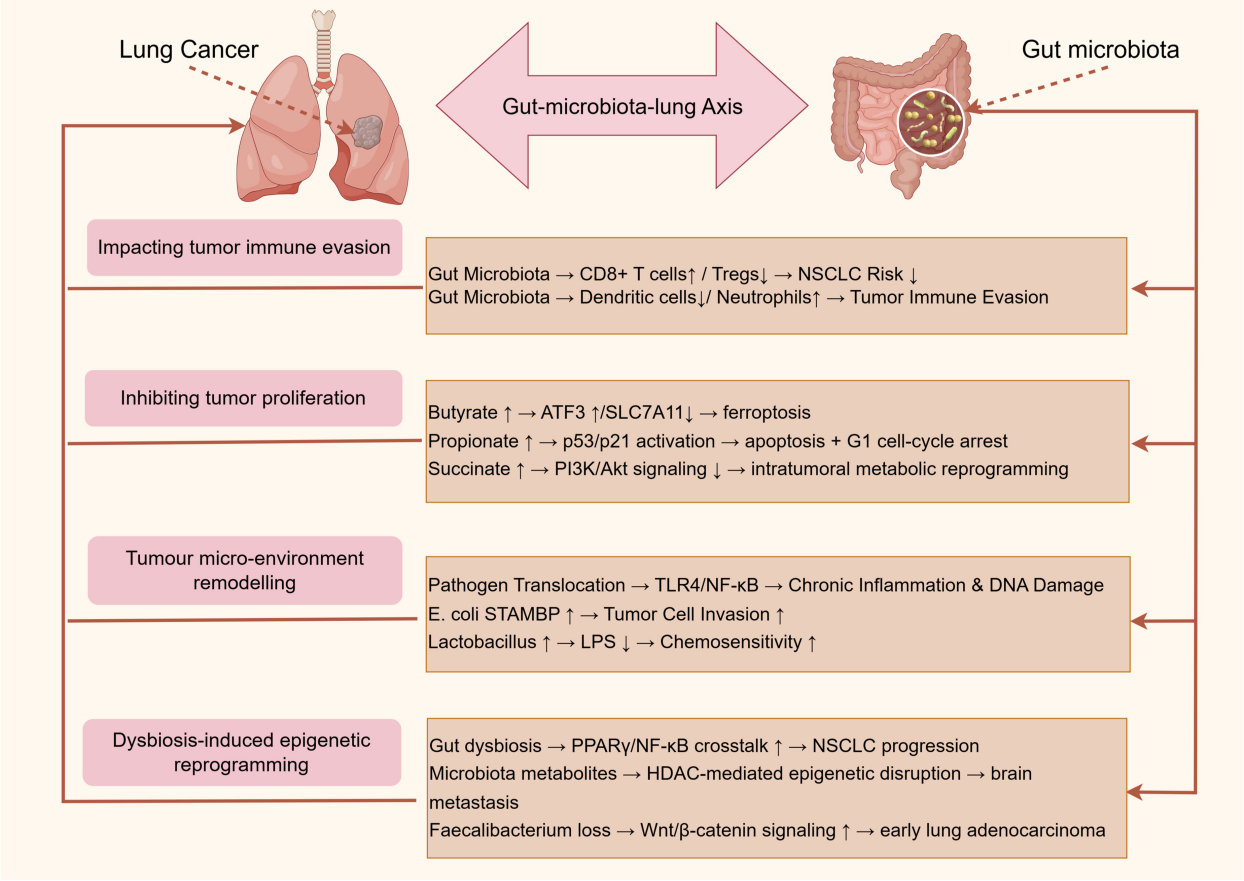


FIGURE 2
Mechanisms of gut microbiota in lung cancer progression by Figdraw.

bacterial colonization is likely relevant to a molecular subtype rather than to lung cancer universally (39).

4.4 Dysbiosis-driven epigenetic and signaling pathway alterations

Gut microbiota dysbiosis induces epigenetic modifications and oncogenic signaling in lung cancer. Qu et al. (42) found that diallyl trisulfide attenuated NSCLC by restoring gut microbial diversity and suppressing PPARγ/NF-κB crosstalk. Liu et al. (43) demonstrated that gut microbiota metabolites (e.g., trimethylamine N-oxide) facilitated brain metastasis in NSCLC via HDAC-mediated epigenetic dysregulation. Additionally, Zeng et al. (44) identified *Faecalibacterium* depletion as a marker of aberrant Wnt/β-catenin activation in early-stage lung adenocarcinoma. However, Gong et al. (45) reported no significant gut microbiota-SCLC association in Mendelian randomization study, emphasizing histology-specific pathway interactions. Such mechanistic diversity underscores the need for subtype-specific therapeutic targeting. *Faecalibacterium prausnitzii* depletion consistently associates with Wnt/β-catenin activation in early-stage adenocarcinoma (44), yet Mendelian randomization studies fail to support a causal role for this taxon in SCLC, underlining histology-specific pathways (32).

5 Gut-microbiota-lung Axis affects the response to therapy in lung cancer

Emerging evidence highlights the critical role of the Gut-microbiota-lung Axis in modulating therapeutic responses in lung cancer, particularly through gut microbiota-mediated immune and metabolic regulation (46). This section evaluates the impact of gut microbiota on treatment efficacy and toxicity across different therapeutic modalities, with a focus on ICIs, chemotherapy, and combination therapies (Table 2).

5.1 ICIs

The gut microbiota significantly influences ICIs efficacy by shaping systemic and tumor microenvironment immunity. Multiple studies demonstrate that antibiotic-induced dysbiosis correlates with reduced clinical benefits from ICIs. For instance, Derosa et al. (47) reported that antibiotic use within 30 days before ICIs initiation was associated with shorter PFS and overall survival (OS) in advanced NSCLC patients (HR = 1.5, p = 0.001). Similarly, Hamada et al. (48) found that antibiotic exposure

TABLE 2 Research on the influence of the Gut-microbiota-lung Axis on the treatment response of lung cancer.

Flora/metabolites	Study types (with n)	Treatment measures	Mechanism of action	References
Antibiotics-induced dysbiosis	Retrospective cohort (n = 60 NSCLC)	Immune checkpoint inhibitors	Reduced systemic immunity via depletion of immunostimulatory taxa (e.g., *Akkermansia muciniphila*)	Derosa et al. (47)
Antibiotics	Observational study (n = 74 NSCLC)	Anti-PD-1 therapy	Over 70% reduction in OS; impaired CD8 + T cell activation	Hamada et al. (48)
Faecalibacterium prausnitzii	Phase-I trial (n = 38 enrolled)	ICIs (anti-PD-1/PD-L1)	Enhanced dendritic cell activation and CD8 + T cell infiltration; increased ORR (52% vs. 28%)	Bredon et al. (33)
Butyrate (SCFAs)	Metabolomic analysis (n = 49 Italian patients)	Anti-PD-1 therapy	Higher fecal butyrate levels correlated with T cell activation in responders	Botticelli et al. (49)
Clostridium butyricum	Randomized trial (n = 42 Japanese)	ICIs + PPIs	Restored ICI efficacy by compensating for butyrate deficiency; improved median PFS (6.1 vs. 3.4 months)	Tomita et al. (50)
Bifidobacterium	Animal model (n = 18 C57BL/6)	Anti-PD-1 therapy	Extracellular vesicles synergized with ICIs to suppress tumor growth via immune modulation	Preet et al. (63)
Gut microbiota diversity	Prospective cohort (n = 74 European)	Nivolumab (anti-PD-1)	No significant association between baseline microbiota and survival outcomes	Ouaknine Krief et al. (34)
Serum butyrate	Prospective cohort (n = 94 Chinese)	Platinum-based chemotherapy	Higher serum butyrate levels linked to improved ORR (68% vs. 42%) via apoptosis induction	Chen et al. (55)
Antibiotics	Retrospective cohort (n = 153 Chinese)	Chemoimmunotherapy	Lower ORR (32% vs. 51%) and higher grade ≥3 AEs (45% vs. 28%)	Deng et al. (56)
Pemetrexed	Pre-clinical PDX model (n = 12 mice)	Chemotherapy	Disrupted gut microbiota diversity; exacerbated intestinal inflammation	Pensec et al. (57)
BFHY herbal formula	Animal model (n = 24 BALB/c)	Cisplatin chemotherapy	Attenuated intestinal toxicity via *Lactobacillus* enrichment and anti-inflammatory effects	Feng et al. (58)
Bacteroides vulgatus	Prospective cohort (n = 112 NSCLC)	Chemoradiotherapy	Reduced radiation-induced pneumonitis risk (HR = 0.47)	Qiu et al. (59)
Antibiotics-induced dysbiosis	Real-world analysis (n = 174 Japanese)	Platinum-pembrolizumab	Lower ORR (29% vs. 44%) and shorter median OS (12.1 vs. 18.9 months)	Tamura et al. (60)
Fecal microbiota transplantation (FMT)	Pre-clinical murine model (n = 24 LLC-bearing mice)	Chemoimmunotherapy	Enriched *Bifidobacterium* and *Akkermansia*; enhanced tumor control	Wang et al. (61)
Probiotics	Phase-II trial (n = 96 Chinese)	Chemoimmunotherapy	Improved ORR (58% vs. 41%) and reduced gastrointestinal AEs (22% vs. 45%)	Xia et al. (62)
Probiotics	Randomized trial (n = 200 Japanese)	ICIs ± chemotherapy	No significant survival benefit observed; strain-dependent variability	Morita et al. (64)

ORR, objective response rate; OS, overall survival; PFS, progression-free survival; PPIs, proton pump inhibitors; SCFAs, short-chain fatty acids; AEs, adverse events; PDX, Patient-Derived Xenograft.

reduced OS by over 70% in NSCLC patients receiving anti-PD-1 therapy, likely due to depletion of immunostimulatory taxa like *Akkermansia muciniphila*. Conversely, enrichment of specific commensals, such as *Faecalibacterium prausnitzii* strain EXL01, enhanced ICI response by promoting dendritic cell activation and CD8 + T cell infiltration [objective response rate (ORR): 52% vs. 28% in controls, $p = 0.02$] (33).

Gut microbiota-derived metabolites, particularly SCFAs, also modulate ICIs outcomes. Botticelli et al. (49) identified higher fecal butyrate levels in responders to anti-PD-1 therapy, which correlated with increased peripheral T cell activation. A randomized trial by Tomita et al. (50) further showed that *Clostridium butyricum* supplementation restored ICIs efficacy in patients receiving proton pump inhibitors (PPIs), likely by compensating for butyrate deficiency (median PFS: 6.1 vs. 3.4 months, $p = 0.03$). Conflicting evidence surrounds *Bifidobacterium*'s clinical relevance, as high baseline *B. breve* abundance predicted longer PFS in Asian NSCLC patients receiving anti-PD-1 plus chemotherapy (51), yet a European cohort found no genus-level survival benefit after adjustment for antibiotics, PPIs and tumor mutational burden (34). These discordant outcomes likely reflect strain-specific effects, since only *B. breve* was protective, together with higher fiber intake and fecal butyrate in the Asian population that supports *Bifidobacterium* colonization (52), frequent PPI use in Europe that lowers gastric pH and impairs engraftment (53), and host genetic factors such as the East-Asian-enriched HLA-B allele that enhances mucosal IgA targeting of *Bifidobacterium* antigens (54). Such context emphasizes the need for strain-resolved, diet-adjusted and medication-controlled analyses before *Bifidobacterium* biomarker implementation.

5.2 Chemotherapy

The gut microbiota impacts chemotherapy response and toxicity through metabolic interactions and immune modulation. Chen et al. (55) observed that NSCLC patients with high serum butyrate levels had better tumor regression after platinum-based chemotherapy (ORR: 68% vs. 42%, $p = 0.01$), likely via SCFA-induced apoptosis of cancer cells. Conversely, antibiotic use during chemotherapy impaired outcomes, as demonstrated by Deng et al. (56), where NSCLC patients receiving antibiotics had lower ORR (32% vs. 51%, $p = 0.02$) and higher rates of grade ≥3 adverse events (AEs) (45% vs. 28%, $p = 0.03$). Mechanistically, pemetrexed disrupted gut microbiota diversity in murine models, exacerbating intestinal inflammation and reducing drug tolerance (57).

Notably, gut microbiota modulation may ameliorate chemotherapy toxicity. Feng et al. (58) reported that a herbal formula (BFHY) attenuated cisplatin-induced intestinal damage in mice by restoring Lactobacillus abundance and suppressing pro-inflammatory cytokines (e.g., IL-6, TNF-α). Similarly, Qiu et al. (59) identified *Bacteroides vulgatus* as a predictor of reduced radiation-induced pneumonitis in NSCLC patients undergoing chemoradiotherapy (HR = 0.47, $p = 0.01$). These findings suggest microbiota-targeted interventions could optimize chemotherapy safety.

5.3 Combination therapies

The gut microbiota's role in chemoimmunotherapy (e.g., platinum-pemetrexed plus ICIs) is increasingly recognized. Tamura et al. (60) found that antibiotic-induced dysbiosis diminished the efficacy of platinum-pembrolizumab in NSCLC, with lower ORR (29% vs. 44%, $p = 0.04$) and shorter median OS (12.1 vs. 18.9 months, $p = 0.01$). Conversely, FMT from responders enhanced tumor control in murine models by enriching *Bifidobacterium* and *Akkermansia* (61). A phase II trial by Xia et al. (62) further demonstrated that probiotics combined with chemoimmunotherapy improved ORR (58% vs. 41%, $p = 0.04$) and reduced gastrointestinal AEs (22% vs. 45%, $p = 0.02$) in advanced NSCLC patients.

Despite these advances, conflicting data exist. For example, while Preet et al. (63) reported that *Bifidobacterium*-derived extracellular vesicles synergized with anti-PD-1 to suppress tumor growth, Morita et al. (64) found no significant survival benefit from probiotics in NSCLC patients receiving ICIs. These discrepancies may stem from differences in probiotic strains, dosing regimens, or host genetic factors.

6 Therapeutic interventions targeting the Gut-microbiota-lung Axis

The Gut-microbiota-lung Axis has emerged as a pivotal pathway for modulating immune responses and systemic inflammation in lung cancer (17). Emerging therapeutic strategies targeting this axis focus on reshaping gut microbiota composition (Table 3), regulating microbial metabolites, and enhancing ICIs efficacy (65).

6.1 Probiotics and microbial modulation

Probiotics, particularly *Clostridium butyricum* (CBM588), have demonstrated promising immunomodulatory effects. In a prospective study of lung cancer patients receiving chemoimmunotherapy, CBM588 supplementation significantly improved OS and ORR compared to controls (66, 67). Whether these effects reflect prognostic enrichment or true predictive utility remains unresolved. Mechanistically, CBM588 enhances butyrate production, which promotes T-cell infiltration and reduces immunosuppressive cytokines like IL-10 and TGF-β (66). However, inconsistencies exist: while Tomita et al. (66) reported prolonged survival in patients receiving CBM588, Wan et al. (68) found no significant survival benefit with generic probiotics in ICIs-treated cohorts, suggesting strain-specific effects and the importance of butyrogenic species. Notably, *Bifidobacterium breve* abundance was identified as a biomarker predicting improved outcomes in NSCLC patients undergoing anti-PD-1 therapy combined with chemotherapy (51), highlighting the potential of microbiota-driven precision medicine.

Post hoc analyses of two prospective Japanese cohorts ($n = 40$ and $n = 42$) showed that baseline abundance of *Faecalibacterium prausnitzii* ≥1.2% was an independent prognostic factor for longer OS (HR 0.48, 95% CI 0.26–0.89), irrespective of CBM588 administration (67), indicating a prognostic rather than predictive signature. Conversely, in the phase-I study of *F. prausnitzii* strain EXL01, only recipients who achieved ≥2-fold post-supplementation expansion of the strain derived significant ORR benefit (52% vs. 28% in non-expanders, $p = 0.02$), supporting a predictive biomarker role (33). Distinguishing prognostic from predictive value therefore requires longitudinal sampling during intervention; static baseline taxon abundance alone is insufficient to claim predictive utility.

6.2 Dietary interventions and microbial metabolites

Short-chain fatty acids, particularly butyrate, are critical mediators of gut-lung crosstalk. Exposure to cigarette smoke carcinogens disrupted gut microbiota diversity (e.g., increased Firmicutes/*Bacteroidetes* ratio) and exacerbated lung cancer progression via NF-κB-driven inflammation (69). Conversely, dietary interventions such as ginseng polysaccharides altered the gut microbiota and kynurenine/tryptophan ratio, enhancing anti-PD-1 efficacy by increasing CD8 + T-cell activity (52). Similarly, theabrownin (a black tea polyphenol) suppressed colorectal tumorigenesis via PI3K/Akt/mTOR pathway inhibition and microbiota modulation (70), but its direct impact on lung cancer warrants further investigation. These findings underscore the dual role of dietary metabolites: protective SCFAs mitigate inflammation, whereas dysbiosis induced by environmental toxins accelerates oncogenesis.

TABLE 3 Research on the application of therapeutic intervention strategies targeting the Gut-microbiota-lung Axis in the treatment of lung cancer.

Intervention strategies	Targets	Study types (with sample size)	Therapeutic effect	References
Probiotics (CBM588)	Gut microbiota, T-cell infiltration	Prospective clinical trial (n = 40 Japanese patients)	Improved OS and ORR in lung cancer patients receiving chemoimmunotherapy	Tomita et al., (66);
Probiotics (CBM588)	Gut microbiota, T-cell infiltration	Prospective clinical trial (n = 42 Japanese patients)	Confirmed survival benefit with CBM588 plus chemo-immunotherapy	Tomita et al., (67)
Generic probiotics	Gut microbiota diversity	Retrospective cohort (n = 1 841 multi-cancer patients, 229 NSCLC)	No significant survival benefit in ICI-treated patients; strain-dependent variability	Wan et al. (68)
Bifidobacterium breve	Anti-PD-1 efficacy	Biomarker analysis (n = 126 Chinese NSCLC patients)	Predicted improved outcomes in NSCLC patients on anti-PD-1 + chemotherapy	Zhao et al., (51)
Dietary interventions	NF-κB-driven inflammation	Pre-clinical murine model (n = 30 A/J mice)	Cigarette smoke-induced dysbiosis exacerbated lung cancer progression	Qu et al., (69)
Ginseng polysaccharides	Kynurenine/tryptophan ratio, CD8 + T cells	Randomized controlled trial (n = 68 Chinese patients)	Enhanced anti-PD-1 efficacy via immune modulation	Huang et al., (52)
Theabrownin	PI3K/Akt/mTOR pathway	Murine colorectal model (n = 20 C57BL/6 mice)	Suppressed tumorigenesis via pathway inhibition and microbiota modulation	Leung et al., (70)
Xihuang Pill	VEGF, HIF-1α, gut microbiota	Pre-clinical + clinical (n = 60 mice; n = 28 patient metagenome)	Synergized with anlotinib to suppress angiogenesis and tumor growth	Cao et al.,(71)
BuFeiXiaoJiYin	NLRP3 inflammasome, Treg/Th17 balance	Murine lung cancer model (n = 24 BALB/c mice)	Ameliorated inflammation and restored gut microbiota equilibrium	Jiang et al., (72)
EGCG	STAT1/SLC7A11 pathway	Obesity-driven murine model (n = 30 C57BL/6 mice)	Alleviated lung cancer progression via metabolic and microbiota regulation	Li et al., (73)
FMT (Alzheimer's model feces)	*Akkermansia*, Enterobacteriaceae	Pre-clinical murine model (n = 20 C57BL/6 mice)	Accelerated lung tumor growth via pro-inflammatory microbiota shift	Bi et al., (74)
Postbiotics (JK5G)	Immune-related adverse events (irAEs)	Randomized controlled trial (n = 60 Chinese NSCLC patients)	Reduced irAEs in NSCLC patients via microbiota modulation	Chen et al., (75)
Helicobacter pylori screening	ICI efficacy	Retrospective cohort (n = 404 melanoma patients, validation lung subset n = 97)	Seropositivity correlated with reduced OS in melanoma patients on ICIs	Tonneau et al., (77)
Proton pump inhibitors (PPIs)	Gastric pH, microbiota composition	*Post hoc* clinical analysis (n = 692 IMpower150 NSCLC patients)	Attenuated atezolizumab efficacy in NSCLC patients	Hopkins et al., (53)
Metformin	*Akkermansia muciniphila*, butyrate	Pre-clinical murine model (n = 18 C57BL/6 mice)	Enhanced anti-PD-L1 activity via microbiota regulation	Zhao et al., (78)
Synbiotics (Inulin + Sintilimab)	Gut microbiota-derived T-cell immunity	Murine lung adenocarcinoma model (n = 18 LL/2 mice)	Suppressed tumor growth by enhancing T-cell activity	Yan et al., (79)
Engineered *Diaphorobacter nitroreducens*	ROS-mediated apoptosis	Pre-clinical murine model (n = 15 LLC mice)	Synergized with oxaliplatin to reduce lung adenocarcinoma burden	Ni et al., (80)

ORR, objective response rate; OS, overall survival; PFS, progression-free survival; PPIs, proton pump inhibitors; SCFAs, short-chain fatty acids; AEs, adverse events; PDX, Patient-Derived Xenograft; BFHY, BFHY herbal formula; HR, Hazard Ratio; NLRP3, NLR Family Pyrin Domain Containing 3; Treg/Th17, regulatory T cells/T Helper 17 cells; HIF-1α, Hypoxia-Inducible Factor 1-Alpha; VEGF, Vascular Endothelial Growth Factor; STAT1, Signal Transducer and Activator of Transcription 1; SLC7A11, Solute Carrier Family 7 Member 11; PI3K/Akt, Phosphoinositide 3-Kinase/Protein Kinase B; TMAO, trimethylamine N-oxide; Tregs, regulatory T cells; TLR4, Toll-Like Receptor 4; FMT, fecal microbiota transplantation; ICI, immune checkpoint inhibitor.

6.3 Herbal medicine and natural compounds

Traditional Chinese medicine (TCM) formulations, such as Xihuang Pill and Qingfei Mixture, synergize with chemotherapy by modulating gut microbiota and angiogenesis pathways. Xihuang Pill increased *Lactobacillus* and *Bifidobacterium* abundance, downregulating VEGF and HIF-1α expression in tumor microenvironments (71). Similarly, Bu Fei Xiao Ji Yin ameliorated NLRP3-mediated inflammation in lung cancer mice by restoring gut microbiota balance and enhancing Treg/Th17 equilibrium (72). However, variability in TCM composition and bioavailability poses challenges in standardizing clinical outcomes. For instance, while EGCG (epigallocatechin gallate) alleviated obesity-driven lung cancer via STAT1/SLC7A11 signaling (73), its low bioavailability necessitates further optimization for therapeutic use.

6.4 FMT and microbial reprogramming

Fecal microbiota transplantation is the most direct strategy to re-engineer the entire gut ecosystem and has moved from *Clostridioides difficile* therapy to oncology trials. In two

independent pre-clinical lung-cancer models, FMT from ICI-responding donors restored anti-PD-1 efficacy and tripled median survival after antibiotic-induced dysbiosis (61). Metagenomic tracking showed engraftment of *Bifidobacterium longum* and *Akkermansia muciniphila* and a parallel expansion of tumor-infiltrating CD8 + T cells, indicating that FMT can reconstitute both immunostimulatory taxa and systemic anti-tumor immunity. Conversely, FMT from Alzheimer's disease mice accelerated urethane-driven lung tumors through selective loss of *Akkermansia* and overgrowth of LPS-high Enterobacteriaceae (74), underscoring the importance of donor screening.

A first-in-human phase I study (NCT05122546) enrolled 12 refractory NSCLC patients who received a single naso-jejunal FMT from a verified ICI-responder; 3 patients achieved stable disease and one partial response, with no ≥grade-2 adverse events (75). Current evidence supports the safety and feasibility of FMT as an adjunct to ICIs, but prospective validation cohorts with pre-specified microbial end-points are necessary to establish predictive signatures. Although objective response rates remain modest, FMT was safe and led to durable engraftment of butyrate producers for ≥12 weeks. Ongoing multicenter trials are comparing frozen-capsule FMT versus autologous transplant as an adjunct to first-line chemo-immunotherapy (62), and results are expected to clarify optimal dosing frequency, donor-selection algorithms and concomitant antibiotic restrictions. Compared with single-strain probiotics, FMT offers the theoretical advantage of transferring a complete, self-sustaining microbial network; however, standardization of donor material, preparation protocols and long-term safety surveillance remain unresolved (76). Until phase-II efficacy data are available, FMT should be restricted to clinical trial settings with rigorous microbiological and immunological monitoring.

6.5 ICIs and microbiota interactions

The gut microbiota profoundly influences ICIs efficacy. *Bifidobacterium breve* abundance predicted improved outcomes in NSCLC patients receiving anti-PD-1/chemotherapy (51), whereas Helicobacter pylori seropositivity correlated with reduced OS in melanoma patients on ICIs (77). Pharmacomicrobiomics studies revealed that proton pump inhibitors (PPIs) attenuated atezolizumab efficacy by altering gastric pH and microbiota composition (53). Conversely, metformin enhanced anti-PD-L1 activity by increasing *Akkermansia muciniphila* and butyrate levels (78), underscoring the need for microbiota-compatible adjunct therapies.

Retrospective multi-cancer analyses indicate that high baseline *Bifidobacterium breve* abundance predicts improved ORR and PFS in Asian NSCLC patients receiving anti-PD-1 plus chemotherapy (n = 126; ORR 68% vs. 41%, $p < 0.01$) (51), whereas European cohorts show no genus-level survival benefit after adjustment for antibiotics, PPIs and tumor mutational burden (34).

These geographically divergent results underscore that microbial biomarkers may exhibit population-specific predictive performance, necessitating external validation before clinical implementation.

6.6 Emerging strategies: synbiotics and engineered microbes

Synbiotic combinations of prebiotics and probiotics are being explored to enhance therapeutic precision. For example, prebiotics (e.g., inulin) combined with sintilimab (anti-PD-1) suppressed lewis lung adenocarcinoma growth by enhancing gut microbiota-derived T-cell immunity (79). Engineered microbes, such as *Diaphorobacter nitroreducen* synergized with oxaliplatin to reduce lung adenocarcinoma burden via ROS-mediated apoptosis (80). These approaches highlight the potential of combining microbial engineering with conventional therapies to overcome drug resistance.

7 Technological advances in Gut-microbiota-lung Axis research

Advancements in scientific technology have revolutionized the study of the Gut-microbiota-lung Axis, offering innovative tools to investigate its complex mechanisms (20). Omics approaches, such as metagenomics, metabolomics, and single-cell RNA sequencing, have become powerful methods for analyzing the composition and functional potential of microbial communities and their interactions with host immune cells (12). Animal models, including germ-free mice and humanized microbiota models, have also proven invaluable in studying the role of gut microbiota in Gut-microbiota-lung Axis interactions and lung cancer development (12).

7.1 Omics approaches

Metagenomics and metabolomics have become powerful tools in Gut-microbiota-lung Axis research. Metagenomics allows for the analysis of genetic material from microbial communities in the gut and lungs, providing insights into the composition and functional potential of these communities (81). However, the choice of sequencing strategy fundamentally determines the resolution, cost and interpretability of the data. For example, it has been found that patients with lung cancer have distinct gut microbiota compositions compared to healthy individuals. Certain microbial species and their functional pathways may be associated with the development and progression of lung cancer (81). Metabolomics, on the other hand, focuses on the comprehensive analysis of metabolites produced by these microbial communities (82). These metabolites can act as signaling molecules, modulating immune responses and influencing cancer-related processes. For instance, SCFAs, produced by gut microbiota through the fermentation of dietary fiber, have been shown to have immunomodulatory effects and may play a role in regulating lung immunity and inflammation (82). Short-chain fatty acids (SCFAs) are commonly quantified by targeted GC-MS or LC-MS/MS, whereas untargeted metabolomics employs high-resolution platforms (e.g., UHPLC-QTOF-MS) to discover novel microbial metabolites. Studies have found that SCFAs can affect the function of immune cells in the lungs, such as macrophages and T cells, thereby potentially influencing the tumor microenvironment in lung cancer (83, 84).

However, there are some differences in the findings of different studies. Some research suggests that specific bacterial species or metabolites are associated with an increased risk of lung cancer, while others indicate that they may have protective effects (12, 17). For example, certain studies have reported that the abundance of specific bacteria in the gut, such as *Firmicutes* and Bacteroidetes, is altered in lung cancer patients, but the exact relationship and underlying mechanisms remain to be fully elucidated (85–87). This inconsistency may be due to differences in study populations, methodologies, and other factors. Therefore, further large-scale, well-designed studies are needed to clarify the specific roles of these microbial components and their metabolites in lung cancer development.

Single-cell RNA sequencing (scRNA-seq) has revolutionized our understanding of immune-microbial interactions in the Gut-microbiota-lung Axis (65). This technology enables the analysis of gene expression at the single-cell level, providing a highly detailed view of the heterogeneity and functional states of immune cells in the gut and lungs (88). For example, scRNA-seq has revealed diverse subsets of immune cells, such as T cells, B cells, and macrophages, and their unique transcriptional profiles in response to microbial stimuli (89). By analyzing these transcriptional changes, researchers can gain insights into how gut microbiota influences the differentiation, activation, and function of immune cells, and how these immune cells, in turn, affect lung cancer development and immune responses (90). Some studies have shown that specific gut microbiota compositions can modulate the tumor-infiltrating immune cell landscape in the lungs, thereby influencing the efficacy of immunotherapy for lung cancer (29, 54). For instance, the presence of certain bacteria in the gut has been associated with increased numbers of cytotoxic T cells and natural killer cells in the lung tumor microenvironment, which may enhance the response to immune checkpoint inhibitors (8, 91).

Nevertheless, there are also discrepancies in the results of different studies. The specific types of immune cells and their functional states influenced by gut microbiota may vary depending on factors such as the composition and function of the microbiota, the genetic background of the host, and the stage of lung cancer (12, 92). Therefore, it is necessary to conduct more in-depth and comprehensive studies to fully understand the complex interactions between gut microbiota and immune cells in the context of lung cancer.

7.2 Animal models

Germ-free (GF) mice, which are raised in a sterile environment and lack exposure to microbiota, have been invaluable in studying the role of gut microbiota in Gut-microbiota-lung Axis interactions (93, 94). By colonizing GF mice with specific microbial communities, researchers can investigate the effects of these microbes on immune system development, lung function, and cancer-related processes (95). For example, studies have shown that the absence of gut microbiota in GF mice leads to impaired immune system development and function, and increased susceptibility to respiratory infections and lung cancer. When these mice are colonized with a normal gut microbiota, their immune systems and lung health are partially restored (96). This suggests that gut microbiota plays a crucial role in maintaining immune homeostasis and protecting against lung diseases.

Humanized microbiota models, which involve transferring human gut microbiota into GF mice or other animal models, further enable the study of the specific effects of human microbiota on Gut-microbiota-lung Axis interactions and lung cancer development (92). These models provide a more clinically relevant system for investigating the mechanistic links between gut microbiota and lung cancer, and for testing potential therapeutic interventions targeting the Gut-microbiota-lung Axis (97). For instance, researchers can use humanized microbiota models to evaluate the impact of specific probiotics or prebiotics on the composition and function of gut microbiota, and subsequently assess their effects on immune responses and tumor growth in the lungs (98).

However, there are also some limitations and differences in the results obtained from different animal models. The gut microbiota of mice differs from that of humans in terms of composition and function, which may affect the translatability of findings to human clinical settings (99). Additionally, the complexity of the Gut-microbiota-lung Axis and the multiple factors involved in its regulation make it challenging to fully recapitulate the human disease conditions in animal models (100). Therefore, it is important to carefully interpret the results from animal studies and to validate them in human clinical studies whenever possible.

8 Challenges and future directions

The manipulation of the gut microbiota holds promise for the treatment of lung cancer, however, the lack of standardized protocols poses a significant challenge (14). Currently, interventions such as FMT, probiotics, and prebiotics are being explored. But the preparation, administration, and quality control of these interventions vary across studies (14). For example, FMT can be administered via different routes, such as nasogastric tubes or capsules, and the donor selection criteria and fecal processing methods also differ. These variations make it difficult to compare results across studies and to translate findings into clinical practice (101). Li et al. (76) demonstrated that FMT could improve the efficacy of immunotherapy in lung cancer patients, but the long-term safety and optimal dosing regimens remain unclear. Similarly, probiotic and prebiotic interventions also lack standardized protocols. Different strains and doses of probiotics may have varying effects on the gut microbiota and immune system (102). Therefore, establishing standardized protocols for microbiota manipulation is crucial for advancing clinical applications.

The gut microbiome varies significantly among individuals due to factors such as genetics, diet, and lifestyle (103). This heterogeneity necessitates the development of personalized microbiome-based therapies for lung cancer patients (20). However, achieving personalization is challenging. First, a comprehensive understanding of the relationship between the gut microbiome and individual clinical outcomes is required (20). Studies have shown that certain microbial signatures are associated with better responses to immunotherapy, but these signatures may not be universal. For instance, some research indicates that a higher abundance of specific bacteria, such as *Akkermansia muciniphila*, is

linked to improved immunotherapy responses, while other studies report different associations (104, 105). Second, the dynamic nature of the gut microbiome further complicates personalization. The microbiome can change over time due to factors like diet and medication use. Therefore, developing personalized therapies requires continuous monitoring and adjustment of the microbiome (90). Additionally, integrating microbiome data with other clinical and molecular data is necessary to create more precise treatment plans (106). Despite these challenges, personalized microbiome-based therapies offer a potential avenue for improving lung cancer treatment outcomes.

The Gut-microbiota-lung Axis involves two-way communication between the gut and lungs, and the lung microbiota plays a crucial role in this process (107). However, the exact role of the lung microbiota in Gut-microbiota-lung Axis dynamics remains poorly understood. Some studies suggest that the lung microbiota influences systemic immunity and inflammation, which in turn affect gut microbiota composition and function (108). For example, Dora et al. (105) found that alterations in the lung microbiota could impact the gut immune system through immune cell trafficking and cytokine signaling. Conversely, gut microbiota-derived metabolites and immune cells can also affect lung health. Research has shown that SCFAs produced by gut microbiota can modulate lung immune responses and influence the development of respiratory diseases (109). However, 16S rRNA profiling is cost-efficient but rarely resolves beyond genus level and cannot predict functional genes; shotgun metagenomics delivers species/strain identification and metabolic pathway data yet requires higher DNA input and bioinformatics load, while both methods yield compositional data that may bias cross-sample comparison of low-abundance taxa (110, 111). Furthermore, the composition and function of the lung microbiota in different lung cancer subtypes and disease stages are not well characterized (112). Zheng et al. (113) revealed distinct lung microbiota profiles in patients with NSCLC compared to healthy individuals, but the functional implications of these differences remain to be elucidated.

Chronic obstructive pulmonary disease is a common comorbidity in lung cancer patients and can significantly influence Gut-microbiota-lung Axis interactions (114). COPD is characterized by chronic inflammation and airflow limitation, and it is associated with alterations in both the gut and lung microbiota (21). However, the impact of COPD on microbiota-immune interactions in the context of lung cancer is not fully understood. Some studies suggest that COPD-related inflammation may exacerbate gut barrier dysfunction and promote the translocation of gut microbial products to the lungs, further intensifying immune responses (23, 114). For example, Bowerman et al. (30) found that patients with COPD had increased gut permeability and altered gut microbiota composition, which were associated with enhanced systemic inflammation. This inflammation could potentially influence lung cancer progression and treatment outcomes. Additionally, the shared risk factors and pathophysiological mechanisms between COPD and lung cancer may also affect microbiota-immune interactions (31, 115). However, more research is needed to clarify these complex relationships and to develop targeted interventions for lung cancer patients with COPD and other comorbidities.

Furthermore, translation of probiotics, FMT, or dietary modulation into thoracic oncology practice faces pragmatic barriers identified by Georgiou 2021 and updated trials. First, regulatory agencies lack harmonized criteria for live-biotherapeutic potency, leading to variable CFU counts between batches of *Clostridium butyricum* CBM588 (67). Second, FMT sourced from ICI-responders requires donor re-screening every 30 days to exclude transmissible pathogens, raising cost to ≈ US $3,500 per infusion in a recent US phase-I NSCLC protocol (NCT05122546), a figure incompatible with universal reimbursement. Third, dietary interventions such as 20 g day ^{1} resistant starch increased fecal butyrate by 2.3-fold in chemo-immunotherapy patients, yet adherence at 12 weeks was 54%, predominantly limited by grade 1–2 bloating (52). Fourth, antibiotic stewardship programs report that 38% of lung cancer admissions receive at least one course of broad-spectrum agents during treatment, potentially abrogating any microbiota-directed benefit; integration of rapid point-of-care pathogen identification could reduce unnecessary prescriptions, but prospective data in oncology are lacking. Collectively, these data indicate that microbiota-based adjuvants are feasible only within clinical trials or specialized centers equipped with GMP-grade biobanks and dietetic support; routine deployment outside such frameworks is currently premature.

9 Conclusion

In conclusion, the Gut-microbiota-lung Axis plays a crucial role in lung cancer development and treatment. Gut microbiota dysbiosis can impact lung health through immune, neural, and humoral pathways, and influence the efficacy of lung cancer therapies. Targeting the Gut-microbiota-lung Axis offers potential for enhancing treatment efficacy and improving patient outcomes. However, challenges such as the lack of standardized protocols and the need for personalized therapies remain. Further research is needed to fully elucidate the mechanisms underlying the Gut-microbiota-lung Axis in lung cancer and to translate these findings into clinical applications.

Author contributions

LL: Writing – original draft, Writing – review & editing. LY: Writing – original draft, Writing – review & editing. HZ: Writing – review & editing. HL: Writing – original draft. TS: Writing – review & editing. LihL: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. This work was supported by the First People's Hospital of Baiyin College Scientific Research Project (project number: 2021YK-01; Project Name: Clinical Research on Thoracoscopic Lobectomy and Aegmentectomy in the Treatment of Early Lung Cancer).

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.

Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

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

Publisher's note

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, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. (2022) 74:229–63. doi: 10.3322/caac.21834

2. Li Y, Yan B, He S. Advances and challenges in the treatment of lung cancer. *Biomed Pharmacother*. (2023) 169:115891. doi: 10.1016/j.biopha.2023.115891

3. Mitsudomi T, Ito H, Okada M, Sugawara S, Shio Y, Tomii K, et al. Neoadjuvant nivolumab plus chemotherapy in resectable non-small-cell lung cancer in Japanese patients from CheckMate 816. *Cancer Sci*. (2024) 115:540–54. doi: 10.1111/cas.16030

4. Zhao Y, Liu Y, Li S, Peng Z, Liu X, Chen J, et al. Role of lung and gut microbiota on lung cancer pathogenesis. *J Cancer Res Clin Oncol*. (2021) 147:2177–86. doi: 10.1007/s00432-021-03644-0

5. Sun J, Song S, Liu J, Chen F, Li X, Wu G. Gut microbiota as a new target for anticancer therapy: from mechanism to means of regulation. *NPJ Biofilms Microbiomes*. (2025) 11:43. doi: 10.1038/s41522-025-00678-x

6. Haldar S, Jadhav SR, Gulati V, Beale DJ, Balkrishna A, Varshney A, et al. Unravelling the gut-lung axis: insights into microbiome interactions and Traditional Indian Medicine's perspective on optimal health. *FEMS Microbiol Ecol*. (2023) 99:fiad103. doi: 10.1093/femsec/fiad103

7. Chen J, Yu X, Wu X, Chai K, Wang S. Causal relationships between gut microbiota, immune cell, and Non-small cell lung cancer: a two-step, two-sample Mendelian randomization study. *J Cancer*. (2024) 15:1890–7. doi: 10.7150/jca.92699

8. Chen Z, Wang Z, Ma H, Bao H, Jiang T, Yang T, et al. Immune cells mediated the causal relationship between the gut microbiota and lung cancer: a Mendelian randomization study. *Front Microbiol*. (2024) 15:1390722. doi: 10.3389/fmicb.2024.1390722

9. Li R, Li J, Zhou X. Lung microbiome: new insights into the pathogenesis of respiratory diseases. *Signal Transduct Target Ther*. (2024) 9:19. doi: 10.1038/s41392-023-01722-y

10. Grenda A, Iwan E, Kuźnar-Kamińska B, Bomba A, Bielińska K, Krawczyk P, et al. Gut microbial predictors of first-line immunotherapy efficacy in advanced NSCLC patients. *Sci Rep*. (2025) 15:6139. doi: 10.1038/s41598-025-89406-1

11. Del Giudice T, Staropoli N, Tassone P, Tagliaferri P, Barbieri V. Gut microbiota are a novel source of biomarkers for immunotherapy in non-small-cell lung cancer (NSCLC). *Cancers*. (2024) 16:1806. doi: 10.3390/cancers16101806

12. Liu X, Cheng Y, Zang D, Zhang M, Li X, Liu D, et al. The role of gut microbiota in lung cancer: from carcinogenesis to immunotherapy. *Front Oncol*. (2021) 11:720842. doi: 10.3389/fonc.2021.720842

13. Eladham MW, Selvakumar B, Saheb Sharif-Askari N, Saheb Sharif-Askari F, Ibrahim SM, Halwani R. Unraveling the gut-Lung axis: exploring complex mechanisms in disease interplay. *Heliyon*. (2024) 10:e24032. doi: 10.1016/j.heliyon.2024.e24032

14. Georgiou K, Marinov B, Farooqi AA, Gazouli M. Gut microbiota in lung cancer: Where do we stand? *Int J Mol Sci*. (2021) 22:10429. doi: 10.3390/ijms221910429

15. Song X, Dou X, Chang J, Zeng X, Xu Q, Xu C. The role and mechanism of gut-lung axis mediated bidirectional communication in the occurrence and development of chronic obstructive pulmonary disease. *Gut Microbes*. (2024) 16:2414805. doi: 10.1080/19490976.2024.2414805

16. Verma A, Bhagchandani T, Rai A, Nikita, Sardarni UK, Bhavesh NS, et al. Short-Chain Fatty Acid (SCFA) as a connecting link between microbiota and gut-lung axis-A potential therapeutic intervention to improve lung health. *ACS Omega*. (2024) 9:14648–71. doi: 10.1021/acsomega.3c05846

17. Druszczynska M, Sadowska B, Kulesza J, Gąsienica-Gliwa N, Kulesza E, Fol M. The intriguing connection between the gut and lung microbiomes. *Pathogens*. (2024) 13:1005. doi: 10.3390/pathogens13111005

18. Enaud R, Prevel R, Ciarlo E, Beaufils F, Wieërs G, Guery B, et al. The gut-lung axis in health and respiratory diseases: a place for inter-organ and inter-kingdom crosstalks. *Front Cell Infect Microbiol*. (2020) 10:9. doi: 10.3389/fcimb.2020.00009

19. Zhang M, Qin Z, Huang C, Liang B, Zhang X, Sun W. The gut microbiota modulates airway inflammation in allergic asthma through the gut-lung axis related immune modulation: a review. *Biomol Biomed*. (2025) 25:727–38. doi: 10.17305/bb.2024.11280

20. Zhang H, Xu Z. Gut-lung axis: role of the gut microbiota in non-small cell lung cancer immunotherapy. *Front Oncol*. (2023) 13:1257515. doi: 10.3389/fonc.2023.1257515

21. Li N, Dai Z, Wang Z, Deng Z, Zhang J, Pu J, et al. Gut microbiota dysbiosis contributes to the development of chronic obstructive pulmonary disease. *Respir Res*. (2021) 22:274. doi: 10.1186/s12931-021-01872-z

22. Kotlyarov S. Role of short-chain fatty acids produced by gut microbiota in innate lung immunity and pathogenesis of the heterogeneous course of chronic obstructive pulmonary disease. *Int J Mol Sci*. (2022) 23:4768. doi: 10.3390/ijms23094768

23. Karakasidis E, Kotsiou OS, Gourgoulianis KI. Lung and gut microbiome in COPD. *J Pers Med*. (2023) 13:804. doi: 10.3390/jpm13050804

24. Borbet TC, Pawline MB, Zhang X, Wipperman MF, Reuter S, Maher T, et al. Influence of the early-life gut microbiota on the immune responses to an inhaled allergen. *Mucosal Immunol*. (2022) 15:1000–11. doi: 10.1038/s41385-022-00544-5

25. Wiertsema SP, van Bergenhenegouwen J, Garssen J, Knippels LMJ. The interplay between the gut microbiome and the immune system in the context of infectious diseases throughout life and the role of nutrition in optimizing treatment strategies. *Nutrients*. (2021) 13:886. doi: 10.3390/nu13030886

26. Yoo JY, Groer M, Dutra SVO, Sarkar A, McSkimming DI. Gut microbiota and immune system interactions. *Microorganisms*. (2020) 8:1587. doi: 10.3390/microorganisms8101587

27. Cheng C, Wang Z, Ding C, Liu P, Xu X, Li Y, et al. Bronchoalveolar lavage fluid microbiota is associated with the diagnosis and prognosis evaluation of lung cancer. *Phenomics*. (2024) 4:125–37. doi: 10.1007/s43657-023-00135-9

28. Zhu Z, Cai J, Hou W, Xu K, Wu X, Song Y, et al. Microbiome and spatially resolved metabolomics analysis reveal the anticancer role of gut Akkermansia muciniphila by crosstalk with intratumoral microbiota and reprogramming tumoral metabolism in mice. *Gut Microbes*. (2023) 15:2166700. doi: 10.1080/19490976.2023.2166700

29. Rahal Z, Liu Y, Peng F, Yang S, Jamal MA, Sharma M, et al. Inflammation mediated by gut microbiome alterations promotes lung cancer development and an immunosuppressed tumor microenvironment. *Cancer Immunol Res*. (2024) 12:1736–52. doi: 10.1158/2326-6066.CIR-24-0469

30. Bowerman KL, Rehman SF, Vaughan A, Lachner N, Budden KF, Kim RY, et al. Disease-associated gut microbiome and metabolome changes in patients with chronic obstructive pulmonary disease. *Nat Commun*. (2020) 11:5886. doi: 10.1038/s41467-020-19701-0

31. Wang L, Cai Y, Garssen J, Henricks PAJ, Folkerts G, Braber S. The bidirectional gut-lung axis in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. (2023) 207:1145–60. doi: 10.1164/rccm.202206-1066TR

32. Liu P, Zhou L, Chen H, He Y, Li G, Hu K. Identification of a novel intermittent hypoxia-related prognostic lncRNA signature and the ceRNA of lncRNA GSEC/miR-873-3p/EGLN3 regulatory axis in lung adenocarcinoma. *PeerJ*. (2023) 11:e16242. doi: 10.7717/peerj.16242

33. Bredon M, Danne C, Pham HP, Ruffié P, Bessede A, Rolhion N, et al. Faecalibaterium prausnitzii strain EXL01 boosts efficacy of immune checkpoint inhibitors. *Oncoimmunology*. (2024) 13:2374954. doi: 10.1080/2162402X.2024.2374954

34. Ouaknine Krief J, Helly de Tauriers P, Dumenil C, Neveux N, Dumoulin J, Giraud V, et al. Role of antibiotic use, plasma citrulline and blood microbiome in advanced non-small cell lung cancer patients treated with nivolumab. *J Immunother Cancer*. (2019) 7:176. doi: 10.1186/s40425-019-0658-1

35. Bi R, Hu R, Jiang L, Wen B, Jiang Z, Liu H, et al. Butyrate enhances erastin-induced ferroptosis of lung cancer cells via modulating the ATF3/SLC7A11 pathway. *Environ Toxicol*. (2024) 39:529–38. doi: 10.1002/tox.23857

36. Kim K, Kwon O, Ryu TY, Jung CR, Kim J, Min JK, et al. Propionate of a microbiota metabolite induces cell apoptosis and cell cycle arrest in lung cancer. *Mol Med Rep*. (2019) 20:1569–74. doi: 10.3892/mmr.2019.10431

37. Feng B, Lu Y, Zhang B, Zhu Y, Su Z, Tang L, et al. Integrated microbiome and metabolome analysis reveals synergistic efficacy of basil polysaccharide and gefitinib in lung cancer through modulation of gut microbiota and fecal metabolites. *Int J Biol Macromol*. (2024) 281(Pt 2):135992. doi: 10.1016/j.ijbiomac.2024.135992

38. Ubachs J, Ziemons J, Soons Z, Aarnoutse R, van Dijk DPJ, Penders J, et al. Gut microbiota and short-chain fatty acid alterations in cachectic cancer patients. *J Cachexia Sarcopenia Muscle*. (2021) 12:2007–21. doi: 10.1002/jcsm.12804

39. Dumont-Leblond N, Veillette M, Racine C, Joubert P, Duchaine C. Non-small cell lung cancer microbiota characterization: prevalence of enteric and potentially pathogenic bacteria in cancer tissues. *PLoS One*. (2021) 16:e0249832. doi: 10.1371/journal.pone.0249832

40. Li X, Mo Y, Shang S, Wu M, Ma S, Zhai Z, et al. Gut *Escherichia coli* promotes lung cancer by increasing circulating STAMBP production. *Discov Oncol*. (2025) 16:459. doi: 10.1007/s12672-025-02206-x

41. Zhao Z, Fei K, Bai H, Wang Z, Duan J, Wang J. Metagenome association study of the gut microbiome revealed biomarkers linked to chemotherapy outcomes in locally advanced and advanced lung cancer. *Thorac Cancer*. (2021) 12:66–78. doi: 10.1111/1759-7714.13711

42. Qu Z, Tian J, Sun J, Shi Y, Yu J, Zhang W, et al. Diallyl trisulfide inhibits 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone-induced lung cancer via modulating gut microbiota and the PPARγ/NF-κB pathway. *Food Funct*. (2024) 15:158–71. doi: 10.1039/d3fo03914e

43. Liu CG, Lin MX, Xin Y, Sun M, Cui J, Liu D, et al. Metagenomics and non-targeted metabolomics reveal the role of gut microbiota and its metabolites in brain metastasis of non-small cell lung cancer. *Thorac Cancer*. (2025) 16:e70068. doi: 10.1111/1759-7714.70068

44. Zeng G, Zeng L, Wang Y, Cao Z, Zeng X, Xue Z, et al. Correlation between gut microbiota characteristics and non-small cell lung cancer based on macrogenomics sequencing. *Hereditas*. (2024) 161:26. doi: 10.1186/s41065-024-00328-w

45. Gong R, Li H. The Role of 418 gut microbiota in small cell lung cancer progression: a Mendelian Randomisation Study. *J Coll Physicians Surg Pak*. (2025) 35:60–5. doi: 10.29271/jcpsp.2025.01.60

46. Nobels A, van Marcke C, Jordan BF, Van Hul M, Cani PD. The gut microbiome and cancer: from tumorigenesis to therapy. *Nat Metab*. (2025) 7:895–917. doi: 10.1038/s42255-025-01287-w

47. Derosa L, Hellmann MD, Spaziano M, Halpenny D, Fidelle M, Rizvi H, et al. Negative association of antibiotics on clinical activity of immune checkpoint inhibitors in patients with advanced renal cell and non-small-cell lung cancer. *Ann Oncol*. (2018) 29:1437–44. doi: 10.1093/annonc/mdy103

48. Hamada K, Yoshimura K, Hirasawa Y, Hosonuma M, Murayama M, Narikawa Y, et al. Antibiotic usage reduced overall survival by over 70% in non-small cell lung cancer patients on Anti-PD-1 immunotherapy. *Anticancer Res*. (2021) 41:4985–93. doi: 10.21873/anticanres.15312

49. Botticelli A, Vernocchi P, Marini F, Quagliariello A, Cerbelli B, Reddel S, et al. Gut metabolomics profiling of non-small cell lung cancer (NSCLC) patients under immunotherapy treatment. *J Transl Med*. (2020) 18:49. doi: 10.1186/s12967-020-02231-0

50. Tomita Y, Goto Y, Sakata S, Imamura K, Minemura A, Oka K, et al. Clostridium butyricum therapy restores the decreased efficacy of immune checkpoint blockade in lung cancer patients receiving proton pump inhibitors. *Oncoimmunology*. (2022) 11:2081010. doi: 10.1080/2162402X.2022.2081010

51. Zhao H, Li D, Liu J, Zhou X, Han J, Wang L, et al. Bifidobacterium breve predicts the efficacy of anti-PD-1 immunotherapy combined with chemotherapy in Chinese NSCLC patients. *Cancer Med*. (2023) 12:6325–36. doi: 10.1002/cam4.5312

52. Huang J, Liu D, Wang Y, Liu L, Li J, Yuan J, et al. Ginseng polysaccharides alter the gut microbiota and kynurenine/tryptophan ratio, potentiating the antitumour effect of antiprogrammed cell death 1/programmed cell death ligand 1 (anti-PD-1/PD-L1) immunotherapy. *Gut*. (2022) 71:734–45. doi: 10.1136/gutjnl-2020-321031

53. Hopkins AM, Kichenadasse G, McKinnon RA, Abuhelwa AY, Logan JM, Badaoui S, et al. Efficacy of first-line atezolizumab combination therapy in patients with non-small cell lung cancer receiving proton pump inhibitors: post hoc analysis of IMpower150. *Br J Cancer*. (2022) 126:42–7. doi: 10.1038/s41416-021-01606-4

54. Hu M, Zhu X, Huang X, Hua L, Lin X, Zhang H, et al. Optimizing anti-PD-1/PD-L1 therapy efficacy and fecal microbiota transplantation donor selection through gut mycobiome-based enterotype. *Cell Rep*. (2025) 44:115589. doi: 10.1016/j.celrep.2025.115589

55. Chen HH, Wu QJ, Zhang TN, Zhao YH. Gut microbiome and serum short-chain fatty acids are associated with responses to chemo- or targeted therapies in Chinese patients with lung cancer. *Front Microbiol*. (2023) 14:1165360. doi: 10.3389/fmicb.2023.1165360

56. Deng F, Du X, Zhang P, Xu J, Li Y, Yang Z. Impact of antibiotic on efficacy and adverse reactions of chemoimmunotherapy in non-small cell lung cancer patients: a retrospective cohort study. *Thorac Cancer*. (2024) 15:2560–9. doi: 10.1111/1759-7714.15490

57. Pensec C, Gillaizeau F, Guenot D, Bessard A, Carton T, Leuillet S, et al. Impact of pemetrexed chemotherapy on the gut microbiota and intestinal inflammation of patient-lung-derived tumor xenograft (PDX) mouse models. *Sci Rep*. (2020) 10:9094. doi: 10.1038/s41598-020-65792-6

58. Feng Y, Jiang Y, Zhou Y, Li ZH, Yang QQ, Mo JF, et al. Combination of BFHY with cisplatin relieved chemotherapy toxicity and altered gut microbiota in mice. *Int J Genomics*. (2023) 2023:3568416. doi: 10.1155/2023/3568416

59. Qiu B, Xi Y, Liu F, Li Y, Xie X, Guo J, et al. Gut microbiome is associated with the response to chemoradiotherapy in patients with non-small cell lung cancer. *Int J Radiat Oncol Biol Phys*. (2023) 115:407–18. doi: 10.1016/j.ijrobp.2022.07.032

60. Tamura K, Okuma Y, Nomura S, Fukuda A, Masuda K, Matsumoto Y, et al. Efficacy and safety of chemoimmunotherapy in advanced non-small cell lung cancer patients with antibiotics-induced dysbiosis: a propensity-matched real-world analysis. *J Cancer Res Clin Oncol*. (2024) 150:216. doi: 10.1007/s00432-024-05649-x

61. Wang X, Geng Q, Jiang H, Yue J, Qi C, Qin L. Fecal microbiota transplantation enhanced the effect of chemoimmunotherapy by restoring intestinal microbiota in LLC tumor-bearing mice. *BMC Immunol*. (2025) 26:30. doi: 10.1186/s12865-025-00710-x

62. Xia Q, Chen G, Ren Y, Zheng T, Shen C, Li M, et al. Investigating efficacy of "microbiota modulation of the gut-lung Axis" combined with chemotherapy in patients with advanced NSCLC: study protocol for a multicenter, prospective, double blind, placebo controlled, randomized trial. *BMC Cancer*. (2021) 21:721. doi: 10.1186/s12885-021-08448-6

63. Preet R, Islam MA, Shim J, Rajendran G, Mitra A, Vishwakarma V, et al. Gut commensal Bifidobacterium-derived extracellular vesicles modulate the therapeutic effects of anti-PD-1 in lung cancer. *Nat Commun*. (2025) 16:3500. doi: 10.1038/s41467-025-58553-4

64. Morita A, Ichihara E, Inoue K, Fujiwara K, Yokoyama T, Harada D, et al. Impacts of probiotics on the efficacies of immune checkpoint inhibitors with or without chemotherapy for patients with advanced non-small-cell lung cancer. *Int J Cancer*. (2024) 154:1607–15. doi: 10.1002/ijc.34842

65. Cao M, Deng Y, Hao Q, Yan H, Wang QL, Dong C, et al. Single-cell transcriptomic analysis reveals gut microbiota-immunotherapy synergy through modulating tumor microenvironment. *Signal Transduct Target Ther*. (2025) 10:140. doi: 10.1038/s41392-025-02226-7

66. Tomita Y, Ikeda T, Sakata S, Saruwatari K, Sato R, Iyama S, et al. Association of probiotic clostridium butyricum therapy with survival and response to immune checkpoint blockade in patients with lung cancer. *Cancer Immunol Res*. (2020) 8:1236–42. doi: 10.1158/2326-6066.CIR-20-0051

67. Tomita Y, Sakata S, Imamura K, Iyama S, Jodai T, Saruwatari K, et al. Association of *Clostridium butyricum* therapy using the live bacterial product CBM588 with the survival of patients with lung cancer receiving chemoimmunotherapy combinations. *Cancers*. (2023) 16:47. doi: 10.3390/cancers16010047

68. Wan L, Wu C, Wu Q, Luo S, Liu J, Xie X. Impact of probiotics use on clinical outcomes of immune checkpoint inhibitors therapy in cancer patients. *Cancer Med*. (2023) 12:1841–9. doi: 10.1002/cam4.4994

69. Qu Z, Zhang L, Hou R, Ma X, Yu J, Zhang W, et al. Exposure to a mixture of cigarette smoke carcinogens disturbs gut microbiota and influences metabolic homeostasis in A/J mice. *Chem Biol Interact*. (2021) 344:109496. doi: 10.1016/j.cbi.2021.109496

70. Leung HKM, Lo EKK, El-Nezami H. Theabrownin alleviates colorectal tumorigenesis in murine AOM/DSS model via PI3K/Akt/mTOR pathway suppression and gut microbiota modulation. *Antioxidants*. (2022) 11:1716. doi: 10.3390/antiox11091716

71. Cao B, Wang S, Li R, Wang Z, Li T, Zhang Y, et al. Xihuang Pill enhances anticancer effect of anlotinib by regulating gut microbiota composition and tumor angiogenesis pathway. *Biomed Pharmacother*. (2022) 151:113081. doi: 10.1016/j.biopha.2022.113081

72. Jiang RY, Wang T, Lan QY, Qin YC, Man TT, Sun H, et al. BuFeiXiaoJiYin ameliorates the NLRP3 inflammation response and gut microbiota in mice with lung cancer companied with Qi-yin deficiency. *Cancer Cell Int*. (2022) 22:121. doi: 10.1186/s12935-022-02543-9

73. Li F, Hao S, Gao J, Jiang P. EGCG alleviates obesity-exacerbated lung cancer progression by STAT1/SLC7A11 pathway and gut microbiota. *J Nutr Biochem*. (2023) 120:109416. doi: 10.1016/j.jnutbio.2023.109416

74. Bi W, Cai S, Hang Z, Lei T, Wang D, Wang L, et al. Transplantation of feces from mice with Alzheimer's disease promoted lung cancer growth. *Biochem Biophys Res Commun*. (2022) 600:67–74. doi: 10.1016/j.bbrc.2022.01.078

75. Chen M, Ma L, Yu H, Huang S, Zhang J, Gong J, et al. JK5G postbiotics attenuate immune-related adverse events in NSCLC patients by regulating gut microbiota: a randomized controlled trial in China. *Front Oncol*. (2023) 13:1155592. doi: 10.3389/fonc.2023.1155592

76. Li L, Liang Y, Yu M, Zhao L, Mei Q, Yu Y, et al. Advances in immune checkpoint inhibitors therapy for small cell lung cancer. *Cancer Med*. (2023) 12:11097–106. doi: 10.1002/cam4.5659

77. Tonneau M, Nolin-Lapalme A, Kazandjian S, Auclin E, Panasci J, Benlaifaoui M, et al. *Helicobacter* pylori serology is associated with worse overall survival in patients with melanoma treated with immune checkpoint inhibitors. *Oncoimmunology*. (2022) 11:2096535. doi: 10.1080/2162402X.2022.2096535

78. Zhao X, Liu C, Peng L, Wang H. Metformin facilitates anti-PD-L1 efficacy through the regulation of intestinal microbiota. *Genes Immun*. (2024) 25:7–13. doi: 10.1038/s41435-023-00234-7

79. Yan Q, Su S, Dai G, He L. Prebiotics modulate gut microbiota-mediated T-cell Immunity to Enhance the inhibitory effect of sintilimab in lewis lung adenocarcinoma model mice. *Anticancer Agents Med Chem*. (2023) 23:1966–73. doi: 10.2174/1871520623666230707112244

80. Ni Y, Li R, Shen X, Yi D, Ren Y, Wang F, et al. Diaphorobacter nitroreducens synergize with oxaliplatin to reduce tumor burden in mice with lung adenocarcinoma. *mSystems*. (2024) 9:e0132323. doi: 10.1128/msystems.01323-23

81. Yiminniyaze R, Zhang Y, Zhu N, Zhang X, Wang J, Li C, et al. Characterizations of lung cancer microbiome and exploration of potential microbial risk factors for lung cancer. *Sci Rep*. (2025) 15:15683. doi: 10.1038/s41598-025-98424-y

82. Bui TNY, Paul A, Guleria S, O'Sullivan JM, Toldi G. Short-chain fatty acids-a key link between the gut microbiome and T-lymphocytes in neonates? *Pediatr Res*. (2025) [Online ahead of print]. doi: 10.1038/s41390-025-04075-0.

83. Feitelson MA, Arzumanyan A, Medhat A, Spector I. Short-chain fatty acids in cancer pathogenesis. *Cancer Metastasis Rev*. (2023) 42:677–98. doi: 10.1007/s10555-023-10117-y

84. Desharnais L, Sorin M, Rezanejad M, Liu B, Karimi E, Atallah A, et al. Spatially mapping the tumour immune microenvironments of non-small cell lung cancer. *Nat Commun*. (2025) 16:1345. doi: 10.1038/s41467-025-56546-x

85. Liu F, Li J, Guan Y, Lou Y, Chen H, Xu M, et al. Dysbiosis of the Gut Microbiome is associated with tumor biomarkers in lung cancer. *Int J Biol Sci*. (2019) 15:2381–92. doi: 10.7150/ijbs.35980

86. Xu N, Wang L, Li C, Ding C, Li C, Fan W, et al. Microbiota dysbiosis in lung cancer: evidence of association and potential mechanisms. *Transl Lung Cancer Res*. (2020) 9:1554–68. doi: 10.21037/tlcr-20-156

87. 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:e0080221. doi: 10.1128/Spectrum.00802-21

88. Sun B, Xun Z, Zhang N, Liu K, Chen X, Zhao H. Single-cell RNA sequencing in cancer research: discovering novel biomarkers and therapeutic targets for immune checkpoint blockade. *Cancer Cell Int*. (2023) 23:313. doi: 10.1186/s12935-023-03158-4

89. Lu H, Qian J, Cheng L, Shen Y, Chu T, Zhao C. Single-cell RNA-sequencing uncovers the dynamic changes of tumour immune microenvironment in advanced lung adenocarcinoma. *BMJ Open Respir Res*. (2023) 10:e001878. doi: 10.1136/bmjresp-2023-001878

90. Feng C, Li N, Gao G, He Q, Kwok LY, Zhang H. Dynamic changes of the gut microbiota and its functional metagenomic potential during the development of non-small cell lung cancer. *Int J Mol Sci*. (2024) 25:3768. doi: 10.3390/ijms25073768

91. Ren S, Feng L, Liu H, Mao Y, Yu Z. Gut microbiome affects the response to immunotherapy in non-small cell lung cancer. *Thorac Cancer*. (2024) 15:1149–63. doi: 10.1111/1759-7714.15303

92. Routy B, Le Chatelier E, Derosa L, Duong CPM, Alou MT, Daillère R, et al. Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. *Science*. (2018) 359:91–7. doi: 10.1126/science.aan3706

93. Bauer JE. On behalf of hermaphrodites and mongrels: refocusing the reception of magnus hirschfeld's critical thought on sexuality and race. *J Homosex*. (2021) 68:777–801. doi: 10.1080/00918369.2019.1661686

94. Amr A, Jaradat S, AlKhatib H, Hamadneh I, Hamadneh L, Hodali H, et al. Extraction of anthocyanins from black grape by-products and improving their stability using Cobalt(II) complexation. *Prev Nutr Food Sci*. (2022) 27:457–63. doi: 10.3746/pnf.2022.27.4.457

95. Taha A, Enodien B, Ochs V, Bachmann M, Gripp M, Adamina M, et al. Analysis of factors relevant to revenue improvement in ventral hernia repair, their influence on surgical training, and development of predictive models: an economic evaluation. *Healthcare*. (2021) 9:1226. doi: 10.3390/healthcare9091226

96. Sencio V, Machado MG, Trottein F. The lung-gut axis during viral respiratory infections: the impact of gut dysbiosis on secondary disease outcomes. *Mucosal Immunol*. (2021) 14:296–304. doi: 10.1038/s41385-020-00361-8

97. Sun Y, Wen M, Liu Y, Wang Y, Jing P, Gu Z, et al. The human microbiome: a promising target for lung cancer treatment. *Front Immunol*. (2023) 14:1091165. doi: 10.3389/fimmu.2023.1091165

98. Lundberg R. Humanizing the gut microbiota of mice: opportunities and challenges. *Lab Anim*. (2019) 53:244–51. doi: 10.1177/0023677218787554

99. Kieser S, Zdobnov EM, Trajkovski M. Comprehensive mouse microbiota genome catalog reveals major difference to its human counterpart. *PLoS Comput Biol*. (2022) 18:e1009947. doi: 10.1371/journal.pcbi.1009947

100. Huang J, Huang J. Microbial biomarkers for lung cancer: current understandings and limitations. *J Clin Med*. (2022) 11:7298. doi: 10.3390/jcm11247298

101. Wei H, Yue Z, Han J, Chen P, Xie K, Sun Y, et al. Oral compound probiotic supplements can improve the quality of life for patients with lung cancer during chemotherapy: a randomized placebo-controlled study. *Thorac Cancer*. (2024) 15:182–91. doi: 10.1111/1759-7714.15177

102. West ML, Hart S, Loughman A, Jacka FN, Staudacher HM, Abbaspour A, et al. Challenges and priorities for researching the gut microbiota in individuals living with anorexia nervosa. *Int J Eat Disord*. (2023) 56:2001–11. doi: 10.1002/eat.24033

103. Duttagupta S, Hakozaki T, Routy B, Messaoudene M. The gut microbiome from a biomarker to a novel therapeutic strategy for immunotherapy response in patients with lung cancer. *Curr Oncol*. (2023) 30:9406–27. doi: 10.3390/curroncol30110681

104. Gopalakrishnan V, Helmink BA, Spencer CN, Reuben A, Wargo JA. The influence of the gut microbiome on cancer, immunity, and cancer immunotherapy. . *Cancer Cell*. (2018) 33:570–80. doi: 10.1016/j.ccell.2018.03.015

105. Dora D, Szõcs E, Soós Á, Halasy V, Somodi C, Mihucz A, et al. From bench to bedside: an interdisciplinary journey through the gut-lung axis with insights into lung cancer and immunotherapy. *Front Immunol*. (2024) 15:1434804. doi: 10.3389/fimmu.2024.1434804

106. Jain N. The need for personalized approaches to microbiome modulation. *Front Public Health*. (2020) 8:144. doi: 10.3389/fpubh.2020.00144

107. Marsland BJ, Trompette A, Gollwitzer ES. The gut-lung axis in respiratory disease. *Ann Am Thorac Soc*. (2015) 12(Suppl 2):S150–6. doi: 10.1513/AnnalsATS.201503-133AW

108. He Y, Wen Q, Yao F, Xu D, Huang Y, Wang J. Gut-lung axis: the microbial contributions and clinical implications. *Crit Rev Microbiol*. (2017) 43:81–95. doi: 10.1080/1040841X.2016.1176988

109. Thome CD, Tausche P, Hohenberger K, Yang Z, Krammer S, Trufa DI, et al. Short-chain fatty acids induced lung tumor cell death and increased peripheral blood CD4+ T cells in NSCLC and control patients ex vivo. *Front Immunol*. (2024) 15:1328263. doi: 10.3389/fimmu.2024.1328263

110. Pasolli E, Truong DT, Malik F, Waldron L, Segata N. Machine Learning Meta-analysis of Large Metagenomic Datasets: tools and Biological Insights. *PLoS Comput Biol*. (2016) 12:e1004977. doi: 10.1371/journal.pcbi.1004977

111. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol*. (2013) 79:5112–20. doi: 10.1128/AEM.01043-13

112. Zhang Y, Chen X, Wang Y, Li L, Ju Q, Zhang Y, et al. Alterations of lower respiratory tract microbiome and short-chain fatty acids in different segments in lung cancer: a multiomics analysis. *Front Cell Infect Microbiol*. (2023) 13:1261284. doi: 10.3389/fcimb.2023.1261284

113. Zheng X, Lu X, Hu Y. Distinct respiratory microbiota associates with lung cancer clinicopathological characteristics. *Front Oncol*. (2023) 13:847182. doi: 10.3389/fonc.2023.847182

114. Vaughan A, Frazer ZA, Hansbro PM, Yang IA. COPD and the gut-lung axis: the therapeutic potential of fibre. *J Thorac Dis*. (2019) 11(Suppl 17):S2173–80. doi: 10.21037/jtd.2019.10.40

115. Zhang D, Li S, Wang N, Tan HY, Zhang Z, Feng Y. The cross-talk between gut microbiota and lungs in common lung diseases. *Front Microbiol*. (2020) 11:301. doi: 10.3389/fmicb.2020.00301

Frontiers in Medicine

Translating medical research and innovation into improved patient care

A multidisciplinary journal which advances our medical knowledge. It supports the translation of scientific advances into new therapies and diagnostic tools that will improve patient care.

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