frontiers | Research Topics

Molecular informatics in precision medicine

Edited by
Abdul Azeez Sayed, J. Francis Borgio and Hari S. Sharma

Published in
Frontiers in Medicine

ISSN 1664-8714
ISBN 978-2-8325-5845-4
DOI 10.3389/978-2-8325-5845-4

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

Molecular informatics in precision medicine

Topic editors

Abdul Azeez Sayed — Imam Abdulrahman Bin Faisal University, Saudi Arabia

J. Francis Borgio — Imam Abdulrahman Bin Faisal University, Saudi Arabia

Hari S. Sharma — Department of Pathology, Erasmus Medical Center, Netherlands

Citation

Sayed, A. A., Borgio, J. F., Sharma, H. S., eds. (2025). *Molecular informatics in precision medicine*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-5845-4

Table of contents

frontiers | Frontiers in Medicine

TYPE Editorial
PUBLISHED 03 December 2024
DOI 10.3389/fmed.2024.1526520

OPEN ACCESS

EDITED AND REVIEWED BY
Alice Chen,
Consultant, Potomac, MD, United States

*CORRESPONDENCE
Sayed AbdulAzeez
asayed@iau.edu.sa

RECEIVED 11 November 2024
ACCEPTED 22 November 2024
PUBLISHED 03 December 2024

CITATION
AbdulAzeez S, Sharma HS and Borgio JF (2024) Editorial: Molecular informatics in precision medicine. *Front. Med.* 11:1526520. doi: 10.3389/fmed.2024.1526520

Editorial: Molecular informatics in precision medicine

Sayed AbdulAzeez[1]*, Hari S. Sharma[2] and J. Francis Borgio[1]

[1]Department of Genetic Research, Institute for Research and Medical Consultations (IRMC), Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia, [2]Department of Pathology and Clinical Bioinformatics, Erasmus MC, University Medical Center, Rotterdam, Netherlands

KEYWORDS

bioinformatics, personalized medicine, next-generation sequence (NGS), sarcoidosis, biomarker, recurrent pregnancy loss

Editorial on the Research Topic
Molecular informatics in precision medicine

1 Introduction

The emerging concept in medicine shifts toward precision medicine personalized to an individual's unique genetic makeup and environmental factors. By integrating advanced molecular technologies such as genomics, transcriptomics, proteomics, metabolomics and microbiomics we can unlock the potential to revolutionize healthcare. Molecular informatics plays a crucial role in this transformation. By analyzing vast amounts of biological data, researchers can identify genetic markers, predict disease risk and develop personalized treatment strategies. This Research Topic deals the latest advancements in molecular informatics, exploring how these technologies can be connected to improve patient outcomes. Key areas of focus include (i) next-generation sequencing, a leveraging cutting-edge sequencing technologies to unravel complex genetic variations, (ii) computer-aided drug discovery for utilizing computational tools to accelerate drug discovery and development, (iii) molecular modeling and simulation for simulating biological processes at the molecular level to gain insights into disease mechanisms and (iv) bioinformatics specially applying computational methods to analyze and interpret biological data. Through a comprehensive exploration of these topics, this collection of articles aims to provide a valuable resource for researchers, clinicians and industry professionals working at the forefront of precision medicine. By understanding the power of molecular informatics, we can move closer to a future where healthcare is truly personalized.

2 Contributions

A total of nine articles including one review and eight original research articles have been published on this Research Topic. Hu et al. revealed the shared pathogenic mechanisms between type 2 diabetes mellitus and ulcerative colitis using comprehensive strategy merging bioinformatics and machine learning. Authors sourced data from the Gene Expression Omnibus database and targets of Gegen Qinlian decoction were identified using PharmMapper and SwissTargetPrediction. Targets associated with type 2 diabetes mellitus and ulcerative colitis were compiled from various databases.

Weighted gene co-expression network analysis, single-cell sequencing analysis, immune infiltration analysis, machine learning, DEG analysis and network pharmacology were the six analysis included in the study. Results showed that the co-morbidity between type 2 diabetes mellitus and ulcerative colitis is primarily associated with immune-inflammatory pathways, including IL-17, TNF, chemokine and toll-like receptor signaling pathways. Machine learning studies identified IGFBP3 as a biomarker for Gegen Qinlian Decoction in treating type 2 diabetes mellitus, while BACE2, EPHB4 and EPHA2 emerged as biomarkers for Gegen Qinlian decoction in ulcerative colitis treatment. The study provides insights into the shared pathogenesis of type 2 diabetes mellitus and ulcerative colitis and proposes novel targets and therapeutic strategies.

Ye et al. reported the presence of mutations in the *ANXA4* gene in patients with recurrent spontaneous abortion (RSA). Variants were annotated and filtered and the pathogenicity of mutations was predicted using various tools. Through whole exome sequencing an *ANXA4* mutation (p.G8D) was identified in one of the 325 samples from recurrent spontaneous abortion patients. This amino acid change was highly conserved among vertebrate species and predicted to be deleterious. Cell adhesion, migration and invasion were all shown to be inhibited by this mutation in functional experiments. The recently discovered *ANXA4* mutation might have important implications for genetic testing and the pathophysiology of recurrent spontaneous abortions. Another interesting bioinformatic meta-analysis by van Wijck et al. exposed a concise gene signature in pathogenesis of sarcoidosis. Expression datasets of sarcoidosis have uncovered both new and previously well-known genes that may be involved in type I and type II interferon-mediated signaling pathways. Cytokines like interferons and STAT1 were upregulated, but eukaryotic initiation factor 2 signaling was downregulated in expression datasets of sarcoidosis according to *in silico* functional analysis. A key role in pathogenesis of sarcoidosis may be attributed to the unique upregulation of matrix metallopeptidase 12 in afflicted tissues. The authors provided additional evidence on the emerging reported evidence to use JAK inhibitors as a targeted treatment strategy in patients with sarcoidosis. Tissue-specific signatures of genes like *MMP12*, *CXCR6* and *SNTB2* observed by the authors suggested that these genes might involve in granuloma formation and progression. It is imperative to do more transcriptome investigations in order to validate the observation from bioinformatic meta-analysis by van Wijck et al..

Shi et al. analyzed the adverse event reports related to immune checkpoint inhibitors such as protein-1 (PD-1) and its ligand (PD-L1). They collected 5,322 reports from the United States food and drug administration adverse event reporting system regarding the adverse event reports related to protein-1 and its ligand inhibitors. The study observed that except for pembrolizumab, five PD-1/PD-L1 inhibitors were associated with serious side effects on the endocrine glands. The majority of patients experienced adverse events between 30 and 365 days with a median time of 61 days. The majority of patients experienced prolonged hospitalization in over 40% and death in over 10% of cases after administration of nivolumab, pembrolizumab, or durvalumab. The authors concluded that men aged ≥65 years should be concerned about endocrine-related adverse events and emphasize the importance of addressing these issues when using these PD-1/PD-L1 inhibitors. Another study by Ni et al. uses a network pharmacology approach to predict the active ingredients of *Ixeris chinensis*, targets of action, and possible interventions in diseases. The authors utilized various databases and software to predict active ingredients, target genes, protein interactions and signaling pathways. The results revealed 12 effective components of *I. chinensis* and 40 key targets, including AKT1, EGFR, TNF, SRC and ESR1. Molecular docking analysis revealed that the main active components of *I. chinensis* can bind well with key targets. The study also provides a basis for research on *I. chinensis* treatment pathways for related diseases and subsequent drug development. The study further demonstrated the feasibility of *I. chinensis* as a therapeutic agent for many diseases and established a foundation for investigating the specific mechanisms of treating diseases and the development of novel pharmaceuticals. However, more *in vitro* and *in vivo* experiments are needed to verify these observations.

Behairy et al. investigated the disease susceptibility of the mannose-binding lectin (*MBL*) mutation (rs1800450) in the development of vitiligo and psoriasis. This observational study examined the *MBL2* gene at codon 54 using real-time PCR and computational modeling of the single nucleotide polymorphism. All genetic association models found no evidence that rs1800450 significantly impact the risk of psoriasis or vitiligo disease. The study also found no significant correlation with rs1800450 on the clinicopathological features of both psoriasis and vitiligo. The rs1800450 SNP on the *MBL2* gene was not associated with autoimmune skin disorders risk in Egyptian adults. The data further supports that *MBL2* is redundant and does not significantly affect autoimmune skin disorders. Hemagglutinin is an important element in influenza virus infection, making it a potential target for therapeutic and vaccine development. Zou et al. aimed to create a computational model for identifying hemagglutinin using a benchmark dataset of 106 hemagglutinin and 106 non-hemagglutinin sequences from UniProt. Using the stacking approach, Zou et al. created an integrated classifier model with an accuracy of 95.85% in 5-fold cross-validation and 93.18% in the independent test. The high prediction accuracy makes it useful for biochemical researchers studying hemagglutinin.

Ping et al. developed an amplification analysis using double allele-specific binding primers for accurate measurement of antihypertensive pharmacogenomics. The researchers used quadruplex quantitative PCR (qPCR) and triplex qPCR analysis for genotyping. Mismatch allele-specific F-primers were validated through agreement analysis/reproducibility evaluation. Seven pairs of primers were successfully selected, with amplification efficiency ranging (except for ADRB1) from 0.9 to 1.1 with the coefficient of variation (CV) was <5%. The study concluded that multiplex amplification analysis using screened allele-specific binding primers is a simple, reliable and accurate tool for guiding drug delivery in antihypertensive personalized treatment. Wen et al. compared the effectiveness of metagenomic next-generation sequencing and conventional microbiological tests in diagnosing pulmonary infections in patients with systemic autoimmune rheumatic diseases receiving immunosuppressant therapy. The study involved reviewing the medical records of 40 patients

with pulmonary infections and systemic autoimmune rheumatic diseases treated with immunosuppressants or corticosteroids. Bronchoalveolar lavage fluid samples were collected and examined by metagenomic next-generation sequencing and conventional microbiological tests. The results showed that metagenomic next-generation sequencing had a higher diagnostic accuracy for detecting co-infections with bacteria and fungi and single infections with fungi. The detection rate of co-infection was significantly higher for metagenomic next-generation sequencing than conventional microbiological tests. The study concluded that metagenomic next-generation sequencing's superior accuracy can help ensure timely adjustment of treatment regimens, improving diagnosis and outcomes in patients with systemic autoimmune rheumatic diseases treated with immunosuppressants.

3 Conclusion

The Research Topic "*Molecular informatics in precision medicine*" presents a compelling collection of research articles that highlight the transformative power of bioinformatics and computational biology in advancing personalized healthcare. One of the key themes explored in this Research Topic is the identification of novel biomarkers and therapeutic targets for complex diseases. For instance, the study by Hu et al. revealed shared pathogenic mechanisms between type 2 diabetes mellitus and ulcerative colitis, suggesting potential therapeutic strategies. Similarly, Ye et al. identified a novel mutation in the *ANXA4* gene associated with recurrent spontaneous abortion, opening new avenues for genetic testing and targeted therapies. Another important theme is the application of advanced computational methods to analyze large-scale genomic and clinical data. The study by van Wijck et al. utilized bioinformatic meta-analysis to identify key genes involved in the pathogenesis of sarcoidosis, providing insights into potential therapeutic interventions. The increasing use of immune checkpoint inhibitors has led to significant advancements in cancer therapy. However drugs used are associated with adverse effects. Shi et al. investigated the adverse event reports related to PD-1 and PD-L1 inhibitors, highlighting the importance of careful monitoring and management of treatment-related side effects. In addition to human diseases, the Research Topic also explores into the application of computational methods in drug discovery and development. Ni et al. used network pharmacology to identify potential therapeutic targets for *Ixeris chinensis*, a traditional Chinese medicine. Finally, the Research Topic highlights the potential of emerging technologies like metagenomic next-generation sequencing in improving clinical diagnosis and treatment. Wen et al. demonstrated the superiority of metagenomic sequencing over conventional methods in diagnosing pulmonary infections in immunosuppressed patients.

In conclusion, the articles in this Research Topic showcase the immense potential of molecular informatics in revolutionizing precision medicine. By integrating computational biology with clinical research, we can unlock new insights into disease mechanisms, identify novel biomarkers, and develop personalized treatment strategies. As technology continues to advance, we can expect to see even more groundbreaking discoveries in the field of precision medicine.

Author contributions

SA: Conceptualization, Writing – original draft, Writing – review & editing. HS: Conceptualization, Writing – original draft, Writing – review & editing. JB: Conceptualization, Writing – original draft, Writing – review & editing.

Acknowledgments

We extend our appreciation to the authors for their substantial contributions to this Research Topic. We express our profound appreciation to the dedicated reviewers for their comprehensive and prompt evaluations, which significantly enhanced the quality of this publication. We would like to convey our appreciation to the editorial staff of the Frontiers in Medicine journal for their continuous support, which has been instrumental in the realization of 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.

Publisher's note

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 06 July 2023
DOI 10.3389/fmed.2023.1161661

OPEN ACCESS

EDITED BY
Enrico Heffler,
Humanitas University, Italy

REVIEWED BY
Davide Borroni,
Riga Stradiņš University, Latvia
Ying Liang,
Peking University Third Hospital, China
Ting Fang Lee,
New York University, United States

*CORRESPONDENCE
Lixia Xu
✉ 15820264379@139.com
Shuangxin Liu
✉ 13543456446@163.com

†These authors have contributed equally to this work

RECEIVED 09 February 2023
ACCEPTED 19 June 2023
PUBLISHED 06 July 2023

CITATION
Wen S, Peng S, Hu X, Jiang N, Li B, Chen B, Deng S, Yuan Y, Wu Q, Tao Y, Ma J, Li S, Lin T, Wen F, Li Z, Huang R, Feng Z, He C, Wang W, Liang X, Shi W, Xu L and Liu S (2023) Validation of metagenomic next-generation sequencing of bronchoalveolar lavage fluid for diagnosis of suspected pulmonary infections in patients with systemic autoimmune rheumatic diseases receiving immunosuppressant therapy. *Front. Med.* 10:1161661. doi: 10.3389/fmed.2023.1161661

Validation of metagenomic next-generation sequencing of bronchoalveolar lavage fluid for diagnosis of suspected pulmonary infections in patients with systemic autoimmune rheumatic diseases receiving immunosuppressant therapy

Sichun Wen[1,2†], Siqi Peng[2,3†], Xuejiao Hu[4†], Nan Jiang[2], Bohou Li[2], Boxi Chen[2], Shuting Deng[2], Ye Yuan[2], Qiong Wu[2], Yiming Tao[2], Jianchao Ma[2], Sijia Li[2], Ting Lin[2], Feng Wen[2], Zhuo Li[2], Renwei Huang[2], Zhonglin Feng[2], Chaosheng He[2], Wenjian Wang[2], Xinling Liang[2], Wei Shi[2], Lixia Xu[2]* and Shuangxin Liu[1,2,3]*

[1]School of Medicine, South China University of Technology, Guangzhou, Guangdong, China, [2]Department of Nephrology, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, China, [3]The Second School of Clinical Medicine, Southern Medical University, Guangzhou, China, [4]Laboratory Medicine, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, China

Background: The accuracy and sensitivity of conventional microbiological tests (CMTs) are insufficient to identify opportunistic pathogens in patients with systemic autoimmune rheumatic diseases (SARDs). The study aimed to assess the usefulness of metagenomic next-generation sequencing (mNGS) vs. CMTs for the diagnosis of pulmonary infections in patients with SARDs receiving immunosuppressant therapy.

Methods: The medical records of 40 patients with pulmonary infections and SARDs treated with immunosuppressants or corticosteroids were reviewed retrospectively. Bronchoalveolar lavage fluid (BALF) samples were collected from all patients and examined by mNGS and CMTs. Diagnostic values of the CMTs and mNGS were compared with the clinical composite diagnosis as the reference standard.

Results: Of the 40 patients included for analysis, 37 (92.5%) were diagnosed with pulmonary infections and 3 (7.5%) with non-infectious diseases, of which two were considered primary diseases and one an asthma attack. In total, 15 pathogens (7 bacteria, 5 fungi, and 3 viruses) were detected by CMTs as compared to 58 (36 bacteria, 12 fungi, and 10 viruses) by mNGS. Diagnostic accuracy of mNGS was superior to that of the CMTs for the detection of co-infections with bacteria and fungi (95 vs. 53%, respectively, $p<0.01$), and for the detection of single infections with fungi (97.5 vs. 55%, respectively, $p<0.01$). Of the 31 patients diagnosed with co-infections, 4 (12.9%) were positive for two pathogens and 27 (87.1%) for three

or more. The detection rate of co-infection was significantly higher for mNGS than CMTs (95 vs. 16%, respectively, $p<0.01$).

Conclusion: The accuracy of mNGS was superior to that of the CMTs for the diagnosis of pulmonary infections in patients with SARDs treated with immunosuppressants. The rapid diagnosis by mNGS can ensure timely adjustment of treatment regimens to improve diagnosis and outcomes.

KEYWORDS

metagenomic next-generation sequencing, systemic autoimmune rheumatic diseases, immunosuppressants, bronchoalveolar lavage fluid, pulmonary infection, antibiotics

Introduction

Systemic autoimmune rheumatic diseases (SARDs) are a group of autoimmune-mediated diseases characterized by the overproduction of autoantibodies. The incidence of SARDs has continued to increase in recent years (1). The propensity for pulmonary infection of patients with SARDs is both an inherent form of disease-related immune dysregulation and acquired by use of immunosuppressants (2). Treatment of SARDs involves using immunosuppressants or immunomodulatory drugs, such as mycophenolate mofetil, azathioprine, and tacrolimus (3). However, immunosuppression and severe activity of the underlying disease often lead to infection by various opportunistic pathogens (4, 5), such as Cryptococcus novelis, which can cause fatal meningoencephalitis in immunocompromised individuals, and Pneumocystis jirovecii, which can cause pneumonia after hematopoietic stem cell and solid organ transplantation and is especially problematic in patients receiving immunosuppressants. Due to the abundance of potential pathogens and the possibility that the symptoms of the primary disease or the treatment regimen may hide an actual pulmonary infection, diagnosis of this population is often challenging (6).

Accurate diagnosis and effective treatment of infections often require testing of pathogens for drug sensitivity. Conventional microbiological tests (CMTs), such as culture-based detection assays, immunological analysis, and polymerase chain reaction (PCR), are commonly used for clinical diagnosis of Pneumocystis jirovecii pneumonia (PJP) (7). However, CMTs are limited by insufficient sensitivity and speed for accurate and rapid identification of pathogens in samples from immunocompromised patients (8). Thus, alternative methods, such as multiplex real-time quantitative fluorescent PCR, biosensors, and metagenomic next-generation sequencing (mNGS), are potentially superior diagnostic options for this patient population. mNGS is a high-throughput nucleic acid sequencing technology that has been widely applied to detect various pathogens (9–13). The advantages of mNGS include shorter detection times and accurate detection of multiple pathogens simultaneously by DNA or mRNA sequencing of clinical samples (14). While mNGS has been used to diagnose pneumonia in immunocompromised patients, with advantages in pathogen detection, particularly in fungal and co-infections, studies exploring its potential application in patients with SARDs remain limited (15, 16). The present study aimed to compare the diagnostic accuracy of mNGS vs. CMTs for the detection of pulmonary infections in immunocompromised patients.

Materials and methods

The cohort of this retrospective study included 40 patients with SARDs and suspected pulmonary infections who were admitted to Guangdong Provincial People's Hospital (Guangzhou, China) from April 2021 to July 2022 and met the following inclusion criteria: (i) confirmed diagnosis of SARDs, included systemic lupus erythematosus (SLE), systemic vasculitis, rheumatoid arthritis(RA), dermatomyositis(DM), primary Sjögren's syndrome, immunoglobulin G4-related disease, Adult-onset Still's disease (AOSD), and mixed connective tissue disease (MCTD). (ii) long-term use of immunosuppressants or corticosteroids, >0.5 mg/kg/day, >1 month; (iii) prior bronchoscopy and collection of bronchoalveolar lavage fluid (BALF); (iv) pathogen detection by mNGS and CMTs with bacterial and fungal smears and cultures; and (v) suspected pulmonary infection confirmed by radiographic images, CT signs included ground glass opacity (GGO), nodules, inflatable sign, parenchymal opacification, reticular or linear shadow, interstitial pneumonia (17).

Data collection

Demographic and clinical data, including age, sex, type of rheumatic disease, use of steroids and immunosuppressants, underlying illness, chest images, results of CMTs and mNGS, changes to antibiotic therapy, and disease regression data, were obtained from electronic medical records.

CMTs

Blood and BALF samples were obtained from all patients. All bronchoalveolar lavage procedures were performed following standard safety protocols. Cultures and smears of the BALF and sputum samples were immediately prepared. All blood samples were immediately assayed by PCR for detection of human immunodeficiency virus (HIV), Epstein–Barr virus (EBV), cytomegalovirus (CMV), and herpes simplex virus in addition to serological analysis of immunoglobulin G, galactomannan, and *Mycobacterium tuberculosis* infection.

mNGS detection using BALF samples

The mNGS procedure for BALF samples included nucleic acid extraction, library construction, sequencing, and bioinformatics analysis. When analyzing and extracting respiratory DNA samples, the Jinshi MicroDNA Kit is mainly used for extraction, followed by DNA fragmentation, splicing, biomolecular labeling (barcode) amplification, and onboard sequencing. The quality of the library is achieved through Qubit® one × dsDNA HS assay was used for evaluation, and then evaluated using Qubit4.0 fluorometer. After the library was quantified by real-time PCR, the MGISEQ-200RS high-throughput sequencing platform was used to sequence the library by the shotgun method. It usually takes 3 days to complete the mNGS analysis and publish the report. The mNGS results were interpreted in reference to the criteria used in previous reports of mNGS for the identification of clinically relevant microorganisms (CRMs) with the Beijing Genomics Institute NGS (BGISEQ-500) platform (15, 18). Bacteria (mycobacteria excluded), fungi (molds excluded), viruses, and parasites with relative abundances at the species level of >30% were considered CRMs. Mycobacteria with a strict mapping read number (SMRN) at the species level of >3 and those with evidence of pulmonary pathogenicity and a SMRN at the species level of >10 were also considered CRMs.

Clinical composite diagnosis as the reference standard

Clinical composite diagnostic criteria were used to identify the pathogens, as determined by two experienced clinicians based on epidemiology, clinical presentation, treatment outcome, laboratory findings, and chest radiology. A true positive was defined as consistency between the results of mNGS and CMTs, while a false positive was defined as detection by mNGS, but not considered pathogenic by the gold standard. If two experts could not reach an agreement, in-depth discussions were held with a third expert.

Statistical analysis

The accuracy, sensitivity, specificity, positive predictive value, and negative predictive value (NPV) of the data were calculated and compared. The chi-square test and Fisher exact test were conducted to identify differences between the results of mNGS and CMTs. All statistical analyses were performed with IBM SPSS Statistics for Windows, version 26.0 (IBM Corporation, Armonk, NY, USA). A probability (p) value <0.05 was considered statistically significant.

Results

General characteristic

The study cohort comprised 40 patients (15 males and 25 females; mean age, 50.82 years) with a confirmed pulmonary infection. The clinical characteristics of the patients are shown in Table 1. All 40 patients had SARDs, which included 22 (55%) with either systemic lupus erythematosus or systemic vasculitis, and were treated with

TABLE 1 The baseline of patients with pneumonia infection.

Characteristic	Clinical value
Age(avg, y)	50.82 ± 16.06
Sex	
Male	37.5% (15/40)
Female	62.5% (25/40)
Primary disease	
SLE	30.0% (12/40)
Systemic vasculitis	25.0% (10/40)
Sjogren's syndrome	12.5% (5/40)
Dermatomyositis	15.0% (6/40)
Other disease	17.5% (7/40)
Immunosuppressant	
Azathioprine	12.5% (5/40)
Cyclophosphamide	35% (14/40)
Mycophenolatemofetil	25% (10/40)
Methotrexate	22.5% (9/40)
Biological agents	10% (4/40)
Leflunomide	2.5% (1/40)
Cyclosporine	15% (6/40)
Corticosteroid	95% (38/40)
Diabetes	
Yes	7.5% (3/40)
No	92.5% (37/40)
Symtom	
Fever	37.5% (15/40)
Cough	25% (10/40)
Anhelation	22.5% (9/40)
Diarrhea	7.5% (3/40)
Asymptomatic	7.5% (3/40)
Lab tests	
WBC(×109/L)	8.63 ± 0.62
Neutrophils%	82.25 (69.58, 89.20)
Lymphocyte(×109/L)	0.83 (0.52,1.18)
Lymphocyte%	0.09 (0.06, 0.18)
Hb(g/L)	101.60 ± 4.08
Erythrocyte sedimentation rate(mm/h)	35.50 (7.25, 61.25)
CRP(mg/L)	22.20 (4.53, 46.53)
PCT(ng/L)	0.09 (0.05, 0.47)
CREA(μmol/L)	77.15 (47.03, 117.25)
ALT(U/L)	16.00 (10.25, 34.50)
Outcomes	
Improved	62.5% (25/40)
Deteriorate	12.5% (5/40)
Loss to follow-up	22.5% (9/40)
Died	2.5% (1/40)

SLE, systemic lupus erythematosus; WBC, white blood cell; Hb, hemoglobin; CRP, C-reaction protein; PCT, procalcitonin; CREA, creatinine; ALT, alanine aminotransferase.

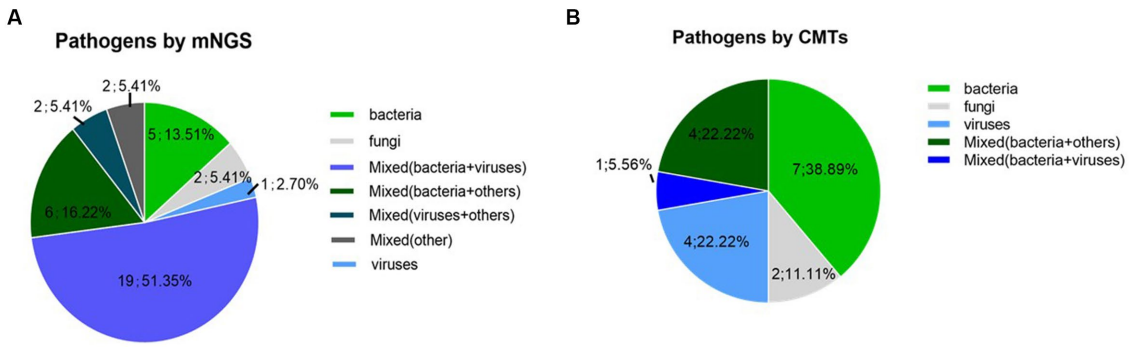

FIGURE 1
Pathogen detection using mNGS and conventional methods **(A)** Detection of mNGS, **(B)** Detection of conventional methods.

immunosuppressants or hormone therapy. The most commonly used immunosuppressants were cyclophosphamide (14/40, 35%), mycophenolate mofetil (10/40, 25%), and methotrexate (9/40, 22.5%).

Diagnostic performance of mNGS vs. CMTs

The pathogens identified by CMTs and mNGS among the 40 patients are shown in Figure 1. Of the 40 patients, 37 (92.5%) had confirmed pulmonary infections by NGS. The isolated pathogens included 36 bacterial, 12 fungal, and 10 viral species. The most common bacterial, fungal, and viral pathogens were *Pseudomonas aeruginosa*, P. jirovecii, and CMV, respectively (Figure 2). The positivity rate of CMTs was significantly lower than that of mNGS (42.5 vs. 92.5%, respectively, $p < 0.01$). The CMTs detected seven bacteria, five fungi (two Cryptococcus), and three viral species.

Consistency of mNGS with CMTs

Of the 40 cases, 16 (40%) were positive for pathogens by both mNGS and CMTs, 21 (52.5%) by only mNGS, and 2 (5%) by only CMTs, while 1 (2.5%) was negative by both mNGS and CMTs (Figure 3). Of 15 double-positive cases, 1 (6.7%) had an exact match between mNGS and CMTs, 2 (13.3%) had a clear mismatch, and 12 (80%) were partial matches, meaning that at least one pathogen was detected by both the CMTs and mNGS.

Diagnostic sensitivity and specificity of mNGS vs. CMTs

For the identification of bacteria, the diagnostic sensitivity and specificity of the CMTs and mNGS were 35%(11/31)/97%(30/31) and 89%(8/9)/89%(8/9), respectively. Notably, the diagnostic accuracy of mNGS for the identification of bacteria was significantly greater than that of the CMTs (95% vs. 53%, respectively, $p < 0.01$). For the detection of fungi, the diagnostic sensitivity and specificity of the CMTs and mNGS were 23%/95% and 94%/100%, respectively. Similarly, the diagnostic accuracy of mNGS for the identification of fungi was significantly greater than that of the CMTs (97.5 vs. 55%, respectively, $p < 0.01$).

Performance of mNGS vs. CMTs for detection of co-infections

Of the 40 patients, 4 (10%) were infected with two pathogens and 27 (67.5%) with three or more. The co-infection detection rate was significantly higher with mNGS than with the CMTs (95 vs. 16%, respectively, $p < 0.01$), and mNGS detected a greater variety of pathogens. Coinfections with fungal-viral-bacterial and bacterial-viral species were the most common. The most common co-infection pathogens involved were P. jirovecii (12/40, 30%) and CMV (13/40, 32.5%).

Diagnostic accuracy of mNGS vs. CMTs

Of the 40 patients, 37 (92.5%) were diagnosed with pneumonia and 3 (7.5%) with non-infectious diseases, of which two were considered primary diseases, and one was an asthma attack. Of the 37 patients diagnosed with pneumonia, 2 (5.4%) had negative mNGS results. All 3 (100%) patients diagnosed with non-infectious diseases had false positive mNGS results, demonstrating sensitivity of 95% (95% confidence interval [CI] = 80–99%), specificity of 33% (95% CI = 1.8–87%), NPV of 33%, and accuracy of 90% (95% CI = 81–99%). In contrast, CMTs misdiagnosed infection in 19 (51.4%) of 37 pneumonia patients, which included one false positive for *P. aeruginosa* infection, demonstrating sensitivity of 47% (95% CI = 30–64%), specificity of 75% (95% CI = 22–99%), NPV of 14%, and accuracy of 50% (95% CI = 35–65%). The NPV and diagnostic accuracy of mNGS were superior to those of the CMTs (Table 2).

Clinical impact of mNGS on diagnosis and treatment

The antibiotic regimen was adjusted based on the mNGS results in 27 (67.5%) of the 40 patients. However, adjustments were abandoned in 9 (22.5%) patients due to compliance with the original antibiotic regimen. Of the 27 cases requiring adjustment of the antibiotic regimen: drugs were added in 19 (70.3%) cases, changes were made in 5 (18.5%) cases, and drugs were discontinued in 3 (11.1%) cases (Figure 4). Sulfamethoxazole was the most commonly adjusted drug before and after the detection of pathogenic bacteria with the mNGS (Table 3). Outcomes were improved in 25 (62.5%) of the 40 patients.

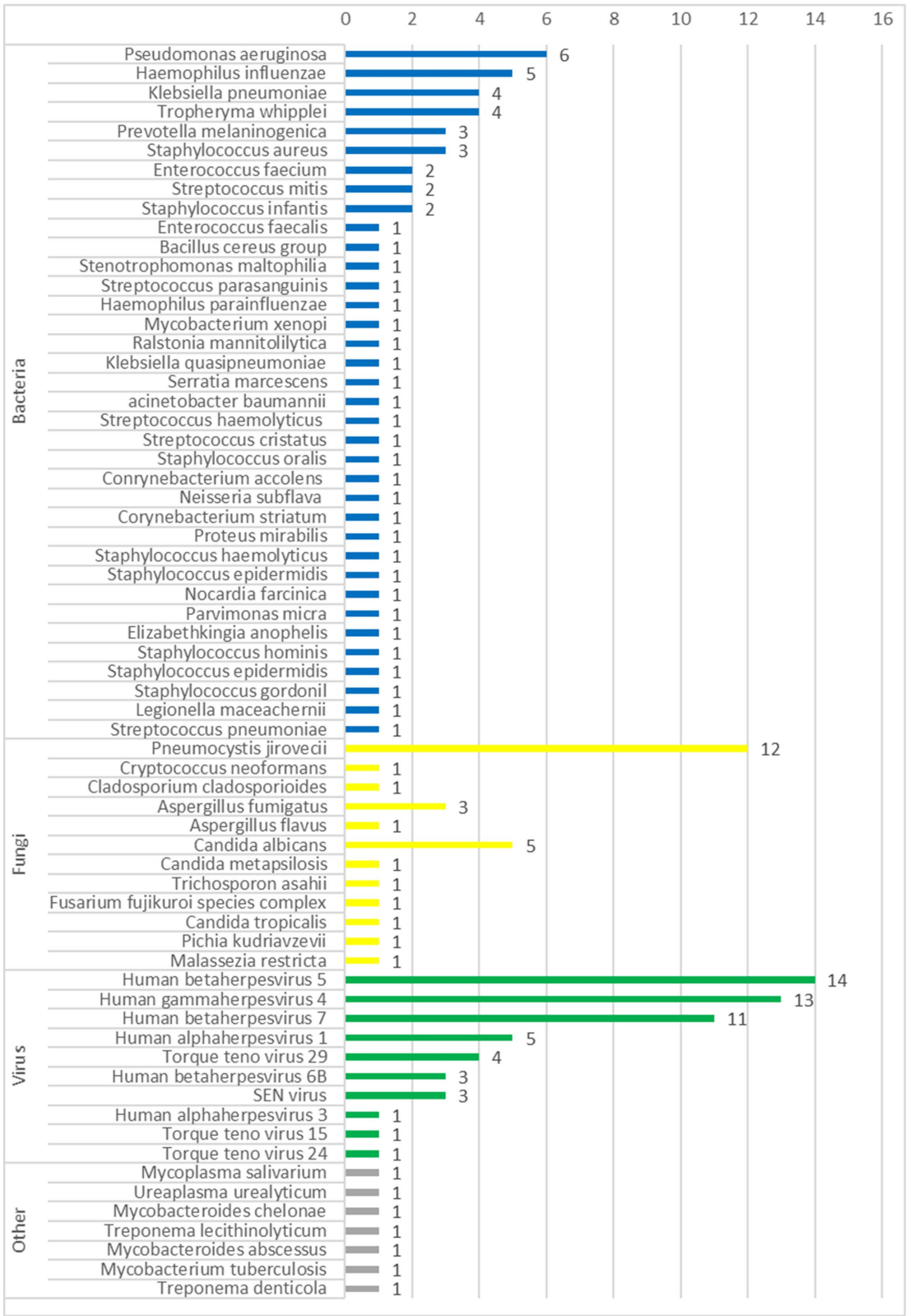

FIGURE 2
The number of pathogens of all cases detected by mNGS.

Discussion

Infection is the leading cause of death in patients with SARDs, and the lungs are the most common site of infection among those receiving immunosuppressant therapy. However, relatively few studies have investigated the use of mNGS for the analysis of BALF samples from patients with SARDs while receiving corticosteroids or immunosuppressants. The present study aimed to compare mNGS vs. CMTs for the identification of pathogens in BALF samples from patients with SARDs treated with corticosteroids or immunosuppressants. Of the 40 patients, pathogens were detected in the BALF samples of 37(92.5%) by mNGS, 18 (45%) by CMTs, and 16(40%) by both methods, while 1 (2.5%) was negative by both mNGS and CMTs. In addition, co-infections were detected in 31 (77.5%) patients by mNGS as compared to only 4 (10%) by CMTs, which led to adjustments of antibiotic regimens for 27 (67.5%) patients. Notably, symptoms improved in 25 (62.5%) of the 40 patients. By comparing mNGS and CMTs, we found that mNGS had the characteristics of high sensitivity and accuracy in pathogen detection, especially for the diagnosis of fungal, viral, specific pathogenic infections, and co-infection, which helped in clinical decision-making and improved prognosis, and was consistent with the findings of a previous report (19).

Among the 40 patients enrolled in this study, 37 (92.5%) received antibiotic treatment before the collection of the BALF samples. Because the detection rate of CMT pathogens is very low, the test results are uncertain. Because CMT has a low detection rate of pathogens, its detection results are unreliable. The results of CMTs usually require 3–5 days as compared to 1 day for mNGS. Compared to CMTs, the results of mNGS are less affected by antibiotics (20). Meanwhile, a wider spectrum of pathogens can be detected by mNGS, which facilitates diagnosis and treatment timely.

In total, 65 pathogens were identified by mNGS and 19 by CMTs. Most of the identified pathogens were bacteria, and only 3 (4.6%) were fungi. The bacteria identified in BALF samples included Haemophilus influenza, *Streptococcus pneumoniae*, and *Pseudomonas aeruginosa*, in addition to the opportunistic pathogens Trophozoites whipplei and *Prevotella melaninogenica*, which are difficult to detect by CMTs.

The most common pathogens detected by mNGS were CMV, EBV, and P. jirovecii. Of the 15 cases of PJP in this study, 14 (93.3%) were diagnosed by mNGS, and only 1 (6.7%) was confirmed by CMTs. Although PJP is traditionally detected by immunofluorescence staining and PCR, the positivity rate of mNGS for the detection of PJP is superior to that of CMTs (21). Moreover, mNGS was comparatively superior for detection of CMV, EBV, and other opportunistic pathogens. Although less virulent in healthy hosts, opportunistic pathogens can cause severe and frequent infections in patients receiving immunosuppressants (22).

In this study, the detection rate of multiple pathogens by mNGS was superior to that by CMTs (72.5% [29/40] vs. 12.5% [5/40], respectively, $p < 0.05$), demonstrating that mNGS is more suitable for detection of mixed infections (23). Deficiencies of CMTs, such as sensitivity to fungi, viruses, and specific pathogens, in addition to interference by empirical antibiotics can be compensated by mNGS (24, 25).

The treatment regimens of 67.5% (27/40) of patients were adjusted based on the results of mNGS and CMTs, which ultimately resulted in improved outcomes for 62.5% of the cohort. Notably, 10 (25%) patients were diagnosed with PJP based on the mNGS results. Of these cases, the antibiotic regimen was terminated or switched to sulfamethoxazole in 7 (70%). These results suggest that mNGS can more accurately identify pathogens causing pulmonary changes, thereby avoiding overuse of antibiotics.

There were some limitations to this study that should be addressed. First, it's difficult to obtain BALF from patients with SARDs as a control group when bronchoscopy is not performed in the patient without lesion in the lung. Second, the patients were recruited from a single medical center, which might have introduced bias. Third, it was difficult to determine whether the pathogens detected by mNGS were collateral contaminants, opportunistic colonizers, or causative pathogens because all patients received immunosuppressant therapy for SARDs. Fourth, there may have been some false positives of patients with mixed infections. Therefore, mNGS requires further development to accurately identify multiple types of pathogens. Finally, some patients did not follow up with outpatient clinics or examinations within the

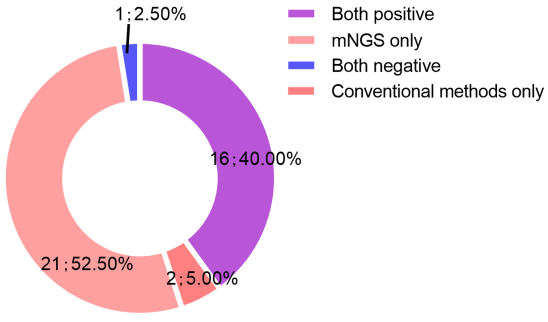

FIGURE 3
Consistency between mNGS and CMTs.

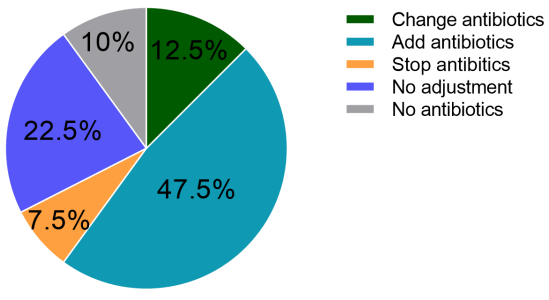

FIGURE 4
Adjustment of antibiotic drug.

TABLE 2 The sensitivity and specificity of mNGS and CMTs.

Assay	Sensitivity, % (95%CI)	Specificity, % (95%CI)	PPV (%)	NPV (%)	PLR	NLR
mNGS	94.59 (0.80 ~ 0.99)	33.33 (0.02 ~ 0.87)	94.59	33.33	1.42	0.16
CMTs	33.33 (0.19 ~ 0.51)	75.00 (0.22 ~ 0.99)	93.33	14.00	1.33	0.89

TABLE 3 Comparison of the proportion of antibiotic use before and after adjustment according to mNGS.

Antimicrobials	Pre-mNGS	After-mNGS	*p* value
Voriconazole	5/40 (12.5%)	8/40 (20%)	0.363
Cefoperazone sodium and sulbactam sodium	8/40 (20%)	7/40 (17.5%)	0.775
Piperacillin sodium tazobactam sodium	3/40 (7.5%)	7/40 (17.5%)	0.310
Moxifloxacin	10/40 (25.00%)	10/40 (25%)	1
Cefuroxime	1/40 (2.50%)	0/40 (0.00%)	1
Cefdizine	0/40 (0.00%)	1/40 (2.5%)	1
Fluconazol	2/40 (5.00%)	3/40 (7.5%)	1
Levofloxacin	1/40 (2.5%)	2/40 (5.00%)	1
Sulfamethoxazole	2/40 (5.00%)	12/40 (30.0%)	0.008
Ertapenem	1/40 (2.50%)	1/40 (2.50%)	1
Ceftazidime	0/40 (0.00%)	1/40 (2.50%)	1
Amphotericin	0/40 (0.00%)	1/40 (2.50%)	1
Caspofungin	0/40 (0.00%)	1/40 (2.50%)	1
imipenem and cilastatin sodium	3/40 (7.50%)	2/40 (5.00%)	1
Linezolid	1/40 (2.50%)	2/40 (5.00%)	1
Cefotaxime sodium	0/40 (0.00%)	1/40 (2.50%)	1
Rifaximin	1/40 (2.50%)	1/40 (2.50%)	1
Metronidazole	1/40 (2.50%)	1/40 (2.50%)	1
Meropenem	2/40 (5.00%)	2/40 (5.00%)	1

McNemar test for table.

prescribed time, resulting in a low follow-up rate. In addition, hospital phone numbers were classified as high-frequency nuisance numbers, making follow-up visits more difficult.

In conclusion, mNGS were significantly better than CMTs for detection of suspected pulmonary infections in patients with SARDs receiving immunosuppressant therapy. Rapid diagnosis by mNGS can ensure timely adjustment of treatment regimens to improve diagnosis and outcomes.

Data availability statement

The data presented in the study are deposited in the NCBI repository, online at https://www.ncbi.nlm.nih.gov/bioproject/PRJNA979827.

Ethics statement

The studies involving human participants were reviewed and approved by the Clinical Trials and Biomedical Ethics Committee of Guangdong Provincial People's Hospital [KY-H-2022-057-03]. The patients/participants provided their written informed consent to participate in this study.

Author contributions

SW, SLiu, and LX were responsible for the overall design and investigation. SW, SP, and SLiu were responsible for manuscript writing and participated in discussion of the results. XH, NJ, BL, BC, SD, YY, QW, SLi, and YT were responsible for data collection and analysis. JM, SLiu, TL, FW, ZL, RH, and ZF were contributed to the patients follow study. CH, WW, XL, and WS were served as lead illustrator of this manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by National Natural Science Foundation of China (Nos. 81870508, 81873616, and 82170730), Science and Technology Planning Project of Guangzhou (Nos. 202102020534 and 20210208040), Guangzhou Municipal Science and Technology Plan Project (Nos. 202102080385 and 201904010026), Natural Science Foundation of Guangdong Province (Nos. 2022A1515012374 and 2023A1515010024), Young Talent Training Program of Guangdong Provincial Association for S&T (No. SKXRC202222), Guangdong Basic and Applied Basic Research Foundation (No. 2021A1515220150), and Guangdong Province High-level Hospital Construction Project (No. DFJH201901).

Conflict of interest

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

Publisher's note

References

1. Gabriel SE, Michaud K. Epidemiological studies in incidence, prevalence, mortality, and comorbidity of the rheumatic diseases. *Arthritis Res Ther*. (2009) 11:229. doi: 10.1186/ar2669

2. Falagas ME, Manta KG, Betsi GI, Pappas G. Infection-related morbidity and mortality in patients with connective tissue diseases: a systematic review. *Clin Rheumatol*. (2007) 26:663–70. doi: 10.1007/s10067-006-0441-9

3. Broen JCA, van Laar JM. Mycophenolate mofetil, azathioprine and tacrolimus: mechanisms in rheumatology. *Nat Rev Rheumatol*. (2020) 16:167–78. doi: 10.1038/s41584-020-0374-8

4. Alarcon GS. Infections in systemic connective tissue diseases: systemic lupus erythematosus, scleroderma, and polymyositis/dermatomyositis. *Infect Dis Clin N Am*. (2006) 20:849–75. doi: 10.1016/j.idc.2006.09.007

5. Zhao Y, Lin X. *Cryptococcus neoformans*: sex, morphogenesis, and virulence. *Infect Genet Evol*. (2021) 89:104731. doi: 10.1016/j.meegid.2021.104731

6. Cunnane G, Doran M, Bresnihan B. Infections and biological therapy in rheumatoid arthritis. *Best Pract Res Clin Rheumatol*. (2003) 17:345–63. doi: 10.1016/s1521-6942(02)00107-9

7. Yang A, Chen C, Hu Y, Zheng G, Chen P, Xie Z, et al. Application of metagenomic next-generation sequencing (mNGS) using bronchoalveolar lavage fluid (BALF) in diagnosing pneumonia of children. *Microbiol Spectr*. (2022) 10:e0148822. doi: 10.1128/spectrum.01488-22

8. Shi Y, Peng JM, Qin HY, Du B. Metagenomic next-generation sequencing: a promising tool for diagnosis and treatment of suspected pneumonia in rheumatic patients with acute respiratory failure: retrospective cohort study. *Front Cell Infect Microbiol*. (2022) 12:941930. doi: 10.3389/fcimb.2022.941930

9. Zhang XX, Guo LY, Liu LL, Shen A, Feng WY, Huang WH, et al. The diagnostic value of metagenomic next-generation sequencing for identifying *Streptococcus pneumoniae* in paediatric bacterial meningitis. *BMC Infect Dis*. (2019) 19:495. doi: 10.1186/s12879-019-4132-y

10. Gu W, Deng X, Lee M, Sucu YD, Arevalo S, Stryke D, et al. Rapid pathogen detection by metagenomic next-generation sequencing of infected body fluids. *Nat Med*. (2021) 27:115–24. doi: 10.1038/s41591-020-1105-z

11. Borroni D, Romano V, Kaye SB, Somerville T, Napoli L, Fasolo A, et al. Metagenomics in ophthalmology: current findings and future prospectives. *BMJ Open Ophthalmol*. (2019) 4:e000248. doi: 10.1136/bmjophth-2018-000248

12. Borroni D, Bonzano C, Sánchez-González JM, Rachwani-Anil R, Zamorano-Martín F, Pereza-Nieves J, et al. Shotgun metagenomic sequencing in culture negative microbial keratitis. *Eur J Ophthalmol*. (2023):11206721221149077. doi: 10.1177/11206721221149077

13. Parekh M, Borroni D, Romano V, Kaye SB, Camposampiero D, Ponzin D, et al. Next-generation sequencing for the detection of microorganisms present in human donor corneal preservation medium. *BMJ Open Ophthalmol*. (2019) 4:e000246. doi: 10.1136/bmjophth-2018-000246

14. Sun L, Zhang S, Yang Z, Yang F, Wang Z, Li H, et al. Clinical application and influencing factor analysis of metagenomic next-generation sequencing (mNGS) in ICU patients with sepsis. *Front Cell Infect Microbiol*. (2022) 12:905132. doi: 10.3389/fcimb.2022.905132

15. Peng JM, Du B, Qin HY, Wang Q, Shi Y. Metagenomic next-generation sequencing for the diagnosis of suspected pneumonia in immunocompromised patients. *J Infect*. (2021) 82:22–7. doi: 10.1016/j.jinf.2021.01.029

16. Zhang X, Qin Y, Lei W, Huang JA. Metagenomic next-generation sequencing of BALF for the clinical diagnosis of severe community-acquired pneumonia in immunocompromised patients: a single-center study. *Exp Ther Med*. (2023) 25:178. doi: 10.3892/etm.2023.11877

17. Hansell DM, Bankier AA, MacMahon H, McLoud TC, Müller NL, Remy J. Fleischner society: glossary of terms for thoracic imaging. *Radiology*. (2008) 246:697–722. doi: 10.1148/radiol.2462070712

18. Li Y, Sun B, Tang X, Liu Y-l, He H-y, Li X-y, et al. Application of metagenomic next-generation sequencing for bronchoalveolar lavage diagnostics in critically ill patients. *Eur J Clin Microbiol Infect Dis*. (2020) 39:369–74. doi: 10.1007/s10096-019-03734-5

19. Wang Q, Wu B, Yang D, Yang C, Jin Z, Cao J, et al. Optimal specimen type for accurate diagnosis of infectious peripheral pulmonary lesions by mNGS. *BMC Pulm Med*. (2020) 20:268. doi: 10.1186/s12890-020-01298-1

20. Miao Q, Ma Y, Wang Q, Pan J, Zhang Y, Jin W, et al. Microbiological diagnostic performance of metagenomic next-generation sequencing when applied to clinical practice. *Clin Infect Dis*. (2018) 67:S231–40. doi: 10.1093/cid/ciy693

21. Liu L, Yuan M, Sun S, Wang J, Shi Y, Yu Y, et al. The performance of metagenomic next-generation sequence in the diagnosis of suspected opportunistic infections in patients with acquired immunodeficiency syndrome. *Infect Drug Resist*. (2022) 15:5645–53. doi: 10.2147/idr.S378249

22. Gugliesi F, Pasquero S, Griffante G, Scutera S, Albano C, Pacheco SFC, et al. Human cytomegalovirus and autoimmune diseases: where are we? *Viruses*. (2021) 13(2):260. doi: 10.3390/v13020260

23. Wang J, Han Y, Feng J. Metagenomic next-generation sequencing for mixed pulmonary infection diagnosis. *BMC Pulm Med*. (2019) 19:252. doi: 10.1186/s12890-019-1022-4

24. Deng W, Xu H, Wu Y, Li J. Diagnostic value of bronchoalveolar lavage fluid metagenomic next-generation sequencing in pediatric pneumonia. *Front Cell Infect Microbiol*. (2022) 12:950531. doi: 10.3389/fcimb.2022.950531

25. Su R, Yan H, Li N, Ding T, Li B, Xie Y, et al. Application value of blood metagenomic next-generation sequencing in patients with connective tissue diseases. *Front Immunol*. (2022) 13:939057. doi: 10.3389/fimmu.2022.939057

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 31 October 2023
DOI 10.3389/fmed.2023.1281880

OPEN ACCESS

EDITED BY
J. Francis Borgio,
Imam Abdulrahman Bin Faisal University, Saudi Arabia

REVIEWED BY
Qiang Tang,
Chengdu University of Traditional Chinese Medicine, China
Jiesi Luo,
Southwest Medical University, China
Ke Han,
Harbin University of Commerce, China
Zhijun Liao,
Fujian Medical University, China

*CORRESPONDENCE
Xiaolong Yu
✉ yuxiaolong@hainanu.edu.cn
Hao Lin
✉ hlin@uestc.edu.cn
Chengbing Huang
✉ 20049607@abtu.edu.cn

†These authors have contributed equally to this work

RECEIVED 23 August 2023
ACCEPTED 16 October 2023
PUBLISHED 31 October 2023

CITATION
Zou X, Ren L, Cai P, Zhang Y, Ding H, Deng K, Yu X, Lin H and Huang C (2023) Accurately identifying hemagglutinin using sequence information and machine learning methods. *Front. Med.* 10:1281880. doi: 10.3389/fmed.2023.1281880

Accurately identifying hemagglutinin using sequence information and machine learning methods

Xidan Zou[1†], Liping Ren[2†], Peiling Cai[3], Yang Zhang[4], Hui Ding[1], Kejun Deng[1], Xiaolong Yu[5*], Hao Lin[1*] and Chengbing Huang[6*]

[1]School of Life Science and Technology, Center for Informational Biology, University of Electronic Science and Technology of China, Chengdu, China, [2]School of Healthcare Technology, Chengdu Neusoft University, Chengdu, China, [3]School of Basic Medical Sciences, Chengdu University, Chengdu, China, [4]Innovative Institute of Chinese Medicine and Pharmacy, Academy for Interdiscipline, Chengdu University of Traditional Chinese Medicine, Chengdu, China, [5]School of Materials Science and Engineering, Hainan University, Haikou, China, [6]School of Computer Science and Technology, Aba Teachers University, Aba, China

Introduction: Hemagglutinin (HA) is responsible for facilitating viral entry and infection by promoting the fusion between the host membrane and the virus. Given its significance in the process of influenza virus infestation, HA has garnered attention as a target for influenza drug and vaccine development. Thus, accurately identifying HA is crucial for the development of targeted vaccine drugs. However, the identification of HA using in-silico methods is still lacking. This study aims to design a computational model to identify HA.

Methods: In this study, a benchmark dataset comprising 106 HA and 106 non-HA sequences were obtained from UniProt. Various sequence-based features were used to formulate samples. By perform feature optimization and inputting them four kinds of machine learning methods, we constructed an integrated classifier model using the stacking algorithm.

Results and discussion: The model achieved an accuracy of 95.85% and with an area under the receiver operating characteristic (ROC) curve of 0.9863 in the 5-fold cross-validation. In the independent test, the model exhibited an accuracy of 93.18% and with an area under the ROC curve of 0.9793. The code can be found from https://github.com/Zouxidan/HA_predict.git. The proposed model has excellent prediction performance. The model will provide convenience for biochemical scholars for the study of HA.

KEYWORDS

hemagglutinin, machine learning, sequence features, feature extraction, stacking

1. Introduction

Influenza is a contagious respiratory disease, posing a significant threat to human health and causing varying degrees of disease burden globally (1, 2). Hemagglutinin (HA), a glycoprotein on the surface of influenza viruses, mediates viral entry and infection by binding to host sialic acid receptors (3). The highly conserved stem or stalk region of HA has been identified as a promising target for the development of a universal influenza vaccine (4). Accurate identification of HA is crucial for targeted vaccine and drug development.

With the increasing maturity of protein sequence coding methods and machine learning algorithms, sequence-based protein recognition has been an effective approach for rapid

identification of protein. It achieves classification and identification of specific proteins using protein sequence coding methods and machine learning algorithms, which has been widely used in the prediction studies of cell-penetrating peptides (5), hemolytic peptide (6), anti-cancer peptides (7), hormone proteins (8), autophagy proteins (9), and Anti-CRISPR proteins (10), etc., because of its high recognition accuracy in the protein identification study.

Despite the pivotal role of HA in influenza virus infection, existing machine learning-based research on HA has primarily focused on influenza virus subtype classification (11, 12), influenza virus host prediction (13), influenza virus mutation and evolution prediction (14), HA structure–function analysis (15), and influenza virus pathogenicity and prevalence prediction (16). However, there are currently no approaches for HA identification based on HA sequence information and machine learning techniques.

In this study, we proposed a machine learning-based prediction model for HA to achieve effective identification. Firstly, we constructed a benchmark dataset based on existing protein databases. Next, we employed feature extraction methods to encode the protein sequences. Subsequently, we fused all the extracted features and utilized the analysis of variance (ANOVA) combined with incremental feature selection (IFS) strategies to obtain the most informative feature subset. Finally, the HA prediction model was developed based on this optimal feature subset. The workflow is shown in Figure 1.

2. Method and materials

2.1. Benchmark dataset

A benchmark dataset is essential for bioinformatics analysis (17, 18). The dataset used in this study was collected from the Universal Protein Resource (UniProt) (19). To ensure the quality of the dataset, several pre-processing steps were performed. Protein sequences containing nonstandard letters (e.g., 'B', 'U', 'X', 'Z') were eliminated. Redundancy removal was done using CD-HIT (20) to remove sequences with high similarity. The cutoff value was set to 80%, and sequences with a similarity higher than 80% were removed. The non-HA dataset was down-sampled to ensure a balanced dataset with equal positive and negative samples. The final benchmark dataset consisted of 212 protein sequences, including 106 HA and 106 non-HA samples. The dataset was randomly split into a training dataset and a test dataset in a 4:1 ratio. The above-mentioned model training set data and test set data are included in https://github.com/Zouxidan/HA_predict.git. At the same time, a dataset named 'predict_data.txt' for testing is also included.

2.2. Feature extraction

Feature extraction plays a crucial role in protein identification and prediction (10, 21–25). However, machine learning algorithms cannot directly process protein sequence information for computation and model construction. Therefore, it is necessary to convert protein sequence information into numerical data that can be understood and utilized by machine learning algorithms (26–29). Here, we employed various methods for feature extraction of protein sequences, including Amino Acid Composition (AAC), Dipeptide Composition (DPC), Tripeptide Composition (TPC), Composition of k-spaced Amino Acid Pairs (CKSAAP), Pseudo-Amino Acid Composition (PseAAC), PseAAC of Distance-Pairs and Reduced Alphabet (DCP). These sequence feature extraction approaches have been widely adopted in the field of bioinformatics (30–32). The implementation of these feature extraction methods was based on iLearnPlus (33).

A protein sequence P of length L can be represented as:

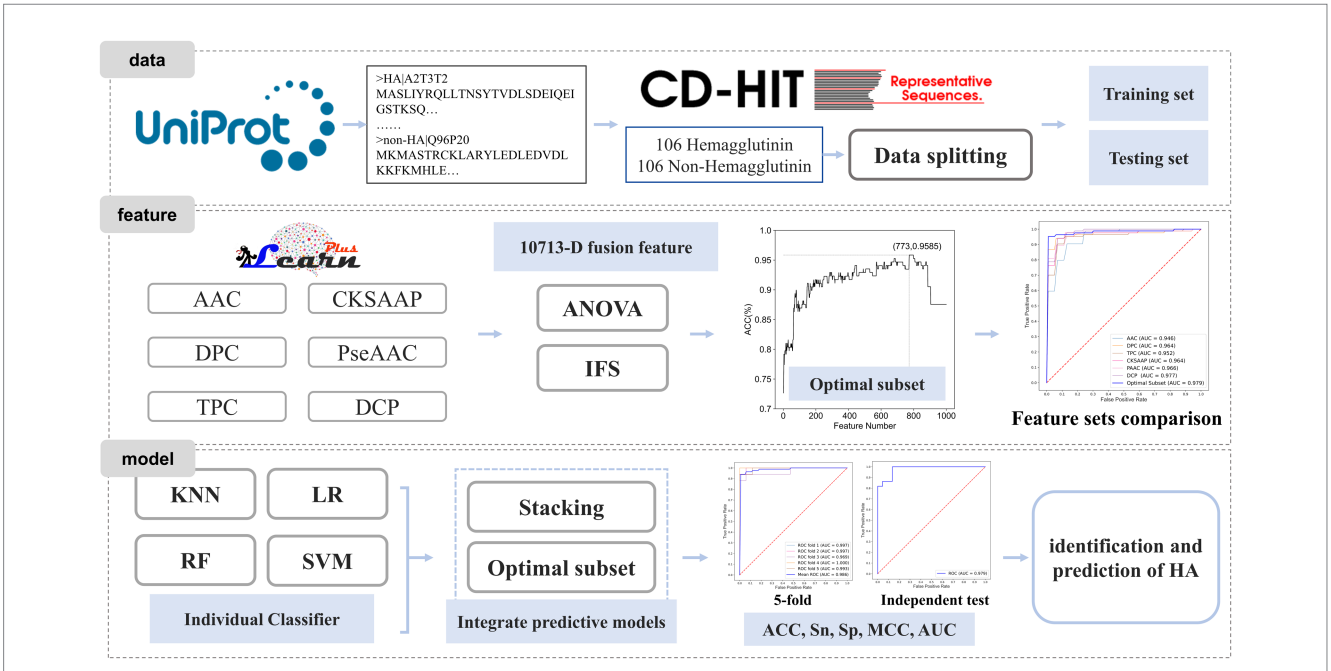

FIGURE 1
Workflow diagram for constructing the HA prediction model.

$$P = R_1R_2R_3R_4R_5R_6\cdots R_L \tag{1}$$

where R_1 denotes the first amino acid of the sequence, R_2 denotes the second amino acid, and so on.

2.2.1. AAC

AAC is a commonly used method for protein sequence feature extraction, which involves 20 feature vectors. AAC was defined as:

$$f_i = \frac{N(a_i)}{\sum_i^{20} N(a_i)} = \frac{N(a_i)}{L} \tag{2}$$

where a_i denotes the i-th natural amino acid and $N(a_i)$ denotes the frequency of amino acid a_i in the protein sequence.

2.2.2. DPC

Similar to AAC, DPC counts the frequency of amino acids, but it focuses on the frequency of two adjacent amino acids in a protein sequence. DPC was defined as:

$$f_{i,j} = \frac{N(a_i,a_j)}{\sum_i^{20}\sum_j^{20} N(a_i,a_j)} = \frac{N(a_i,a_j)}{L-1} \tag{3}$$

where (a_i, a_j) denotes two adjacent amino acids and $N(a_i, a_j)$ denotes the frequency of the amino acid pair (a_i, a_j) in the protein sequence.

2.2.3. TPC

TPC is another feature extraction method that considers the relationship among three adjacent amino acids, providing more protein sequence information compared to AAC and DPC. TPC was defined as:

$$f_{i,j,z} = \frac{N(a_i,a_j,a_z)}{\sum_i^{20}\sum_j^{20}\sum_z^{20} N(a_i,a_j,a_z)} = \frac{N(a_i,a_j,a_z)}{L-2} \tag{4}$$

where (a_i, a_j, a_z) denotes the combination of three adjacent amino acids, and $N(a_i, a_j, a_z)$ denotes the frequency of the tripeptide combination (a_i, a_j, a_z) in the protein sequence.

2.2.4. CKSAAP

To obtain further sequence information, Chen et al. proposed CKSAAP (34) which was defined as:

$$f_{i,j,k} = \frac{N(a_i,x_k,a_z)}{\sum_i^{20}\sum_j^{20} N(a_i,x_k,a_z)} = \frac{N(a_i,x_k,a_z)}{L-k-1} \tag{5}$$

where k denotes the number of amino acids spaced between two amino acids, x_k denotes k arbitrary amino acids, (a_i, x_k, a_j) denotes the spaced amino acid pair, and $N(a_i, x_k, a_j)$ denotes the frequency of the spaced amino acid pair (a_i, x_k, a_j) in the protein sequence.

2.2.5. PseAAC

To incorporate protein sequence ordinal information and improve prediction quality, a powerful feature, called PseAAC, was proposed, which incorporated the physicochemical characteristics of amino acids. PseAAC was defined as:

$$f_i = \begin{cases} \dfrac{x_i}{\sum_{i=1}^{20} x_i + \omega\sum_{j=1}^{\lambda}\theta_j} & (0 < i \le 20) \\ \dfrac{\omega\theta_{i-20}}{\sum_{i=1}^{20} x_i + \omega\sum_{j=1}^{\lambda}\theta_j} & (20 < i \le 20+\lambda) \end{cases} \tag{6}$$

where x_i denotes the normalized amino acid frequency, ω denotes the weight factor for short-range and long-range, and θ_j denotes the j-th sequence correlation factor.

θ_j was calculated as:

$$\theta_j = \frac{1}{L-j}\sum_{i=1}^{L-j}\Theta(R_i + R_{i+j}) \tag{7}$$

$\Theta(R_i + R_{i+j})$ was defined as:

$$\Theta(R_i + R_{i+j}) = \frac{\left[H_1(R_{i+j}) - H_1(R_i)\right]^2 + \left[H_2(R_{i+j}) - H_2(R_i)\right]^2 + \left[M(R_{i+j}) - M(R_i)\right]^2}{3} \tag{8}$$

where $H_1(R_i)$, $H_2(R_i)$, and $M(R_i)$ denote the standardized hydrophobicity, standardized hydrophilicity, and standardized side chain mass of the amino acid R_i, respectively.

The hydrophobicity, hydrophilicity, and side chain mass of amino acids were standardized using the following equations:

$$\begin{cases} H_1(R_i) = \dfrac{H_1^0(R_i) - \sum_{i=1}^{20}\frac{H_1^0(R_i)}{20}}{\sigma(H_1^0)} \\ H_2(R_i) = \dfrac{H_2^0(R_i) - \sum_{i=1}^{20}\frac{H_2^0(R_i)}{20}}{\sigma(H_2^0)} \\ M(R_i) = \dfrac{M^0(R_i) - \sum_{i=1}^{20}\frac{M^0(R_i)}{20}}{\sigma(M^0)} \end{cases} \tag{9}$$

where $H_1(R_i)$, $H_2(R_i)$, and $M(R_i)$ denote the standardized hydrophobicity, standardized hydrophilicity, and standardized side chain mass of amino acids, respectively, and $H_1^0(R_i)$, $H_2^0(R_i)$, and $M^0(R_i)$ denote the corresponding raw physicochemical properties of amino acids.

2.2.6. DCP

To incorporate more protein sequence order information and reduce the impact of high-dimensional features, Liu et al. proposed

DCP (35). Based on a validated amino acid simplification alphabet scheme (36), three simplified amino acid alphabets were defined as:

$$\begin{cases} cp(13) = \begin{Bmatrix} MF; IL; V; A; C; WYQHP; G; T; S; N; \\ RK; D; E \end{Bmatrix} \\ cp(14) = \begin{Bmatrix} IMV; L; F; WY; G; P; C; A; S; T; N; \\ HRKQ; E; D \end{Bmatrix} \\ cp(19) = \begin{Bmatrix} P; G; E; K; R; Q; D; S; N; T; H; C; I; \\ V; W; YF; A; L; M \end{Bmatrix} \end{cases} \quad (10)$$

For any simplified amino acid alphabet, DCP was defined as:

$$f_{i,j,z} = \frac{N_d(cp_i, cp_j)}{\sum_i^z \sum_j^z N_d(cp_i, cp_j)} \quad (11)$$

where z denotes the number of amino acid clusters in the simplified alphabet, and $N_d(cp_i, cp_j)$ denotes the frequency of any two amino acid clusters with distance d in the protein sequence.

In this study, the following parameters were used for protein sequence feature extraction: $k = 1$ for CKSAAP (amino acid spacing value), $\lambda = 10$ for PseAAC (number of amino acid theoretical properties), and $\omega = 0.7$ for the weight factor for short-range and long-range. Consequently, we extracted features that include 20-dimensional AAC, 400-dimensional DPC, 8000-dimensional TPC, 800-dimensional CKSAAP, 30-dimensional PseAAC, and 1,463-dimensional DCP.

2.3. Feature fusion and selection

Different feature extraction methods offer diverse interpretations and representations of protein sequences. Relying solely on a single feature extraction method may limit the information provided by a single feature. To obtain a more comprehensive and reliable interpretation of protein sequences, we fused all features to create a fused feature set, resulting in a 10,713-dimensional feature set ($20 + 400 + 8{,}000 + 800 + 30 + 1{,}463$). We then selected the optimal feature subset using ANOVA and IFS.

ANOVA, a widely used feature selection tool, tests the difference in means between groups to determine whether the independent variable influences the dependent variable. Its high accuracy has made it an effective choice for feature selection (8). For a feature f, its F-value was calculated based on the principle of ANOVA as follows:

$$F(f) = \frac{SSA / (K-1)}{SSE / (N-K)} \quad (12)$$

where $F(f)$ represents the F-value of feature f, SSA represents the sum of squares between groups, SSE represents the sum of squares within groups, K-1 and N-K denote the degrees of freedom between and within groups, respectively. N is the total number of samples, and K is the number of groups.

SSA and SSE were calculated as follows:

$$\begin{cases} SSA = \sum_{i=1}^{K} \sum_{j=1}^{k_i} \left(f(i,j) - \frac{\sum_{j=1}^{k_i} f(i,j)}{k_i} \right)^2 \\ SSE = \sum_{i=1}^{K} k_i \left(\frac{\sum_{j=1}^{k_i} f(i,j)}{k_i} - \frac{\sum_{i=1}^{K} \sum_{j=1}^{k_i} f(i,j)}{\sum_{i=1}^{K} k_i} \right)^2 \end{cases} \quad (13)$$

where $f(i,j)$ denotes the j-th feature of the i-th group, K represents the number of groups, and k_i represents the total number of samples in the i-th group.

A larger F-value indicates a stronger influence of the feature on data classification, thereby contributing more to the data classification results. In the feature set, the large amount of data, redundant data and noise will not only result in higher computational costs, but also cause the phenomenon of overfitting or reduced accuracy of the prediction model. The above fusion feature set contains 10,713 features, which is a large number of features. For saving computational time and reducing computational cost, we firstly use ANOVA to initially filter to obtain the 1,000 features which have the greatest influence on the classification results.

Next, the optimal subset of features was determined by searching the top 1,000 features ranked by F-value using IFS. IFS is a frequently employed feature selection method in the field of bioinformatics (37, 38). The specific process of IFS is as follows. Firstly, all features were sorted in descending order according to their F-values obtained from ANOVA. Then, each feature was sequentially added to the feature set, and a model was constructed using support vector machine (SVM) for each newly formed feature subset. Grid search was utilized to obtain optimal models, and their performance was evaluated using 5-fold cross-validation. The optimal feature subset was defined as the set of features that maximized the model's accuracy.

2.4. Machine learning methodology and modeling

The advancement of machine learning has provided an effective approach to solving biological problems (39–42). Utilizing machine learning techniques to identify proteins based on sequence features has proven to be a rapid and widely applied method in various studies (43–45).

Constructing appropriate models is crucial for achieving accurate and robust predictions. In this study, we selected four commonly used machine learning algorithms, namely K-nearest neighbor (KNN) (46), logistic regression (LR) (47), random forest (RF) (48), and SVM (49), to build the fundamental classifier model for the HA dataset. The optimal parameters for each algorithm were obtained using grid search. To further enhance the model's accuracy and generalization ability, we developed an integrated classifier model by combining the four basic classifier models. The Stacking algorithm was employed, with logistic regression serving as the second-layer classifier. All the machine learning models utilized in this study were implemented using scikit-learn (50).

KNN is a simple yet effective machine learning algorithm based on the implementation of the distance between data and data. LR is a binary classification algorithm based on the sigmoid function,

which classifies samples by their corresponding output values. In RF, the result of prediction is determined by the vote or average of decision trees. The basic principle of SVM is to separate two classes of training data by defining a hyperplane and maximizing the distance between the two classes.

The Stacking algorithm is one of the widely used integrated learning methods, which obtains predictive models with higher accuracy and better generalization ability by combining basic classifier models. The Stacking algorithm was initially proposed by Wolpert (51). Its basic idea is to obtain an optimal integrated classifier model by training and combining multiple basic classifier models. In the Stacking algorithm, machine learning algorithms with strong learning and fitting capabilities are frequently used to construct basic classifier models for adequate learning and interpretation of training data. To reduce the degree of overfitting, simple algorithms with strong interpretations are commonly used to construct integrated classifier models.

2.5. Performance evaluation

To assess the effectiveness of the constructed models, we employed 5-fold cross-validation and independent testing. The performance of the proposed model was evaluated using several metrics, including accuracy (*ACC*), sensitivity (*Sn*), specificity (*Sp*), Matthew's correlation coefficient (*MCC*), and the area under the receiver operating characteristic curve (*AUC*) (27, 52–56). *ACC*, *Sn*, *Sp*, and *MCC* were expressed as:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

$$Sn = \frac{TP}{TP + FN} \tag{15}$$

$$Sp = \frac{TN}{TN + FP} \tag{16}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FN)(TP + FP)(TN + FP)(TN + FN)}} \tag{17}$$

where *TP*, *TN*, *FP*, and *FN* represent the following respectively: correctly identified positive samples, correctly identified negative samples, incorrectly identified negative samples, and incorrectly identified positive samples.

Additionally, we utilized the receiver operating characteristic (ROC) curve to evaluate model performance. A higher *AUC* value indicates better model performance, as it reflects the proximity to 1 according to the underlying principle.

3. Results and discussion

3.1. Optimal feature subset

We constructed optimal feature subsets using ANOVA and IFS and evaluated the models for each subset using the *ACC*. Figure 2A shows the IFS curve for the fusion feature set. When the feature set contained 773 features, the prediction model achieved a maximum *ACC* value of 0.9585.

In optimal feature subset, 13-dimensional AAC, 73-dimensional DPC, 629-dimensional TPC, 29-dimensional CKSAAP, and 29-dimensional DCP features are included. Notably, PseAAC is not included in this subset, suggesting that it is less effective in classifying HA compared to the other features. Furthermore, TPC has the highest proportion in optimal feature subset, indicating that TPC provides the best identification and differentiation ability among the six methods for feature extraction.

To demonstrate the impact of optimal feature subsets on model performance, we compared the performance of SVM prediction models constructed with optimal feature subsets to those constructed with six single feature sets. Each model was optimized using grid search within the same parameter range, and all models were evaluated using 5-fold cross-validation. Table 1 presents the results of the comparison, and Figure 2B shows the ROC curves for the 5-fold cross-validation of these models. The model constructed with the optimal feature subset achieved an *ACC* of 94.06% and an *AUC* of 0.970, outperforming the models constructed with other single feature sets. These results indicate that the optimal feature subset significantly improved the model's prediction performance.

3.2. Model construction and evaluation

We constructed four basic models and an integrated model based on the optimal feature subsets. The optimal parameters for each algorithm were as follows: $K = 52$ for KNN, $n = 62$ for RF, $f = 6$ for the number of features considered during best-split search, $\xi = 4$ for the SVM kernel parameter, and $C = 32$ for the regularization parameter.

Table 2 presents the performance comparison of different classifier models using two testing methods. Figures 2C,D show the ROC curve of the constructed integration model using these wo testing methods. With 5-fold cross-validation, the proposed integrated model achieved an *ACC* of 95.85% and an *AUC* of 0.9863. On the independent test set, the integrated model achieved an *ACC* of 93.18% and an *AUC* of 0.9793. These results demonstrate that the proposed integrated model exhibited better HA prediction capability, improved model performance, and enhanced generalization ability compared to a single model.

3.3. Comparison of other machine learning algorithms

We have created two models based on optimal feature subsets and compared their performance to demonstrate the superiority of our proposed model. The comparison results are presented in Table 3, where we compared the model constructed with the XGboost algorithm with our proposed model. The main parameters of the model constructed based on the XGboost algorithm are as follows: *max_depth* = 3, *learning_rate* = 0.16, *colsample_bytree* = 0.85, *subsample* = 0.75. The results in Table 3 show that our model has good classification performance.

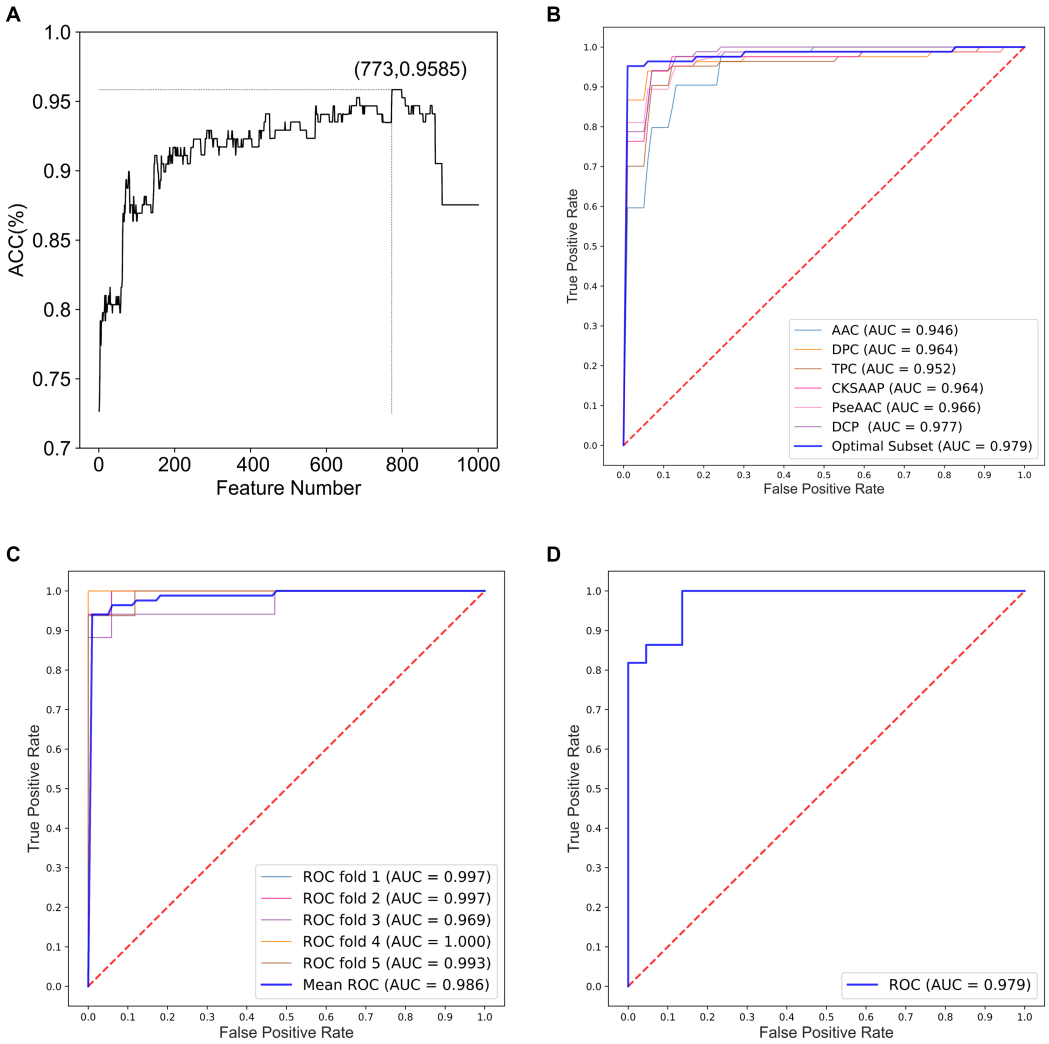

FIGURE 2
Performance analysis for optimal feature subsets and HA prediction models. **(A)** IFS curve for fusion features. **(B)** ROC curves of models constructed based on optimal feature subsets and six single feature sets. **(C)** ROC curves of the integrated classifier model with 5-fold cross-validation. **(D)** ROC curves of the integrated classifier model with independent testing.

TABLE 1 **Performance of models constructed based on optimal feature subsets and six single feature sets.**

Feature type	*Sn*(%)	*Sp*(%)	*MCC*	*ACC*(%)	*AUC*
AAC	86.99	84.56	0.7324	85.74	0.9463
DPC	93.97	91.62	0.8597	92.83	0.9643
TPC	92.87	91.62	0.8512	92.30	0.9773
CKSAAP	92.79	89.26	0.8296	91.07	0.9636
PseAAC	85.66	92.79	0.7941	89.29	0.9518
DCP	86.99	89.26	0.7650	88.13	0.9658
Optimal subset	94.04	94.04	0.8825	94.06	0.9790

3.4. Leave-one-out validation of the model

Due to the small sample data size, model robustness may be questioned. To ensure credible results, we use the leave-one-out method to re-validate model performance. The results of the model performance evaluation based on the leave-one-out method are shown in Table 4. In the performance evaluation of the model using the leave-one-out method, the model achieves an *ACC* of 93.45% and

TABLE 2 Performance of the integrated classifier model and the four basic classifier models.

category	Classifiers	*Sn*(%)	*Sp*(%)	*MCC*	*ACC*(%)	*AUC*
5-fold	KNN	91.69	83.38	0.7603	87.49	0.9250
	LR	91.69	72.57	0.6629	82.09	0.9266
	RF	95.29	89.34	0.8482	92.30	0.9645
	SVM	94.04	96.47	0.9070	95.26	0.9790
	Stacking	95.22	96.47	0.9179	95.85	0.9863
Independent test	KNN	90.91	86.36	0.7735	88.64	0.9483
	LR	95.45	81.82	0.7800	88.64	0.9566
	RF	90.91	90.91	0.8182	90.91	0.9793
	SVM	100.00	81.82	0.8321	90.91	0.9752
	Stacking	100.00	86.36	0.8718	93.18	0.9793

TABLE 3 Performance of the stacking classifier model and the XGboost classifier models.

category	Classifiers	*Sn*(%)	*Sp*(%)	*MCC*	*ACC*(%)	*AUC*
5-fold	XGboost	92.94	88.01	0.8188	90.48	0.9675
	Stacking	95.22	96.47	0.9179	95.85	0.9863
Independent test	XGboost	100.00	81.82	0.8321	90.91	0.9917
	Stacking	100.00	86.36	0.8718	93.18	0.9793

TABLE 4 Performance evaluation based on the leave-one-out method.

category	Classifiers	*Sn*(%)	*Sp*(%)	*MCC*	*ACC*(%)	*AUC*
Leave-one-out	Stacking	92.86	94.05	0.8691	93.45	0.9846

an *AUC* of 0.9846. The model shows good performance on both cross-validation methods, signifying its stability and the reliability of its classification outcomes.

4. Conclusion

Hemagglutinin (HA) is a vital glycoprotein found on the surface of influenza viruses, and accurately identifying HA is crucial for the development of targeted vaccine drugs. In this study, we proposed a prediction model based on HA protein sequence features. The model was constructed using the Stacking algorithm, incorporating an optimal subset of features and a basic classifier model. Our results demonstrated that the constructed model exhibits excellent predictive capacity and generalization ability.

We anticipate that the model will prove valuable in the effective identification and prediction of HA. Moving forward, we plan to explore additional feature extraction methods and optimize our prediction model to further enhance its performance. Additionally, we are committed to developing an accessible web server to facilitate the identification and prediction of HA.

In summary, our research provides a promising approach to accurately identifying HA and lays the foundation for the development of targeted vaccine drugs. We believe that our findings contribute to the advancement of influenza research and offer valuable insights for future studies in this field.

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

XZ: Data curation, Formal analysis, Methodology, Visualization, Writing – original draft. LR: Formal analysis, Investigation, Validation, Writing – review & editing. PC: Investigation, Validation, Writing – review & editing. YZ: Investigation, Software, Validation, Writing – review & editing. HD: Supervision, Validation, Writing – review & editing. KD: Data curation, Investigation, Supervision, Writing – review & editing. XY: Conceptualization, Writing – review & editing. HL: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. CH: Conceptualization, Methodology, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Natural Science Foundation of China

(62250028 and 62202069), the Natural Science Foundation of Sichuan Province (2022NSFSC1610 and 2023NSFSC0678), and the China Postdoctoral Science Foundation (2023M730507).

Conflict of interest

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

The reviewer QT declared a shared affiliation with the author YZ to the handling editor at the time of review.

Publisher's note

References

1. Krammer F, Smith GJD, Fouchier RAM, Peiris M, Kedzierska K, Doherty PC, et al. Influenza. *Nat Rev Dis Primers*. (2018) 4:21. doi: 10.1038/s41572-018-0002-y

2. Uyeki TM, Hui DS, Zambon M, Wentworth DE, Monto AS. Influenza. *Lancet*. (2022) 400:693–706. doi: 10.1016/S0140-6736(22)00982-5

3. Skehel JJ, Wiley DC. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu Rev Biochem*. (2000) 69:531–69. doi: 10.1146/annurev.biochem.69.1.531

4. Nuwarda RF, Alharbi AA, Kayser V. An overview of influenza viruses and vaccines. *Vaccine*. (2021) 9:27. doi: 10.3390/vaccines9091032

5. Qiang X, Zhou C, Ye X, Du PF, Su R, Wei L. CPPred-FL: a sequence-based predictor for large-scale identification of cell-penetrating peptides by feature representation learning. *Brief Bioinform*. (2020) 21:11–23. doi: 10.1093/bib/bby091

6. Hasan MM, Schaduangrat N, Basith S, Lee G, Shoombuatong W, Manavalan B. HLPpred-fuse: improved and robust prediction of hemolytic peptide and its activity by fusing multiple feature representation. *Bioinformatics*. (2020) 36:3350–6. doi: 10.1093/bioinformatics/btaa160

7. Wei L, Zhou C, Chen H, Song J, Su R. ACPred-FL: a sequence-based predictor using effective feature representation to improve the prediction of anti-cancer peptides. *Bioinformatics*. (2018) 34:4007–16. doi: 10.1093/bioinformatics/bty451

8. Tang H, Zhao YW, Zou P, Zhang CM, Chen R, Huang P, et al. HBPred: a tool to identify growth hormone-binding proteins. *Int J Biol Sci*. (2018) 14:957–64. doi: 10.7150/ijbs.24174

9. Jiao S, Chen Z, Zhang L, Zhou X, Shi L. ATGPred-FL: sequence-based prediction of autophagy proteins with feature representation learning. *Amino Acids*. (2022) 54:799–809. doi: 10.1007/s00726-022-03145-5

10. Dao FY, Liu ML, Su W, Lv H, Zhang ZY, Lin H, et al. AcrPred: A hybrid optimization with enumerated machine learning algorithm to predict anti-CRISPR proteins. *Int J Biol Macromol*. (2023) 228:706–14. doi: 10.1016/j.ijbiomac.2022.12.250

11. Cacciabue M, Marcone DN. INFINITy: A fast machine learning-based application for human influenza A and B virus subtyping. *Influenza Other Respir Viruses*. (2023) 17:e13096. doi: 10.1111/irv.13096

12. Ao C, Jiao S, Wang Y, Yu L, Zou Q. Biological sequence classification: A review on data and general methods. *Research*. (2022) 2022:0011. doi: 10.34133/research.0011

13. Xu Y, Wojtczak D. Dive into machine learning algorithms for influenza virus host prediction with hemagglutinin sequences. *Biosystems*. (2022) 220:104740. doi: 10.1016/j.biosystems.2022.104740

14. Yin R, Thwin NN, Zhuang P, Lin Z, Kwoh CK. IAV-CNN: A 2D convolutional neural network model to predict antigenic variants of influenza a virus. *IEEE/ACM Trans Comput Biol Bioinform*. (2022) 19:3497–506. doi: 10.1109/tcbb.2021.3108971

15. Wang H, Zang Y, Zhao Y, Hao D, Kang Y, Zhang J, et al. Sequence matching between hemagglutinin and neuraminidase through sequence analysis using machine learning. *Viruses*. (2022) 14:469. doi: 10.3390/v14030469

16. Kargarfard F, Sami A, Hemmatzadeh F, Ebrahimie E. Identifying mutation positions in all segments of influenza genome enables better differentiation between pandemic and seasonal strains. *Gene*. (2019) 697:78–85. doi: 10.1016/j.gene.2019.01.014

17. Su W, Liu ML, Yang YH, Wang JS, Li SH, Lv H, et al. PPD: A manually curated database for experimentally verified prokaryotic promoters. *J Mol Biol*. (2021) 433:166860. doi: 10.1016/j.jmb.2021.166860

18. Wei Y, Zou Q, Tang F, Yu L. WMSA: a novel method for multiple sequence alignment of DNA sequences. *Bioinformatics*. (2022) 38:5019–25. doi: 10.1093/bioinformatics/btac658

19. Bateman A, Martin MJ, Orchard S, Magrane M, Ahmad S, Alpi E, et al. UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res*. (2022) 51:D523–31. doi: 10.1093/nar/gkac1052

20. Fu LM, Niu BF, Zhu ZW, Wu ST, Li WZ. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. (2012) 28:3150–2. doi: 10.1093/bioinformatics/bts565

21. Manavalan B, Patra MC. MLCPP 2.0: an updated cell-penetrating peptides and their uptake efficiency predictor. *J Mol Biol*. (2022) 434:167604. doi: 10.1016/j.jmb.2022.167604

22. Shoombuatong W, Basith S, Pitti T, Lee G, Manavalan B. THRONE: A new approach for accurate prediction of human RNA N7-Methylguanosine sites. *J Mol Biol*. (2022) 434:167549. doi: 10.1016/j.jmb.2022.167549

23. Thi Phan L, Woo Park H, Pitti T, Madhavan T, Jeon YJ, Manavalan B. MLACP 2.0: an updated machine learning tool for anticancer peptide prediction. *Comput Struct Biotechnol J*. (2022) 20:4473–80. doi: 10.1016/j.csbj.2022.07.043

24. Bupi N, Sangaraju VK, Phan LT, Lal A, Vo TTB, Ho PT, et al. An effective integrated machine learning framework for identifying severity of tomato yellow leaf curl virus and their experimental validation. *Research*. (2023) 6:0016. doi: 10.34133/research.0016

25. Ao C, Ye X, Sakurai T, Zou Q, Yu L. m5U-SVM: identification of RNA 5-methyluridine modification sites based on multi-view features of physicochemical features and distributed representation. *BMC Biol*. (2023) 21:93. doi: 10.1186/s12915-023-01596-0

26. Wang YH, Zhang YF, Zhang Y, Gu ZF, Zhang ZY, Lin H, et al. Identification of adaptor proteins using the ANOVA feature selection technique. *Methods*. (2022) 208:42–7. doi: 10.1016/j.ymeth.2022.10.008

27. Lv H, Dao F-Y, Lin H. DeepKla: an attention mechanism-based deep neural network for protein lysine lactylation site prediction. *iMeta*. (2022) 1:e11. doi: 10.1002/imt2.11

28. Yang K, Li M, Yu L, He X. Repositioning linifanib as a potent anti-necroptosis agent for sepsis. *bioRxiv*. (2023) 9:57. doi: 10.1101/2022.03.24.485557

29. Wang Y, Zhai Y, Ding Y, Zou Q. SBSM-pro: support bio-sequence machine for proteins. *arXiv Preprint*. (2023). doi: 10.48550/arXiv.2308.10275

30. Zhang D, Xu ZC, Su W, Yang YH, Lv H, Yang H, et al. iCarPS: a computational tool for identifying protein carbonylation sites by novel encoded features. *Bioinformatics*. (2020) 37:171–7. doi: 10.1093/bioinformatics/btaa702

31. Manavalan B, Basith S, Shin TH, Wei L, Lee G. Meta-4mCpred: A sequence-based Meta-predictor for accurate DNA 4mC site prediction using effective feature representation. *Mol Ther Nucleic Acids*. (2019) 16:733–44. doi: 10.1016/j.omtn.2019.04.019

32. Hasan MM, Manavalan B, Khatun MS, Kurata H. i4mC-ROSE, a bioinformatics tool for the identification of DNA N4-methylcytosine sites in the Rosaceae genome. *Int J Biol Macromol*. (2020) 157:752–8. doi: 10.1016/j.ijbiomac.2019.12.009

33. Chen Z, Zhao P, Li C, Li FY, Xiang DX, Chen YZ, et al. iLearnPlus: a comprehensive and automated machine-learning platform for nucleic acid and protein sequence analysis, prediction and visualization. *Nucleic Acids Res*. (2021) 49:e60. doi: 10.1093/nar/gkab122

34. Chen K, Kurgan L, Rahbari M. Prediction of protein crystallization using collocation of amino acid pairs. *Biochem Biophys Res Commun*. (2007) 355:764–9. doi: 10.1016/j.bbrc.2007.02.040

35. Liu B, Xu JH, Lan X, Xu RF, Zhou JY, Wang XL, et al. iDNA-Prot vertical bar dis: identifying DNA-binding proteins by incorporating amino acid distance-pairs and reduced alphabet profile into the general Pseudo amino acid composition. *PLoS One*. (2014) 9:12. doi: 10.1371/journal.pone.0106691

36. Peterson EL, Kondev J, Theriot JA, Phillips R. Reduced amino acid alphabets exhibit an improved sensitivity and selectivity in fold assignment. *Bioinformatics*. (2009) 25:1356–62. doi: 10.1093/bioinformatics/btp164

37. Dao FY, Lv H, Zhang ZY, Lin H. BDselect: A package for k-mer selection based on the binomial distribution. *Curr Bioinforma*. (2022) 17:238–44. doi: 10.2174/1574893616666211007102747

38. Yang H, Luo Y, Ren X, Wu M, He X, Peng B, et al. Risk prediction of diabetes: big data mining with fusion of multifarious physical examination indicators. *Inf. Fusion*. (2021) 75:140–9. doi: 10.1016/j.inffus.2021.02.015

39. Charoenkwan P, Chiangjong W, Nantasenamat C, Hasan MM, Manavalan B, Shoombuatong W. StackIL6: a stacking ensemble model for improving the prediction of IL-6 inducing peptides. *Brief Bioinform*. (2021) 22:bbab172. doi: 10.1093/bib/bbab172

40. Basith S, Lee G, Manavalan B. STALLION: a stacking-based ensemble learning framework for prokaryotic lysine acetylation site prediction. *Brief Bioinform*. (2022) 23:bbab376. doi: 10.1093/bib/bbab376

41. Hasan MM, Tsukiyama S, Cho JY, Kurata H, Alam MA, Liu X, et al. Deepm5C: A deep-learning-based hybrid framework for identifying human RNA N5-methylcytosine sites using a stacking strategy. *Mol Ther*. (2022) 30:2856–67. doi: 10.1016/j.ymthe.2022.05.001

42. Jeon YJ, Hasan MM, Park HW, Lee KW, Manavalan B. TACOS: a novel approach for accurate prediction of cell-specific long noncoding RNAs subcellular localization. *Brief Bioinform*. (2022) 23:bbac243. doi: 10.1093/bib/bbac243

43. Yuan SS, Gao D, Xie XQ, Ma CY, Su W, Zhang ZY, et al. IBPred: A sequence-based predictor for identifying ion binding protein in phage. *Comput Struct Biotechnol J*. (2022) 20:4942–51. doi: 10.1016/j.csbj.2022.08.053

44. Zhang ZY, Ning L, Ye X, Yang YH, Futamura Y, Sakurai T, et al. iLoc-miRNA: extracellular/intracellular miRNA prediction using deep BiLSTM with attention mechanism. *Brief Bioinform*. (2022) 23:bbac395. doi: 10.1093/bib/bbac395

45. Yang Y, Gao D, Xie X, Qin J, Li J, Lin H, et al. DeepIDC: A prediction framework of injectable drug combination based on heterogeneous information and deep learning. *Clin Pharmacokinet*. (2022) 61:1749–59. doi: 10.1007/s40262-022-01180-9

46. Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Trans Inf Theory*. (1967) 13:21–7. doi: 10.1109/TIT.1967.1053964

47. Freedman DA. *Statistical models: theory and practice*. New York: Cambridge University Press (2005).

48. Breiman L. Random forests. *Mach Learn*. (2001) 45:5–32. doi: 10.1023/a:1010933404324

49. Cortes C, Vapnik V. Support-Vector Networks. *Mach Learn*. (1995) 20:273–97. doi: 10.1007/bf00994018

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

51. Breiman L. Stacked regressions. *Mach Learn*. (1996) 24:49–64. doi: 10.1007/BF00117832

52. Sun Z, Huang Q, Yang Y, Li S, Lv H, Zhang Y, et al. PSnoD: identifying potential snoRNA-disease associations based on bounded nuclear norm regularization. *Brief Bioinform*. (2022) 23:bbac240. doi: 10.1093/bib/bbac240

53. Basith S, Hasan MM, Lee G, Wei L, Manavalan B. Integrative machine learning framework for the identification of cell-specific enhancers from the human genome. *Brief Bioinform*. (2021) 22:bbab252. doi: 10.1093/bib/bbab252

54. Manavalan B, Basith S, Shin TH, Lee G. Computational prediction of species-specific yeast DNA replication origin via iterative feature representation. *Brief Bioinform*. (2021) 22:bbaa304. doi: 10.1093/bib/bbaa304

55. Wei L, He W, Malik A, Su R, Cui L, Manavalan B. Computational prediction and interpretation of cell-specific replication origin sites from multiple eukaryotes by exploiting stacking framework. *Brief Bioinform*. (2021) 22:bbaa275. doi: 10.1093/bib/bbaa275

56. Yu L, Zheng YJ, Gao L. MiRNA-disease association prediction based on meta-paths. *Brief Bioinform*. (2022) 23:bbab571. doi: 10.1093/bib/bbab571

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 20 December 2023
DOI 10.3389/fmed.2023.1269221

OPEN ACCESS

EDITED BY
Abdul Azeez Sayed,
Imam Abdulrahman Bin Faisal University, Saudi Arabia

REVIEWED BY
Fernando Marqués-García,
Hospital Germans Trias i Pujol, Spain
Sheng Ding,
Chengdu University, China

*CORRESPONDENCE
Zhang Jing
✉ zhangjing@qlu.edu.cn
Lan Wenjun
✉ lanwenjun0522@qlu.edu.cn

†These authors share first authorship

RECEIVED 29 July 2023
ACCEPTED 28 November 2023
PUBLISHED 20 December 2023

CITATION
Ping Y, Quanlin S, Yue H, Jing Z and Wenjun L (2023) Screening and validation of double allele-specific binding F-primers for the measurement of antihypertensive pharmacogenomics. *Front. Med.* 10:1269221. doi: 10.3389/fmed.2023.1269221

Screening and validation of double allele-specific binding F-primers for the measurement of antihypertensive pharmacogenomics

Yang Ping†, Su Quanlin†, Hu Yue, Zhang Jing* and Lan Wenjun*

Institute of Biomedical Engineering, Qilu University of Technology (Shandong Academy of Sciences), Jinan, Shandong, China

Objective: Previous studies have proposed that genetic polymorphisms of CYP2D6*10, ADRB1, NPPA, CYP3A5*3, ACE, CYP2C9*3, and AGTR1 are involved in antihypertensive pharmacogenomics. The purpose of this study is to develop an amplification analysis using double allele-specific (AS) binding primers for accurate measurement of antihypertensive pharmacogenomics.

Methods: To establish a quadruplex quantitative PCR (qPCR) analysis for genotyping of CYP2D6*10, ADRB1 (1165 G>C), NPPA (2238 T>C) and CYP3A5*3, and a triplex qPCR analysis for genotyping of ACE (I/D), CYP2C9*3 and AGTR1 (1166 A>C), mismatch AS F-primers were screened by detection of plasmid/gDNA, and were validated by agreement analysis/reproducibility evaluation, in which the ΔCq (differences in threshold cycles between the wild-type F-primer-based amplification assay and the mutant-type F-primer-based amplification assay) was employed to determine genotypes.

Results: Seven pairs of primers were successfully selected through three rounds of F-primers screening. Except for ADRB1, the robustness assessment showed the amplification efficiency ranging from 0.9 to 1.1. In agreement analysis, two specimens in the training set (n = 203) were defined by the triplex analysis rather than NGS as heterozygotes for ACE, which was evidenced by gel electrophoresis. Reproducibility evaluation demonstrated that the coefficient of variation (CV) was <5%.

Conclusion: Multiplex amplification analysis using screened AS binding primers is a simple, reliable, and accurate tool to guide drug delivery in antihypertensive personalized treatment.

KEYWORDS

allele-specific (AS), F-primers, polymorphism, antihypertensive, genotyping

1 Introduction

Hypertension is becoming the main cause of cardiovascular disease (1–4). Patients with hypertension reached 1.278 billion worldwide in 2019 (5). A previous study reported that intensive hypertension control could avoid 2.209 million coronary heart disease (CHD) events, 4.409 million stroke events, and 75,100 cardiovascular deaths in 10 years (6). However, hypertension control is still poor because of insufficient clinical experience and apparent drug resistance (2).

Clinical practices have displayed the heterogeneity of patients' responses to antihypertensive drugs (7, 8). Associated with the efficacy of antihypertensive drugs covering beta-blockers, diuretics, calcium channel blockers (CCBs), angiotensin-converting enzyme (ACE) inhibitors, and angiotensin receptor antagonists (ARBs), the hypertensive pharmacogenomics of CYP2D6*10, adrenergic receptor beta 1 (ADRB1, 1165 G>C), natriuretic peptide type A (NPPA, 2238 T>C), CYP3A5*3, ACE (I/D), CYP2C9*3, and angiotensin II receptor 1 (AGTR1, 1166 A>C) have been documented. For example, genetic polymorphisms of AGTR1 and cytochrome P450 oxidase (CYP2C9) impact ARB drug-target affinity and drug metabolism, respectively. ARBs exert their antihypertensive effect by blocking the binding of the AGTR1 receptor with angiotensin II ligands. A carrier with a C-type allele for AGTR1 appears more sensitive to ARBs (9–11). Metabolizing most ARBs, CYP2C9 is a member of the cytochrome P450 oxidase superfamily (12). Patients with the CYP2C9*3 (*1/*1) allele extend their fast metabolic response to ARBs (13, 14).

Discrimination of single nucleotide polymorphisms (SNPs) has become a field of intense investigation, and various technologies have been suggested (15). It is not suitable for next-generation sequencing (NGS) technology to measure SNPs due to expensive equipment and time consumption (16). Presenting high sensitivity, high throughput, and high reliability, quantitative PCR (qPCR) analysis has become a mainstream technique for SNP measurement (17–19). Uniplex amplification analysis for CYP2D6*10, ADRB1 (1165 G>C), NPPA (2238 T>C), CYP3A5*3, ACE, CYP2C9*3, and AGTR1 (1166A>C) has been reported (7, 20–23). In this study, we described a multiplex qPCR analysis using double allele-specific (AS) binding F-primers that included wild-type and mutated-type F-primers to discriminate above genetic polymorphisms.

2 Materials and methods

2.1 Design strategy

To detect the genotypes of CYP2D6*10, ADRB1 (1165 G>C), NPPA (2238 T>C), CYP3A5*3, ACE (I/D), CYP2C9*3, and AGTR1 (1166 A>C), AS F-primers with the second, third, or fifth mismatched base at the 3′-terminal, co-hydrolysis probe, and reverse primers were designed.

First, F-primers were screened by plasmid determination using uniplex qPCR. Second, selected F-primers were screened by detecting genomic DNA using multiplex qPCR, including a quadruplex analysis for CYP2D6*10, ADRB1, NPPA, and CYP3A5*3 and a triplex analysis for ACE (I/D), CYP2C9*3, and AGTR1. Each of the multiplex analyses contained two reactions: the wild-type F-primer-based amplification assay and the mutant-type F-primer-based amplification assay, in which ΔCq (differences in threshold cycles between the wild-type F-primer-based assay and the mutant-type F-primer-based assay) was employed to determine genotypes. Third, the F-primers were screened by robustness assessment. Finally, the screened F-primer-based assay was validated by concordance analysis, using NGS as the reference method. The sterilized distilled water instead of DNA was used as the negative control during the whole experiment process. The design strategy of this study is shown in Figure 1.

2.2 Sample collection and DNA extraction

A total of 203 oral swab samples were collected from Chinese volunteers. Genomic DNA was isolated using the QIAamp DNA Mini kit (Cat No.51304, QIAGEN, Dusseldorf, Germany), according to the manufacturer's instructions. The quality and quantity of DNA were determined by the NanoPhotometer P360 (Implen GmbH, Munich, Germany). Sanger sequencing was conducted by Personal company (Qingdao Personal Biotechnology Co., Ltd., China). NGS sequencing was conducted by the Center for Molecular Diagnosis at Shandong Provincial Hospital, affiliated with Shandong First Medical University (Jinan, China).

2.3 Primers and probe design

Based on the nucleotide sequences of these seven genes, AS F-primers, co-reverse primers, and hydrolysis probes were designed using Primer Express 3.0. The probes for CYP2D6*10, ADRB1 (1165 G>C), NPPA (2238 T>C), and CYP3A5*3 in the quadruplex analysis were labeled with the fluorescent dyes FAM, VIC, NED, and CY5 at their 5′ ends and the quencher BHQ1, BHQ1, BHQ2, and BHQ3 at their 3′ ends, respectively. The probes for ACE (I/D), CYP2C9*3, and AGTR1 (1166 A>C) in

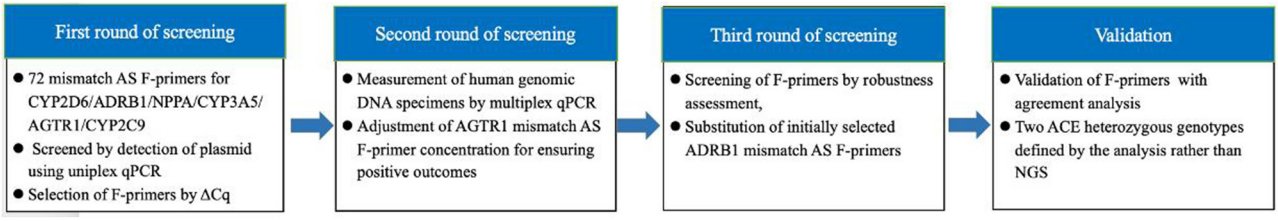

FIGURE 1
Design strategy of this study. F-primers as polymorphism-binding oligonucleotides were optimized through three rounds of screening and validated with agreement analysis considering next-generation sequencing as a reference method.

the triplex analysis were labeled with the fluorescent dyes FAM, NED, and CY5 at their 5′ ends, and the quencher BHQ1, BHQ2, and BHQ3 at their 3′ ends, respectively. Primers were synthesized by Sangon Biotech (Shanghai) Co. Ltd. (Shanghai, China), and TaqMan probes were synthesized by Invitrogen Corporation (Shanghai, China). According to NCBI, each approximately 400-bp fragment of CYP2D6*10 (rs1065852), ADRB1 (rs1801253), NPPA (rs5065), CYP3A5*3 (rs776746), ACE (rs4646994), CYP2C9*3 (rs1057910), and AGTR1 (rs5186) was chemically synthesized and cloned into pUC57 plasmid vectors by Sangon Biotech Co., Ltd. (Shanghai, China).

2.4 Multiplex AS qPCR

Optimized multiplex AS qPCR was executed in a total of 20 μL reaction mixture containing 10 μL AceQ® Universal U+ Probe Master Mix V2 (Vazyme, Nanjing, China), 0.2 μM of each wild/mutated-type F-primer (0.4 μM for AGTR1), 0.2 μM of each reverse primer, 0.1 μM of each hydrolysis probe, and 10 ng gDNA. The uniplex amplification analysis was conducted according to the same protocol. The qPCR protocols started with a contamination digestion step for 2 min at 37°C, and a pre-denaturation step for 5 min at 95°C, followed by 45 cycles of 95°C for 10 s and 60°C

TABLE 1 Primer and probe sequences.

Analysis	Gene loci	Sequences (5′→ 3′)	5′ label	3′ label
Quadruplex qPCR	CYP2D6*10	WF: GCTGGGCTGCACGCTAA*C		
		MF: GCTGGGCTGCACGCTAG*T		
		R: CCTCCCTCACCTGGTCGAA		
		P: ACCAGGCCCCCTGCCACTGC	FAM	BHQ1
	ADRB1 (1165 G>C)	WF1: GCAAGGCCTTCCAGG		
		MF1: AAGGCCTTCCAGC		
		R1: CGCGGCCGGTCTCC		
		P1: TGCTCTGCTGCGCGCGC	VIC	BHQ1
		WF2: CGCAAGGCCTTG*CAGG		
		MF2: CGCAAGGCCTTG*CAGC		
		R2: TGGGTCGCGTGGCG		
		P2: ACTGCTCTGCTGCGCGCGC	VIC	BHQ1
	NPPA (2238 T>C)	WF: AGATATGTCTGTGTTCTCTTTGCAGTG*CT		
		MF: GATATGTCTGTGTTCTCTTTGCAGTG*CC		
		R: GGCAACAAGATGACACAAATGC		
		P: CAGACTGCAAGAGGCTCCTGTCCCC	NED	BHQ2
	CYP3A5*3	WF: GTGGTCCAAACAGGGAAGAGATG*T		
		MF: GTGGTCCAAACAGGGAAGAGATG*C		
		R: CATTATGGAGAGTGGCATAGGAGAT		
		P: CATTCGTTAAGCTGGGTGGTACATACGTGG	CY5	BHQ3
Triplex qPCR	ACE (D)	WF: ACCTGCTGCCTATACAGTCACTTTTA		
	ACE (I)	MF: GCTGGGATTACAGGCGTGATAC		
	ACE (I/D)	R: GGGACGTGGCCATCACA		
	ACE (I/D)	P: CAAGGCATTCAAACCCCTACCAGATCTG	FAM	BHQ1
	CYP2C9*3	WF: GTGCACGAGGTCCAGAGC*TACA		
		MF: GTGCACGAGGTCCAGAGG*TACC		
		R: CGAAAACATGGAGTTGCAGTGT		
		P: TGACCTTCTCCCCACCAGCCTGC	NED	BHQ2
	AGTR1 (1166 A>C)	WF: CAGCACTTCACTACCAAATC*AGCA		
		MF: AGCACTTCACTACCAAATGAT*CC		
		R: TTCATCGAGTTTCTGACATTGTTCT		
		P: TTGCATTAGACAGATGACGGCTGCTCG	CY5	BHQ3

WF, wild-type F-primer; MF, mutated-type F-primer; R, reverse primer; P, probe.
*Mismatched base.

for 35 s. The fluorescence signal was collected at 60°C. These amplifications were performed on the ABI7500 Real-Time PCR Instrument (Thermo Fisher Scientific Inc., MA, USA).

2.5 Data analysis

Data analysis and graphing were carried out using GraphPad Prism software version 9 (GraphPad Software, Inc., San Diego, CA).

3 Results

3.1 First round of F-primers screening by measurement of plasmid using uniplex qPCR

AS F-primers with the second, third, or fifth 3′-terminal mismatched base, co-hydrolysis probes, and reverse primers were designed for each SNP. Uniplex qPCR analysis and plasmid

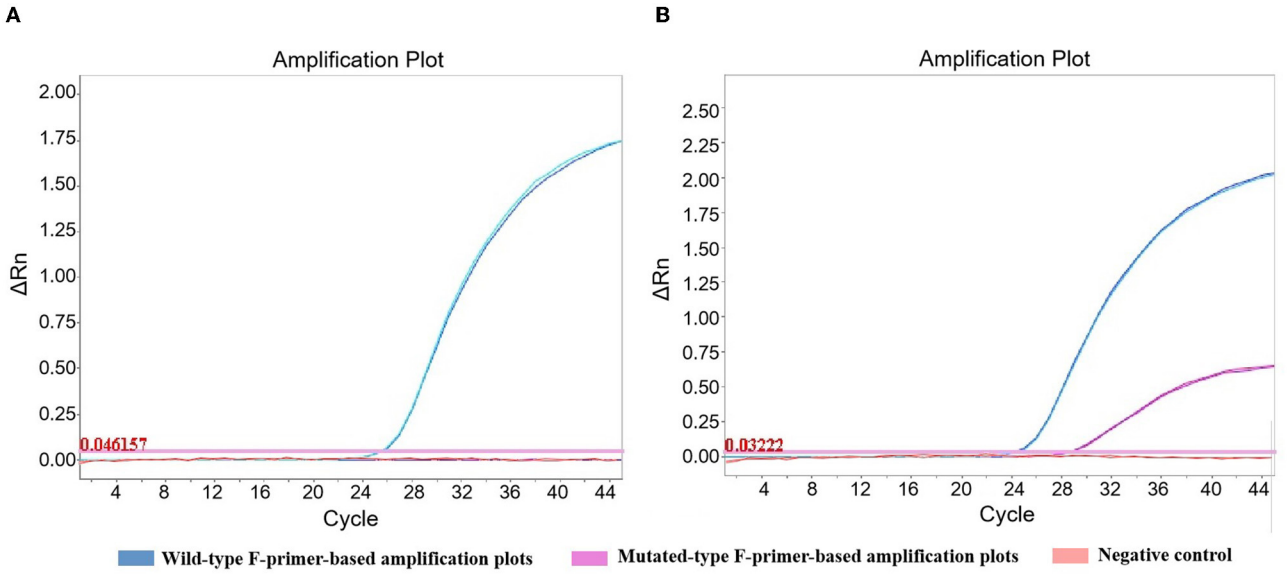

FIGURE 2
Amplification plots of AGTR1 in the triplex amplification analysis. The concentration of mutated-type F-primer for AGTR1 was adjusted from 0.2 **(A)** to 0.4 μM **(B)** to ensure positive outcomes. Wild-type homozygote was detected in duplicate by the triplex amplification analysis.

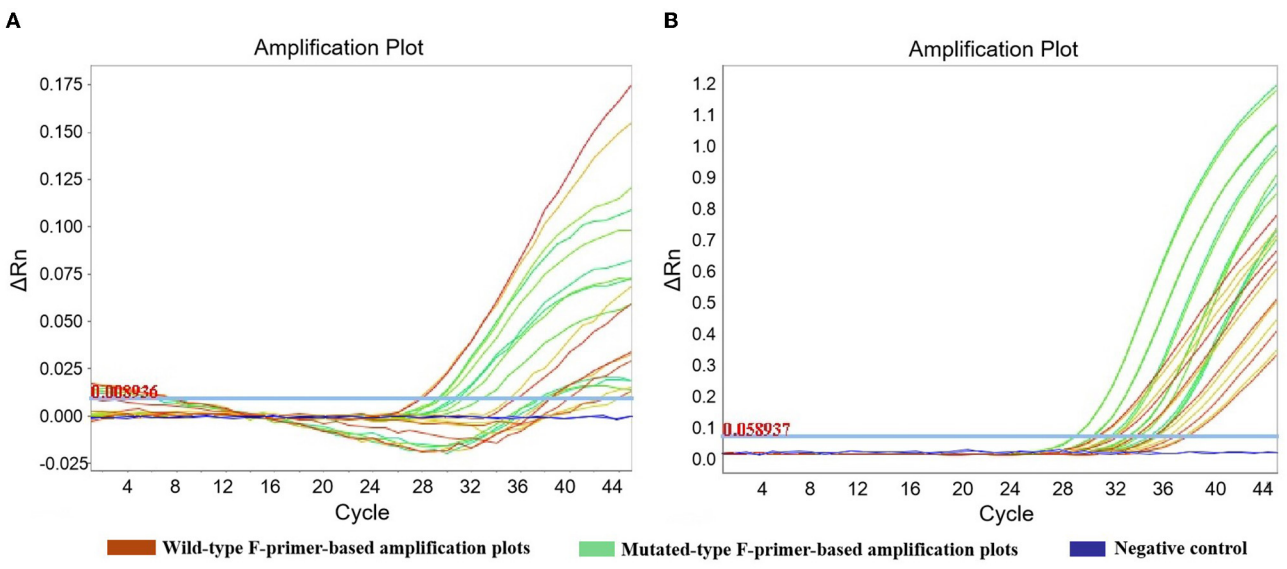

FIGURE 3
Amplification plots of ADRB1 in robustness assessment. To improve the robustness of ADRB1 measurement, F-primer WF2/MF2, probe P2, and reverse primer R2 were substituted for F-primer WF1/MF1, probe P1, and reverse primer R1, respectively, in the quadruplex amplification analysis. **(A)** Unoptimized amplification plots. **(B)** Optimized amplification plots. To assess robustness, serial dilutions of heterozygote (1.25–40 ng) were measured by the quadruplex amplification analysis, including the wild-type F-primer-based amplification assay and the mutated-type F-primer-based amplification assay. Reactions were run in duplicate.

models covering homozygotes and heterozygotes were employed to screen mismatch AS forward primers ($n = 72$) to roughly enable maximization of ΔCq (differences in threshold cycles between the wild-type F-primer-based assay and the mutated-type F-primer-based assay) (Supplementary material S1). The originally selected F-primers are shown in Table 1.

3.2 Second round of F-primers screening by detection of gDNA using multiplex qPCR

Selected F-primers were in succession screened by examination of human gDNA, comprising homozygotes and heterozygotes. To omit the positive control set in antihypertensive pharmacogenomic measurement, positive outcomes obtained from the wild- or mutated-type F-primer-based assay were required. Because undetermined results were observed in gDNA scans, the concentration of mutated-type F-primer for AGTR1 was adjusted from 0.2 to 0.4 μM to ensure positive outcomes (Figure 2).

3.3 Third round of F-primers screening by robustness assessment

Six concentrations (40 ng, 20 ng, 10 ng, 5 ng, 2.5 ng, and 1.25 ng) of heterozygotes from oral swabs were prepared. Amplification

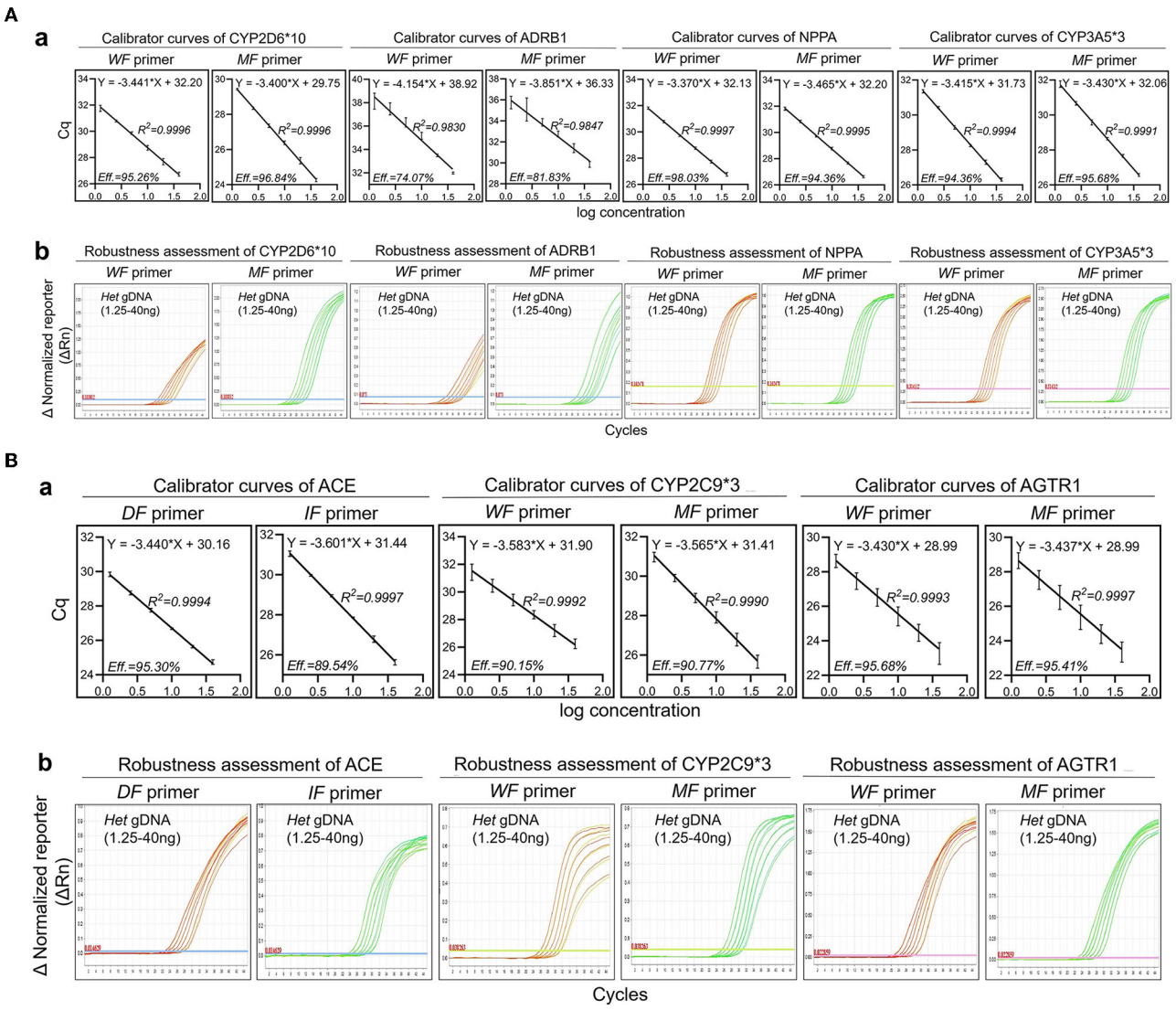

FIGURE 4
Robustness of optimized amplification analysis. The robustness assessment was executed by employing mismatch allele-specific F-primers targeting single nucleotide polymorphisms to simultaneously detect heterozygotes. Serial dilutions of heterozygotes (1.25–40 ng) were measured by multiplex amplification analysis containing a wild-type F-primer-based amplification assay and a mutated-type F-primer-based amplification assay. **(A)** (a) Calibrator curves of quadruplex amplification analysis. Amplification efficiency (Eff) % and R^2 are shown; (b) Amplification plots of the robustness assessment in quadruplex amplification analysis. Representative amplification plots are shown. **(B)** (a) Calibrator curves of triplex amplification analysis. Amplification efficiency (Eff) % and R^2 are shown; (b) Amplification plots of the robustness assessment in triplex amplification analysis. Representative amplification plots are shown. Reactions were run in duplicate with three independent experiments. Data are expressed as mean ± SE. *DF*, deletion-type F-primer; *IF*, insertion-type F-primer; *WF* primer, wild-type F-primer; *MF* primer, mutated-type F-primer; *Het*, heterozygote.

efficiency was calculated using the generated calibrator curve: $10^{-1/\text{slope}}-1$, with the logarithm of the template copies plotted on the X-axis and Cq plotted on the Y-axis (24). The reactions were conducted in duplicate with three dependent experiments. As the calibrator curve did not appear in a dose-dependent manner, F-primer WF2/MF2, probe P2, and reverse primer R2 for ADRB1 were substituted for F-primer WF1/MF1, probe P1, and reverse primer R1, respectively, in the quadruplex amplification analysis (Table 1 and Figure 3). Except for ADRB1, optimized calibrator curves demonstrated amplification efficiencies ranging from 0.9 to 1.1 and analytical sensitivities of at least 1.25 ng (Figure 4).

3.4 Verification of screened F-primers-based analysis by agreement analysis

Considering NGS as a reference method, we examined 203 gDNA samples extracted from oral swabs to evaluate the accuracy of the multiplex analyses using double allele-specific binding F-primers. The results showed that, besides ACE, the coincidence rate was 100%. Two specimens (No.001 and No.056) in the training set ($n = 203$) were defined by the analysis rather than NGS as heterozygotes for ACE, which was evidenced by gel electrophoresis (Figure 5). The cutoff values for genotyping are shown in Table 2.

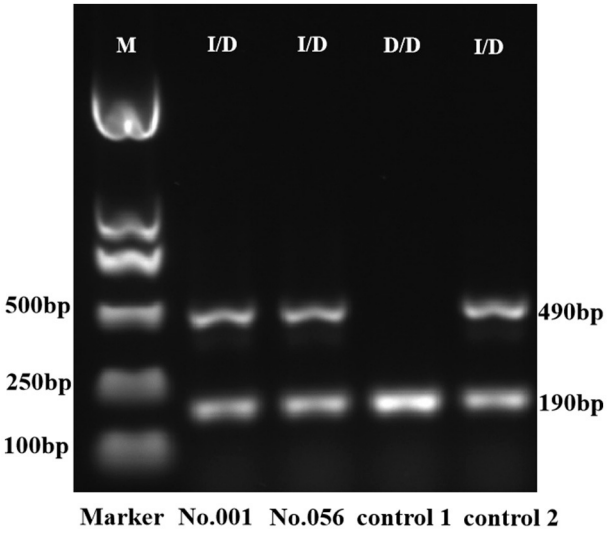

FIGURE 5
Gel electrophoresis of PCR amplicon for ACE. The D/D genotype was identified by the presence of a single 190-bp amplicon, and the I/D genotype extended both 490-bp and 190-bp amplicons. The ACE genotypes of controls 1 and 2 were D/D and I/D, respectively, defined by Sanger sequencing. M, DNA marker; I/D, heterozygote; D/D, deletion homozygote.

3.5 Substantiation of screened F-primers-based analysis by producibility evaluation

To evaluate the producibility of the analysis, each heterozygote was tested in eight-plicates by two operators, using two different reagent lots every 5 days ($n = 80$/specimen) at one site. A total of 80 Cq values were collected to calculate the coefficient of variation (CV). The results showed that the CV values for reproducibility were within 4.00% for all days, specimens, replicates, operators, and reagent lots combined (Figure 6).

4 Discussion

In this study, multiplex amplification analysis was established for the measurement of hypertensive pharmacogenomics. Due to genetic polymorphism, only about one-third of patients with hypertension accept effective treatment (25–30). Therefore, this study is helpful for hypertension patients to take more effective and well-tolerated medication.

Compared to other methods (31), multiplex qPCR behaves as a simple and effective approach to the detection of pharmacogenomic SNPs (32). Polymorphism-specific binding molecules in PCR-based analysis comprise dsDNA-binding dye, AS probe, and primer (33, 34). The dsDNA-binding dye-based high-resolution dissolution curve (HRM) assay needs a specific equipment module. In addition to the diseconomy, it is time-consuming and laborious to discover an appropriate minor groove binder (MGB) probe (35). The wild-type allele reaction probably outcompetes the mutated-type allele reaction when two AS probes barcoded with different fluorophores are utilized to identify genetic polymorphism (36). Sometimes, it is difficult to accurately discriminate SNPs using single-color melting curve analysis. For enhancement of AS primer specificity, base mismatch is more economical than locked nucleic acid (LNA) decoration (37). In this study, mismatch AS primers were screened and validated. The combination of a wild-type AS F-primer-based amplification assay with a mutated-type AS F-primer-based amplification assay was utilized to obtain ΔCq to define the genotype. The results evidenced that screened F-primer-based amplification analysis is a simple, accurate, and reliable approach to measure antihypertensive pharmacogenomics.

As the definition of ACE (I/D) genotype for two specimens differed between triplex analysis and NGS, we utilized PCR-gel electrophoresis to substantiate the outcomes. Located on chromosome 17, the ACE gene consists of 26 exons and appears as a polymorphism in the form of either insertion (I) or deletion (D) of a 287-bp Alu repeat sequence in intron 16. The ACE (I/D) allele can be detected by PCR using the primers flanking the 287 bp insertion sequence (38). In gel electrophoresis, the I/I genotype can be identified by the presence of a single 490 bp amplicon, the D/D genotype can be recognized by the presence of a single 190 bp product, and the I/D genotype extends both 490 and 190 bp amplicons (39). The results of gel electrophoresis validated the accuracy of the triplex analysis for ACE (I/D) measurement (Figure 5).

Based on hypertensive pharmacogenomics of CYP2C9*3, ADRB1(1165 G>C), AGTR1 (1166 A>C), CYP2D6*10, ACE (I/D), CYP3A5*3, and NPPA (2238 T>C), the principle of personalized drug delivery was proposed as follows: (a) doubling the standard dose is suggested when the hypertension is

TABLE 2 Genotyping cutoff values.

Analysis	Genetic polymorphism	Wild-type F-primer-based assay	Mutated-type F-primer-based assay	Heterozygote type
Quadruplex qPCR	CYP2D6*10	mCq > wCq and mCq - wCq > 3	wCq > mCq and wCq - mCq > 3	| wCq - mCq | ≤ 3
	NPPA (2238 T>C)			
	CYP3A5*3			
	ADRB1 (1165 G>C)	mCq > wCq and mCq - wCq > 4	wCq > mCq and wCq - mCq > 4	| wCq - mCq | ≤ 4
Triplex qPCR	ACE (I/D)[a]	mCq > wCq and mCq - wCq > 2	wCq > mCq and wCq - mCq > 2	| wCq - mCq | ≤ 2
	AGTR1 (1166 A>C)			
	CYP2C9*3	mCq > wCq and mCq - wCq > 1.5	wCq > mCq and wCq - mCq > 1.5	| wCq - mCq | ≤ 1.5

mCq, Cq value of the mutated-type F-primer-based assay; wCq, Cq value of the wild-type F-primer-based assay.
[a]Deletion type for ACE was expressed as wild type, while insertion type for ACE was expressed as mutated type.

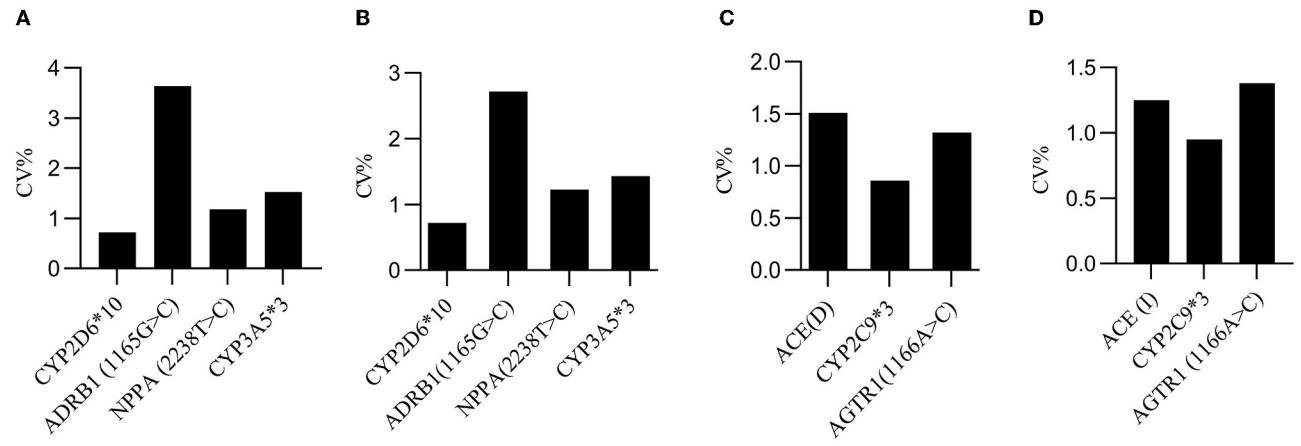

FIGURE 6
Producibility evaluation. **(A)** The inter-day CV values for the wild-type F-primers-based amplification assay in quadruplex analysis. **(B)** The inter-day CV values for the mutated-type F-primers-based amplification assay in quadruplex analysis. **(C)** The inter-day CV values for the wild-type F-primers-based amplification assay in triplex analysis. **(D)** The inter-day CV values for the mutated-type F-primers-based amplification assay in triplex analysis. The inter-day CV value was <4% for all days, specimens, replicates, operators, and reagent lots combined.

moderately sensitive to certain anti-hypertensive drugs; and (b) the minimum dose is recommended to initiate treatment when the hypertension is highly sensitive to certain anti-hypertensive drugs (7). Following the above principle, clinical studies evidenced that, compared to clinic experience-guided anti-hypertensive therapy, genotype-guided treatment appeared more effective and had fewer side effects. Herein, we established a simple, efficient, and accurate method for simultaneously detecting the genotypes of CYP2D6*10, ADRB1 (1165 G>C), NPPA (2238 T>C), CYP3A5*3, ACE, CYP2C9*3, and AGTR1 (1166 A>C) by screening and verification of mismatched AS F-primers.

5 Conclusion

As an accurate and reliable approach, the analysis described in this study is a valuable tool to determine the genotypes for CYP2D6*10, ADRB1, NPPA, CYP3A5*3, ACE, CYP2C9*3, and AGTR1, which can guide drug delivery in antihypertensive treatment to ensure curative effect. Employing the similar technique verified in this study, our laboratory will design and develop a multiplex amplification analysis for guiding aspirin delivery in the future.

Data availability statement

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

Ethics statement

This study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and were approved by the Biomedical Research Ethic Committee of Shandong Provincial Hospital.

Author contributions

YP: Data curation, Validation, Writing—original draft, Formal analysis. SQ: Writing—original draft, Data curation. HY: Software, Writing—original draft, Methodology. ZJ: Supervision, Writing—review & editing, Project administration, Methodology. LW: Funding acquisition, Project administration, Supervision, Writing—review & editing, Methodology.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by grants from the technological development of Yinfeng Biological Engineering Group (YFSW01-2021-B07-235) and the Key Project of International Cooperation of Qilu University of Technology (No. QLUTGJHZ2018008).

Acknowledgments

The authors wish to acknowledge Dr. Liu Yiqing, Department of Clinical Laboratory, Shandong Provincial Hospital Affiliated to Shandong First Medical University, for his help in NGS sequencing.

Conflict of interest

LW, ZJ, and HY are inventors of the detection kit for antihypertensive- associated genotyping (China Patent Application Number: 202210835232.3).

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2023.1269221/full#supplementary-material

References

1. Roth GA, Johnson C, Abajobir A, Abd-Allah F, Abera SF, Abyu G, et al. Global, regional, and national burden of cardiovascular diseases for 10 causes, 1990 to 2015. *J Am Coll Cardiol.* (2017) 70:1–25. doi: 10.1016/j.jacc.2017.04.052

2. DiPette DJ, Skeete J, Ridley E, Campbell NRC, Lopez-Jaramillo P, Kishore SP, et al. Fixed-dose combination pharmacologic therapy to improve hypertension control worldwide: Clinical perspective and policy implications. *J Clin Hypertens.* (2019) 21:4–15. doi: 10.1111/jch.13426

3. Comeau KD, Shokoples BG, Schiffrin EL. Sex differences in the immune system in relation to hypertension and vascular disease. *Can J Cardiol.* (2022) 38:1828–43. doi: 10.1016/j.cjca.2022.05.010

4. Mills KT, Bundy JD, Kelly TN, Reed JE, Kearney PM, Reynolds K, et al. Global disparities of hypertension prevalence and control: a systematic analysis of population-based studies from 90 countries. *Circulation.* (2016) 134:441–50. doi: 10.1161/CIRCULATIONAHA.115.018912

5. NCD Risk Factor Collaboration. Worldwide trends in hypertension prevalence and progress in treatment and control from 1990 to 2019: a pooled analysis of 1201 population-representative studies with 104 million participants. *Lancet.* (2021) 398:957–80. doi: 10.1016/S0140-6736(21)01330-1

6. The writing committee of the report on cardiovascular health and diseases in china. Report on cardiovascular health and diseases in China 2021: an updated summary. *Biomed Environ Sci.* (2022) 35:573–603. doi: 10.3967/bes2022.079

7. Xiao Z-L, Yang M, Chen X-B, Xie X-M, Chen M-F. Personalized antihypertensive treatment guided by pharmacogenomics in China. *Cardiovasc Diagn Ther.* (2022) 12:635–45. doi: 10.21037/cdt-22-154

8. Bosch TM. Pharmacogenomics of drug-metabolizing enzymes and drug transporters in chemotherapy. *Methods Mol Biol.* (2008) 448:63–76. doi: 10.1007/978-1-59745-205-2_5

9. Brown MJ. Heterogeneity of blood pressure response to therapy. *Am J Hyperten.* (2010) 23:926–8. doi: 10.1038/ajh.2010.139

10. Liu Y, Kong X, Jiang Y, Zhao M, Gao P, Cong X, et al. Association of AGTR1 A1166C and CYP2C9*3 gene polymorphisms with the antihypertensive effect of valsartan. *Int J Hypertens.* (2022) 2022:7677252. doi: 10.1155/2022/7677252

11. Zotova TY, Kubanova AP, Azova MM, Aissa AA, Gigani OO, Frolov VA. Analysis of polymorphism of angiotensin system genes (ACE, AGTR1, and AGT) and gene ITGB3 in patients with arterial hypertension in combination with metabolic syndrome. *Bull Exp Biol Med.* (2016) 161:334–8. doi: 10.1007/s10517-016-3408-0

12. Fekete F, Mangó K, Déri M, Incze E, Minus A, Monostory K. Impact of genetic and non-genetic factors on hepatic CYP2C9 expression and activity in Hungarian subjects. *Sci Rep.* (2021) 11:17081. doi: 10.1038/s41598-021-96590-3

13. Kamide K, Kawano Y, Rakugi H. Pharmacogenomic approaches to study the effects of antihypertensive drugs. *Hypertens Res.* (2012) 35:796–9. doi: 10.1038/hr.2012.82

14. Kirchheiner J, Seeringer A. Clinical implications of pharmacogenetics of cytochrome P450 drug metabolizing enzymes. *Biochim Biophys Acta.* (2007) 1770:489–94. doi: 10.1016/j.bbagen.2006.09.019

15. Diakosavvas M, Fasoulakis ZN, Kouroupi M, Theodora M, Inagamova L, Tsatsaris G, et al. Primary malignant melanoma of the cervix: a case report and a review of the literature. *Case Rep Oncol Med.* (2020) 2020:7206786. doi: 10.1155/2020/7206786

16. De Cario R, Kura A, Suraci S, Magi A, Volta A, Marcucci R, et al. Sanger validation of high-throughput sequencing in genetic diagnosis: still the best practice? *Front Genet.* (2020) 11:592588. doi: 10.3389/fgene.2020.592588

17. Saijo M, Morikawa S, Kurane I. Real-time quantitative polymerase chain reaction for virus infection diagnostics. *Expert Opin Med Diagn.* (2008) 2:1155–71. doi: 10.1517/17530059.2.10.1155

18. Zhao K, Shi K, Zhou Q, Xiong C, Mo S, Zhou H, et al. The development of a multiplex real-time quantitative PCR assay for the differential detection of the wild-type strain and the MGF505-2R, EP402R and I177L gene-deleted strain of the african swine fever virus. *Animals.* (2022) 12:1754. doi: 10.3390/ani12141754

19. Baris I, Etlik O, Koksal V, Ocak Z, Baris ST. SYBR green dye-based probe-free SNP genotyping: introduction of T-Plex real-time PCR assay. *Anal Biochem.* (2013) 441:225–31. doi: 10.1016/j.ab.2013.07.007

20. Candong W, Yuxiong K, Jianzhong J, Li J. Genotyping of essential hypertension candidate gene-ACE by TaqMan. *J Fudan Univ Nat Sci.* (2002) 6:614–7. doi: 10.15943/j.cnki.fdxb-jns.2002.06.004

21. Xiuqing L, Yanwu L, Lei H. Establishment of real - time PCR assay for detection of hypertension drug - related gene polymorphism. *J Mol Diagnost Therapy.* (2021) 13:707–11. (In Chinese). doi: 10.19930/j.cnki.jmdt.2021.05.008

22. Henderson SO, Haiman CA, Mack W. Multiple Polymorphisms in the renin-angiotensin-aldosterone system (ACE, CYP11B2, AGTR1) and their contribution to hypertension in African Americans and Latinos in the multiethnic cohort. *Am J Med Sci.* (2004) 328:266–73. doi: 10.1097/00000441-200411000-00006

23. Galaviz-Hernández C, Lazalde-Ramos BP, Lares-Assef I, Macías-Salas A, Ortega-Chavez MA, Rangel-Villalobos H, et al. Influence of genetic admixture components on CYP3A5*3 allele-associated hypertension in amerindian populations from Northwest Mexico. *Front Pharmacol.* (2020) 11:638. doi: 10.3389/fphar.2020.00638

24. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem.* (2009) 55:611–22. doi: 10.1373/clinchem.2008.112797

25. Oliveira-Paula GH, Pereira SC, Tanus-Santos JE, Lacchini R. Pharmacogenomics and hypertension: current insights. *Pharmgenomics Pers Med.* (2019) 12:341–59. doi: 10.2147/PGPM.S230201

26. Shin J, Johnson JA. Pharmacogenetics of beta-blockers. *Pharmacotherapy.* (2007) 27:874–87. doi: 10.1592/phco.27.6.874

27. Johnson JA. Advancing management of hypertension through pharmacogenomics. *Ann Med.* (2012) 44 Suppl 1:S17–22. doi: 10.3109/07853890.2011.653399

28. Arwood MJ, Cavallari LH, Duarte JD. Pharmacogenomics of hypertension and heart disease. *Curr Hypertens Rep.* (2015) 17:586. doi: 10.1007/s11906-015-0586-5

29. Rysz J, Franczyk B, Rysz-Górzyńska M, Gluba-Brzózka A. Pharmacogenomics of hypertension treatment. *Int J Mol Sci.* (2020) 21:4709. doi: 10.3390/ijms21134709

30. Cooper-DeHoff RM, Johnson JA. Hypertension pharmacogenomics: in search of personalized treatment approaches. *Nat Rev Nephrol.* (2016) 12:110–22. doi: 10.1038/nrneph.2015.176

31. Nak-On S, Tejangkura T, Chontananarth T. Multi-detection for paramphistomes using novel manually designed PAR-LAMP primers and its application in a lateral flow dipstick (LAMP-LFD) system. *Vet Parasitol.* (2023) 317:109905. doi: 10.1016/j.vetpar.2023.109905

32. Lee J, Kim S, Park J, Kang D-H. Detection of Salmonella enterica serovar Montevideo in food products using specific PCR primers developed by comparative genomics. *LWT.* (2022) 165:113677. doi: 10.1016/j.lwt.2022.113677

33. Nikodem D, Clapa T, Narozna D. [HRM-PCR in medical diagnostic]. *Postepy Biochem.* (2021) 67:59–63. doi: 10.18388/pb.2021_373

34. Matsuda K. PCR-based detection methods for single-nucleotide polymorphism or mutation: real-time PCR and its substantial contribution toward technological refinement. *Adv Clin Chem.* (2017) 80:45–72. doi: 10.1016/bs.acc.2016.11.002

35. Cheng H, Liu Y, Fu M, Liu J, Chen G. Taqman-MGB FQ-PCR for Human GPIIIa, GP1BA and PEAR1 SNPs. *Clin Lab.* (2022) 68. doi: 10.7754/Clin.Lab.2022.220132

36. Lang AH, Drexel H, Geller-Rhomberg S, Stark N, Winder T, Geiger K, et al. Optimized allele-specific real-time PCR assays for the detection of common mutations in KRAS and BRAF. *J Mol Diagn.* (2011) 13:23–8. doi: 10.1016/j.jmoldx.2010.11.007

37. Navarro E, Serrano-Heras, G, Castaño MJ, Solera J. Real-time PCR detection chemistry. *Clin Chim Acta.* (2015) 439:231–50. doi: 10.1016/j.cca.2014.10.017

38. Gupta S, Agrawal BK, Goel RK, Sehajpal PK. Angiotensin-converting enzyme gene polymorphism in hypertensive rural population of Haryana, India. *Emerg Trauma Shock.* (2009) 2:150–4. doi: 10.4103/0974-2700.55323

39. Johnston JD,Winder AF, Breimer LH. An MboI polymorphism at codon 192 of the human tyrosinase gene is present in Asians and Afrocaribbeans. *Nucleic Acids Res.* (1992) 20:1433. doi: 10.1093/nar/20.6.1433

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 09 February 2024
DOI 10.3389/fmed.2023.1340703

OPEN ACCESS

EDITED BY
Abdul Azeez Sayed,
Imam Abdulrahman Bin Faisal University, Saudi Arabia

REVIEWED BY
Padmaja Mummaneni,
United States Food and Drug Administration, United States
Rami M. Elshazli,
Horus University, Egypt

*CORRESPONDENCE
Mohammed Y. Behairy
✉ mohamedyehya950@gmail.com
Dalal Sulaiman Alshaya
✉ dsalshaya@pnu.edu.sa

RECEIVED 18 November 2023
ACCEPTED 28 December 2023
PUBLISHED 09 February 2024

CITATION
Behairy MY, Tawfik NZ, Eid RA, Nasser Binjawhar D, Alshaya DS, Fayad E, Elkhatib WF and Abdallah HY (2024) Mannose-binding lectin gene polymorphism in psoriasis and vitiligo: an observational study and computational analysis.
Front. Med. 10:1340703.
doi: 10.3389/fmed.2023.1340703

Mannose-binding lectin gene polymorphism in psoriasis and vitiligo: an observational study and computational analysis

Mohammed Y. Behairy[1]*, Noha Z. Tawfik[2], Refaat A. Eid[3], Dalal Nasser Binjawhar[4], Dalal Sulaiman Alshaya[5]*, Eman Fayad[6], Walid F. Elkhatib[7,8] and Hoda Y. Abdallah[9,10]

[1]Department of Microbiology and Immunology, Faculty of Pharmacy, University of Sadat City, Sadat City, Egypt, [2]Department of Dermatology, Venereology and Andrology, Faculty of Medicine, Suez Canal University, Ismailia, Egypt, [3]Pathology Department, College of Medicine, King Khalid University, Abha, Saudi Arabia, [4]Department of Chemistry, College of Science, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia, [5]Department of Biology, College of Science, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia, [6]Department of Biotechnology, College of Sciences, Taif University, Taif, Saudi Arabia, [7]Microbiology and Immunology Department, Faculty of Pharmacy, Ain Shams University, African Union Organization St., Abbassia, Cairo, Egypt, [8]Department of Microbiology and Immunology, Faculty of Pharmacy, Galala University, Suez, Egypt, [9]Department of Histology and Cell Biology (Genetics Unit), Faculty of Medicine, Suez Canal University, Ismailia, Egypt, [10]Center of Excellence in Molecular and Cellular Medicine, Faculty of Medicine, Suez Canal University, Ismailia, Egypt

Introduction: Psoriasis and vitiligo are inflammatory autoimmune skin disorders with remarkable genetic involvement. Mannose-binding lectin (MBL) represents a significant immune molecule with one of its gene variants strongly linked to autoimmune diseases. Therefore, in this study, we investigated the role of the MBL variant, rs1800450, in psoriasis and vitiligo disease susceptibility.

Methods: The study comprised performing *in silico* analysis, performing an observational study regarding psoriasis patients, and performing an observational study regarding vitiligo patients. Various *in silico* tools were used to investigate the impact of the selected mutation on the function, stability, post-translational modifications (PTMs), and secondary structures of the protein. In addition, a total of 489 subjects were enrolled in this study, including their demographic and clinicopathological data. Genotyping analysis was performed using real-time PCR for the single nucleotide polymorphism (SNP) rs1800450 on codon 54 of the MBL gene, utilizing TaqMan genotyping technology. In addition, implications of the studied variant on disease susceptibility and various clinicopathological data were analyzed.

Results: Computational analysis demonstrated the anticipated effects of the mutation on MBL protein. Furthermore, regarding the observational studies, rs1800450 SNP on codon 54 displayed comparable results in our population relative to global frequencies reported via the 1,000 Genomes Project. This SNP showed no significant association with either psoriasis or vitiligo disease risk in all genetic association models. Furthermore, rs1800450 SNP did not significantly correlate with any of the demographic or clinicopathological features of both psoriasis and vitiligo.

Discussion: Our findings highlighted that the rs1800450 SNP on the *MBL2* gene has no role in the disease susceptibility to autoimmune skin diseases, such as psoriasis and vitiligo, among Egyptian patients. In addition, our analysis advocated the notion of the redundancy of MBL and revealed the lack of significant impact on both psoriasis and vitiligo disorders.

KEYWORDS

psoriasis, vitiligo, MBL, innate immunity, SNP, rs1800450

1 Introduction

Psoriasis and vitiligo are inflammatory autoimmune diseases with genetic factors playing a remarkable role in both diseases (1). Psoriasis is an inflammatory skin disease that is characterized by the development of scaly and erythematous plaques with a global prevalence of approximately 2% (2, 3). Its pathogenesis comprises intricate interactions between the adaptive immune system and the innate immune system (4). Meanwhile, vitiligo represents an autoimmune disease of the skin characterized by autoimmune melanocyte destruction, resulting in the related depigmentation patches affecting the skin in addition to mucosa (5). The prevalence of vitiligo is estimated at approximately 0.2–2% in varied populations (6). Genetic factors have a remarkable role in vitiligo and the risk related to these factors is estimated to reach 75–83%, leaving only approximately 20% for the other environmental factors (7).

One of the important components of innate immunity is mannose-binding lectin (MBL), and the defects of this molecule have been linked with different autoimmune diseases (8). This protein is one of the pattern-recognition molecules that are responsible for activating the complement system through the lectin pathway (9, 10). It is encoded by the *MBL2* gene located on chromosome 10, and the presence of certain missense single nucleotide polymorphisms (SNPs) in this gene was associated with notable low levels of MBL (11). One of these specific mutations is rs1800450 in codon 54, which leads to the change of glycine to aspartic acid (12), with the variant allele represented by allele B and the wild-type one represented by allele A (13). Moreover, this missense mutation could impact the binding activity of MBL, leading to implications on its function (14). The presence of this SNP was linked with several autoimmune diseases such as rheumatoid arthritis (15) and systemic lupus erythematosus (16). However, the relationship between this SNP and autoimmune skin diseases, specifically vitiligo and psoriasis, was investigated by only a few studies showing conflicting results (17–20).

More studies are required to investigate the impact of this missense mutation on psoriasis and vitiligo diseases, especially in light of the increasing interest in the genetic association studies with different human diseases (21–24). By deciphering the nature of the relationship between this mutation and such chronic autoimmune diseases and their features, hopes could be raised to improve the guidelines for the prediction and management of these autoimmune conditions.

2 Materials and methods

The study comprised an *in silico* analysis, an observational study regarding psoriasis patients, and an observational study regarding vitiligo patients. A chart demonstrating the outline of the study research is displayed in Figure 1.

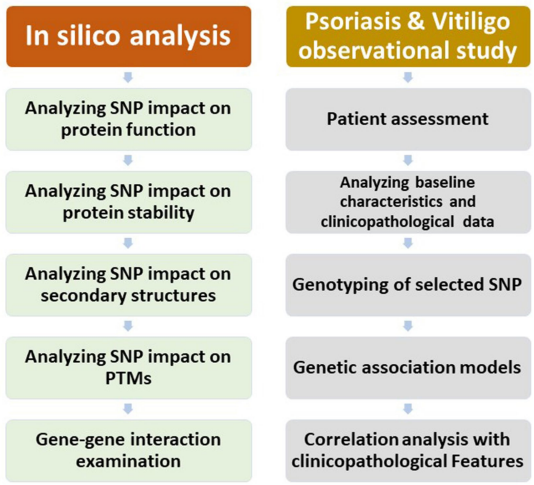

FIGURE 1
A chart demonstrating the outline of the study work.

2.1 Analyzing the variant's effect on the MBL protein function

For estimating how the SNP might affect protein function, MutPred2 was used.[1] MutPred2 represents a software package and machine learning technique that combines genetic and molecular data for reasoning probabilistically regarding the pathogenicity of substitutions of amino acids by providing (1) a general prediction regarding pathogenicity and (2) a list describing certain molecular changes that may affect the phenotype (25).

2.2 Analyzing the variant's effect on the MBL protein stability

The stability of the protein was examined with regard to the chosen mutation using the Mu-Pro tool. The Mu-Pro tool utilizes a strong support vector machine technique that, when used for cross-validation and verification, demonstrates an accuracy of 84%[2] (26).

2.3 Secondary structure analysis

The secondary structure of the studied protein was analyzed using the SOPMA server, and the precise alignment regarding the modified amino acids in the secondary structure was figured out.[3] In addition, the secondary structures with the selected SNP were examined as well. The SOPMA server could forecast the secondary structures of a given protein by examining the numerous alignments regarding the protein's sequence (27).

2.4 Conducting post-translational modification analysis

The MusiteDeep server[4] was employed to forecast the positions of various types of post-translational modifications. Due to the

1 http://mutpred2.mutdb.org/index.html

2 http://mupro.proteomics.ics.uci.edu/

3 https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html

4 https://www.musite.net

significant role of PTMs in controlling how proteins function, the recognition of PTMs is crucial for understanding the pathogenesis of diseases (28–30).

2.5 Gene–gene interaction examination

The GeneMANIA tool was employed to produce the *MBL2* gene–gene interaction network.[5] Using a range of resources and data types, GeneMANIA could predict the genes that strongly interact with a given gene (31).

2.6 Protein–protein interaction analysis

The STRING database[6] was adopted to predict the MBL protein–protein interaction network. The STRING database represents a database into which protein–protein interactions, encompassing both functional and physical relationships, are methodically gathered and integrated (32). The prediction was restricted to the 10 proteins representing the most interactive ones.

2.7 Study design of the experimental work

This is a case–control study in which the study participants were recruited from the outpatient dermatology department clinic, Suez Canal University Hospital (SCUH), Ismailia, Egypt, with the study participants classified into three main groups: Group 1 (vitiligo patient group), Group 2 (psoriasis patient group), and Group 3 (control group). All subjects or their next of kin provided informed consent.

2.7.1 Group 1 (vitiligo patients)

After being diagnosed with vitiligo through clinical examination, supplemented with Wood's lamp, 90 patients from both sexes were included. All the necessary clinicopathological data, such as sex, age, BMI, family history, previous history regarding other autoimmune diseases (e.g., Hashimoto's thyroiditis, diabetes mellitus, psoriasis, and Addison's disease), age of the disease onset, severity, disease duration, and treatment, were collected depending on patients' history. Furthermore, a thorough dermatological examination was applied to all patients to assess the size, site, pattern, and distribution of individual lesions. In addition, disease severity assessment was performed depending on the criteria of the Vitiligo Area Severity Index (VASI) in addition to the Vitiligo European Task Force (VETF).

2.7.2 Group 2 (psoriasis patients)

After being diagnosed with chronic plaque psoriasis, 99 patients of both sexes were included in the study, with the exclusion of those suffering from other autoimmune diseases. All needed clinicopathological data, such as sex, age, BMI, family history, severity, age of the disease onset, lines of treatment, and disease duration, were collected depending on patients' history. Moreover, a thorough dermatological examination was carried out on all patients to assess the size, site, pattern, and distribution of individual lesions. In addition, disease severity assessment was performed depending on the Psoriasis Area Severity Index (PASI) score. According to the European consensus, PASI is interpreted as mild in cases where the PASI score is less than 10, moderate in cases where the score is between 10 and 20, and severe in cases where the score is greater than 20.

2.7.3 Group 3 (control)

A total of 300 healthy volunteers were recruited from the SCUH blood bank. The members of this third group were matched with the members of the other two groups in terms of sex and age.

2.8 Genotyping

The molecular analysis was conducted at the Center of Excellence of Molecular and Cellular Medicine at the Faculty of Medicine, Suez Canal University and Hospital. The extraction of DNA was conducted from venous blood depending on the QIAamp DNA Blood Mini Kit (Cat. No. 51104, QIAGEN; Hilden, Germany). The concentration and purity of DNA were checked depending on the NanoDrop 2000 1^C spectrophotometer (NanoDrop Tech., Inc. Wilmington, DE, USA). The DNA samples were stored at −20°C until further analysis. *MBL2* SNP rs1800450 was identified by relying on real-time PCR (RT-PCR) utilizing TaqMan genotyping assays (assay ID: C___2336609_20). The reaction components were obtained from Applied Biosystems (Foster City, CA, USA). A 25-mL reaction volume was used for the PCR, containing 1.25 mL of TaqMan SNP genotyping assay mix, 12.5 mL of TaqMan genotyping master mix, no AmpErase UNG (2×), and 20 ng of genomic DNA, which was diluted to 11.25 mL, utilizing DNase-RNase-free water. After that, the amplification was done using the StepOne™ real-time PCR equipment (Applied Biosystems, Foster City, CA, USA). SDS software version 1.3.1 (Applied Biosystems) was utilized for allelic discrimination. The procedures were conducted blindly with respect to the vitiligo/psoriasis/control groups.

2.9 Statistical analyses

For the statistical analysis, we depended on Statistical Package for the Social Sciences (SPSS) software, version 24 in addition to Microsoft® Excel 2010. Percentage and frequency were adopted to express qualitative variables, with the usage of the chi-squared (χ^2) test besides Fisher's exact tests to compare between groups. Mean ± standard deviation (SD) was adopted to express quantitative variables, with the use of Student's *t*-test and one-way ANOVA tests to compare quantitative variables with a normal distribution, while the Mann–Whitney U test and Kruskal–Wallis test were used to compare quantitative variables with a non-normal distribution. Statistical significance was determined by a value of p below 0.05. Moreover, the calculation of odds ratios (OR) was performed using a 95% confidence interval (CI). Hardy–Weinberg equilibrium (HWE) was calculated as well. In addition, SNPSTAT was used for genetic models.

5 http://www.genemania.org

6 https://string-db.org/

3 Results

3.1 The analysis of the variant's effect on the MBL protein function

The utilization of the MutPred2 tool revealed a score of 0.849 with rs1800450 SNP, which suggested pathogenicity. Furthermore, molecular alterations were predicted, including losing acetylation at K59 with a probability of 0.27 and a value of *p* of 7.3e-03, gaining SUMOylation at K56 with a probability of 0.25 and a value of *p* of 7.8e-03, losing methylation at K56 with a probability of 0.23 and a value of *p* of 3.1e-03, gaining ubiquitylation at K49 with a probability of 0.17 and a value of *p* of 0.02, and gaining proteolytic cleavage at D53 with a probability of 0.13 and a value of *p* of 0.02.

3.2 Analyzing the variant's effect on the MBL protein stability

The analysis of how the nominated SNP could affect MBL stability was conducted using the Mu-Pro tool. The SNP was predicted by the Mu-Pro tool to decrease the stability with ΔΔG of −0.69653508.

3.3 Secondary structure analysis

The SOPMA tool was used to analyze the anticipated secondary structures of MBL, as demonstrated in Figure 2A. SOPMA revealed that 75 residues were linked with an alpha helix (30.24%), 39 with an extended strand (15.37%), 122 with a random coil (49.19%), and 12 with a beta turn (4.84%). Position 54 was found to be associated with the random coil. Furthermore, the secondary structures with the presence of the selected mutation were predicted, as shown in Figure 2B. In this case, SOPMA exposed that 69 residues were linked with an alpha helix (27.82%), 38 with an extended strand (15.32%), 133 with a random coil (53.63%), and 8 with a beta turn (3.23%). Position 54 was found to be associated with the random coil in this case as well.

3.4 The post-translational modification analysis

To identify the expected post-translational modification sites, the MusiteDeep server was utilized; the anticipated post-translational modification sites with MBL protein are shown in Figure 3A. Moreover, the anticipated post-translational modification sites with MBL in the case of the nominated mutation are shown in Figure 3B, and loss of SUMOylation at position 56 could be noticed in this case.

3.5 Gene–gene interaction examination

Utilizing the GeneMANIA tool, the *MBL2* gene's gene–gene interaction was examined, and the results showed 20 genes with

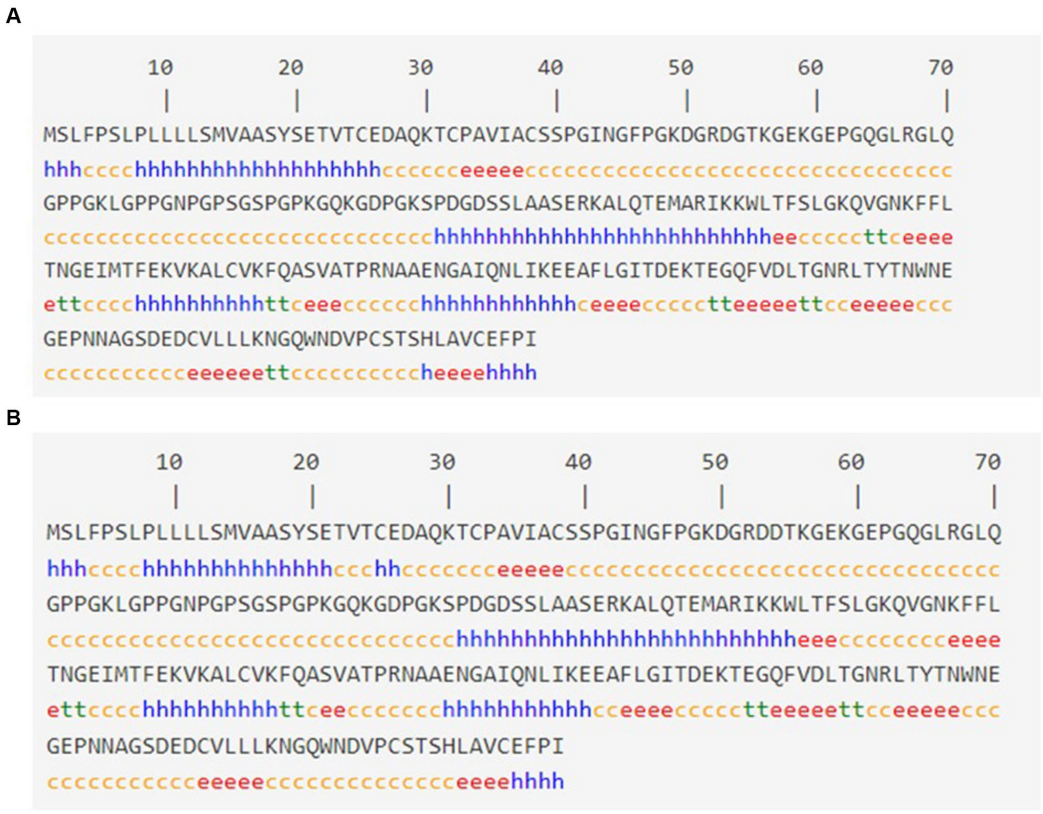

FIGURE 2
Secondary structure analysis: **(A)** Secondary structure with wild MBL protein and **(B)** secondary structure with MBL protein in case of the selected mutation, produced by the SOPMA server. Alpha helix: (h), Extended strand: (e), Beta turn: (t), Random coil: (c).

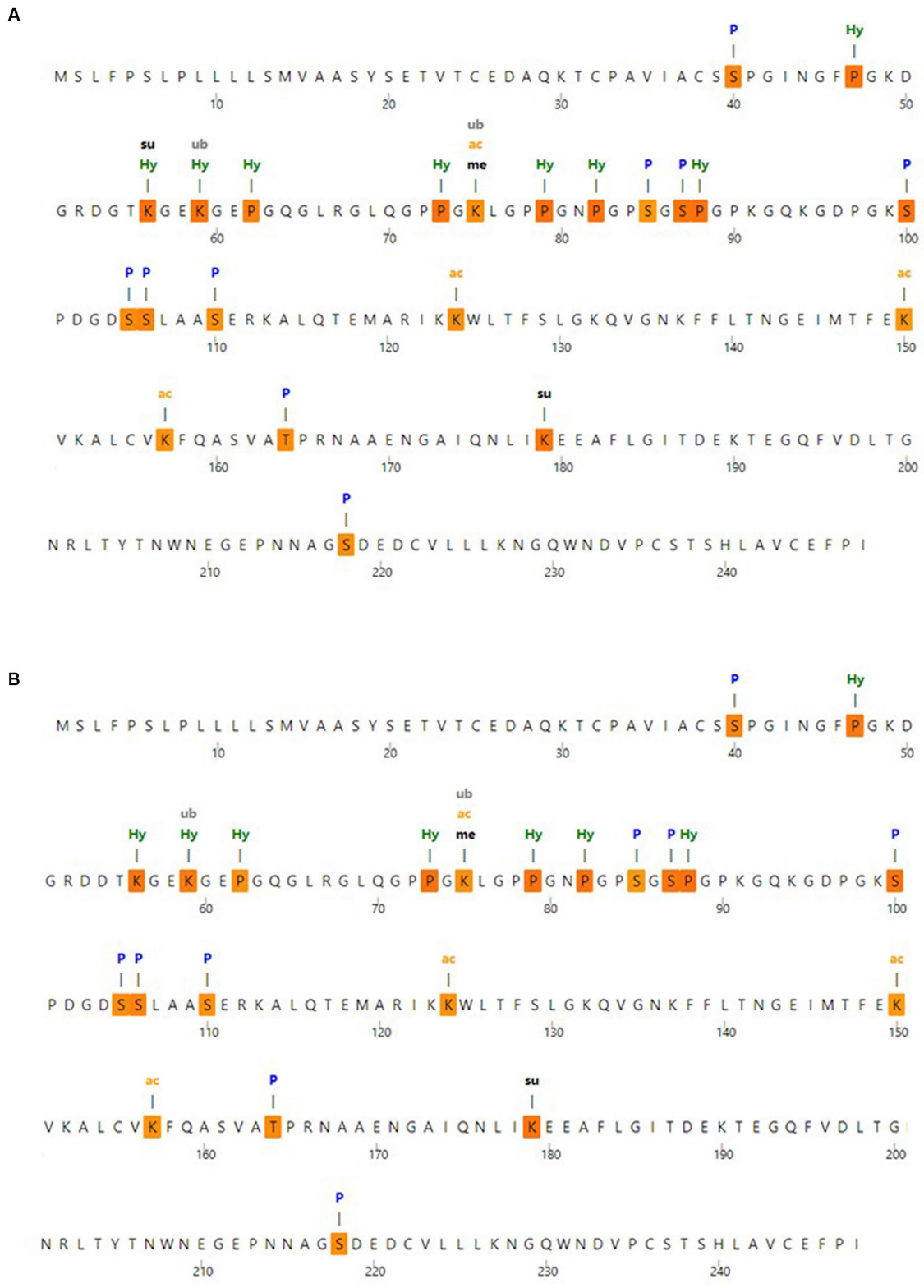

FIGURE 3
Post-translational modification analysis **(A)** post-translational modification sites with wild MBL protein and **(B)** post-translational modification sites with MBL protein in case of the selected mutation, produced by the MusiteDeep server. P, Phosphorylation; me, Methylation; gl, Glycosylation; Su, SUMOylation; Ub, Ubiquitination; ac, Acetyllysine; pa, Palmitoylation; Pc, Pyrrolidone carboxylic acid; Hy, Hydroxylation.

the strongest connections to the *MBL2* gene (Figure 4). Among these genes, the mannan-binding lectin serine peptidase 1 gene (MASP1) ranked first. Then, the mannan-binding lectin serine peptidase 2 gene (MASP2) held the second rank. After that, complement C2 (C2) held the third rank.

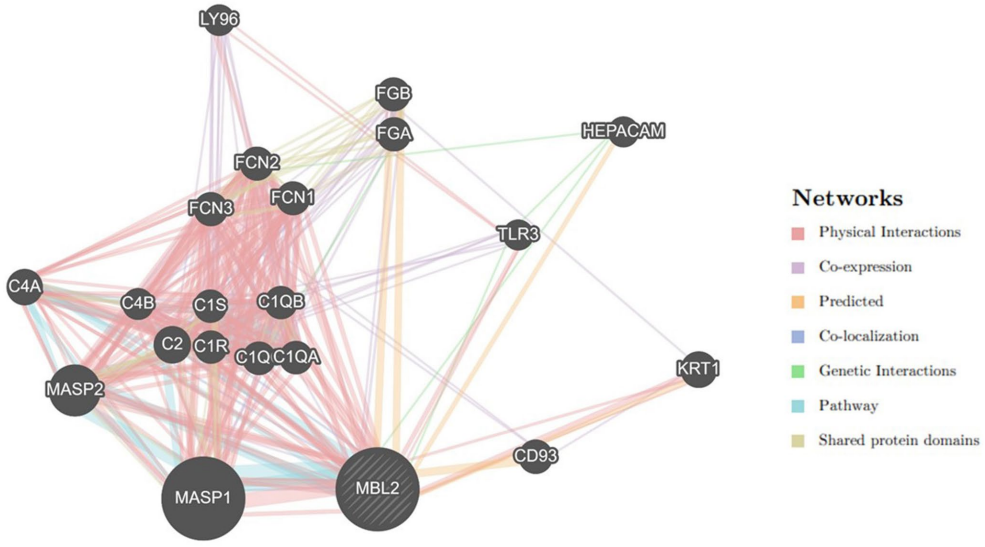

FIGURE 4
Network of *MBL2* gene–gene interactions, produced by GeneMANIA.

3.6 Protein–protein interaction analysis

Depending on the STRING database, the MBL protein–protein interaction network was generated, and the 10 most interactive proteins were predicted, as could be shown in detail in Figure 5.

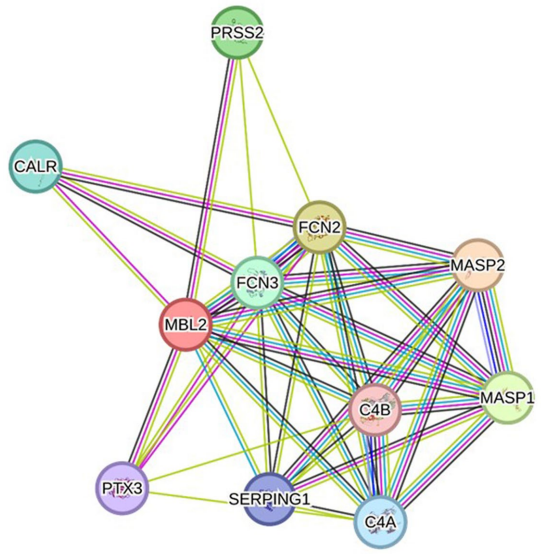

FIGURE 5
Protein–protein interaction network related to MBL protein, generated by the STRING databases (version 12). The nodes signify proteins, whereas the edges indicate protein–protein associations. The edges could possess any of the seven distinct colored lines, with distinct indications. Black lines: co-expression evidence; light blue lines: database evidence; yellow lines: text-mining evidence; purple lines: experimental evidence; blue lines: co-occurrence evidence; green lines: neighborhood evidence; red lines: fusion evidence. MASP2: Mannan-binding lectin serine protease 2 A chain, FCN2: Ficolin-2, MASP1: Mannan-binding lectin serine peptidase 1, PRSS2: Trypsin-2, FCN3: Ficolin-3, CALR: Calreticulin, C4A: Complement C4-A alpha chain, SERPING1: Plasma protease C1 inhibitor, PTX3: Pentraxin-related protein PTX, C4B: Complement C4-B alpha chain (https://string-db.org/ accessed on 23 December 2023).

3.7 Baseline characteristics of population understudy

Our study comprised a total of 489 subjects, including 90 vitiligo patients, 99 psoriasis patients, and a comparable number of healthy control subjects, i.e., 300. The baseline features of enrolled subjects are shown in Table 1. The age of the control, vitiligo, and psoriasis groups ranged from 18 to 63 years, from 6 to 60 years, and from 18 to 60 years, respectively. Meanwhile, the percentage of men in these three groups was 50.3, 61.1, and 45.5%. Considering special habits among the study groups, 53.8% of the control group participants were smokers, 64.4% of vitiligo patients were smokers, and 56.6% of psoriasis patients were smokers. Regarding family history, 18.9% of vitiligo patients had a positive family history, while 15.2% of psoriasis patients had a positive history. No significant statistical difference was found with these baseline characteristics except with the age variable.

3.8 Clinicopathological data among patient groups understudy

Different clinicopathological features between patient groups are shown in Table 2. The mean of disease duration for vitiligo patients and psoriasis patients was 9.2 ± 9.61 and 7.0 ± 6.25, respectively, with no statistical significance. Furthermore, the mean age regarding disease onset for these two groups was 25.93 ± 15.98 and 35.07 ± 13.34, respectively, with statistically significant differences found between these groups. Studying Vitiligo Disease Activity (VIDA) revealed the presence of 37.5, 3.4, 13.6, 20.5, 22.7, and 2.3% of vitiligo patients in stages −1, 0, 1, 2, 3, and 4, respectively, showing statistical significance. Meanwhile, studying psoriasis severity revealed the presence of 45% of psoriasis patients in the mild subgroup, 31% in

TABLE 1 Description of the population understudy.

Variable	Controls ($n = 300$)	Vitiligo ($n = 90$)	Psoriasis ($n = 99$)	Value of p
Age, mean (year) *t*	32.0 ± 9.5	34.4 ± 16.6	42.0 ± 14.0	0.001*
Min. – Max.	18–63	6–60	18–60	
Sex M				
Female	149 (49.7%)	35 (38.9%)	54 (54.5%)	0.75
Male	151 (50.3%)	55 (61.1%)	45 (45.5%)	
Smoking M				
Smoker	161 (53.8%)	58 (64.4%)	56 (56.6%)	0.38
Non-smoker	138 (46.2%)	32 (35.6%)	43 (43.4%)	
Family history M				
Positive	-	17 (18.9%)	15 (15.2%)	0.562
Negative	-	73 (81.1%)	84 (84.8%)	

Data are shown as mean ± SD or number (percentage). A value of p of < 0.05 was considered statistically significant. *t*, independent *t*-test between the psoriasis and vitiligo groups (parametric). M, the Mann–Whitney test between the psoriasis and vitiligo groups (non-parametric).

TABLE 2 Clinicopathological data among patient groups' understudy.

Variable	Vitiligo patients ($n = 90$)	Psoriasis patients ($n = 99$)	Value of p
Disease duration *t*			
Mean ± SD	9.2 ± 9.61	7.0 ± 6.25	0.514
Min. – Max.	0.1–39	1.0–30	
Age of onset *t*			
Mean ± SD	25.93 ± 15.98	35.07 ± 13.34	<0.001*
Min. – Max.	5–59	3–59	
VIDA			
Stage −1	33 (37.5%)		
Stage 0	3 (3.4%)		<0.001*
Stage 1	12 (13.6%)		
Stage 2	18 (20.5%)		
Stage 3	20 (22.7%)		
Stage 4	2 (2.3%)		
Psoriasis severity			
Mild		45 (45%)	0.029*
Moderate		31 (31%)	
Severe		24 (24%)	
VASI *t*			
Mean ± SD	0.51 ± 0.30		<0.001*
Min. – Max.	0.10–1.0		
PASI *t*			
Mean SD		13.66 ± 9.05	<0.001*
Min. – Max.		2.0–41	

Data are shown as mean ± SD or number (percentage). *A significant value of p of < 0.05 was considered statistically significant. One-sample chi-squared test was used. *t*, Independent *t*-test between the psoriasis and vitiligo groups (parametric). PASI, Psoriasis Area Severity Index; VASI, Vitiligo Area Severity Index; VIDA, Vitiligo Disease Activity.

the moderate subgroup, and 24% in the severe subgroup, with statistical significance found as well. The mean of the Vitiligo Area Severity Index (VASI) was 0.51 ± 0.30 in the vitiligo group, while the mean of the Psoriasis Area Severity Index (PASI) was 13.66 ± 9.05 in the psoriasis group, with statistical significance indicated by a value of p less than 0.001.

3.9 *MBL2* genotype analysis

Table 3 shows the genotype and allele frequencies of *MBL2* gene rs1800450 SNP among the studied skin autoimmune diseases. As shown, the allele frequencies in the control group were 514 (85.67%) and 86 (14.33%) for allele A and allele B, respectively. Regarding the psoriasis group, allele frequencies were 174 (87.88%) and 24 (12.12%), respectively, while with the vitiligo group, allele frequencies were 164 (91.11%) and 16 (8.89%), respectively (Figure 6A).

Meanwhile, the genotype frequencies for wild genotype, heterozygous genotype, and mutant genotype with the control group were 222 (74%), 70 (23.3%), and 8 (2.7%), respectively. Regarding the psoriasis group, the genotype frequencies were 76 (76.8%), 22 (22.2%), and 1 (1%), respectively, while with the vitiligo group, the genotype frequencies were 75 (83.3%), 14 (15.6%), and 1 (1.1%), respectively (Figure 6B). The genotype distribution in study groups showed consistency with Hardy–Weinberg equilibrium. In addition, allele frequencies for allele B in different populations for the rs1800450 SNP are shown in Figure 7, according to the 1,000 Genomes project, along with the allele frequencies in the study groups.

3.10 Genetic association models analysis

The analysis of *MBL2* gene rs1800450 SNP with psoriasis risk was performed under all genetic association models, as shown in Table 4; nevertheless, no statistical significance was uncovered. In addition, the analysis of the rs1800450 variant with vitiligo risk was conducted under all genetic association models as well; however, no significant statistical difference was found, as shown in Table 5.

TABLE 3 Genotype and allele frequencies of the *MBL2* gene rs1800450 SNP in the study population.

	MBL2 (rs1800450)			
	All ($n = 489$)	Control ($n = 300$)	Psoriasis ($n = 99$)	Vitiligo ($n = 90$)
Genotypes				
A/A	373 (76.3%)	222 (74%)	76 (76.8%)	75 (83.3%)
A/B	106 (21.7%)	70 (23.3%)	22 (22.2%)	14 (15.6%)
B/B	10 (2.0%)	8 (2.7%)	1 (1%)	1 (1.1%)
Alleles				
A	852 (87.1%)	514 (85.67%)	174 (87.88%)	164 (91.11%)
B	126 (12.9%)	86 (14.33%)	24 (12.12%)	16 (8.89%)
P HWE	0.74981	0.85667	0.91214	0.93175

Data are shown as numbers (percentage). The Hardy–Weinberg equilibrium was calculated using an online calculator (https://gene-calc.pl/hardy-weinberg-page accessed on 12 July 2023).

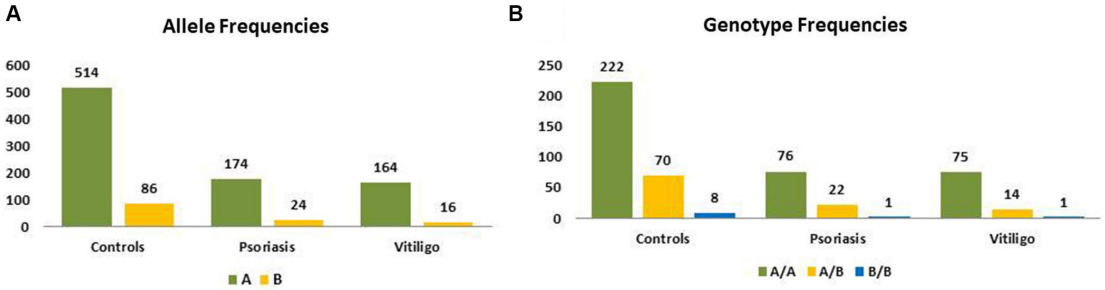

FIGURE 6
Genotype and allele frequencies of *MBL2* gene rs1800450 SNP among skin autoimmune diseases. **(A)** Allele frequencies and **(B)** genotype frequencies (controls = 300 subjects, psoriasis = 99 patients, vitiligo = 90 patients).

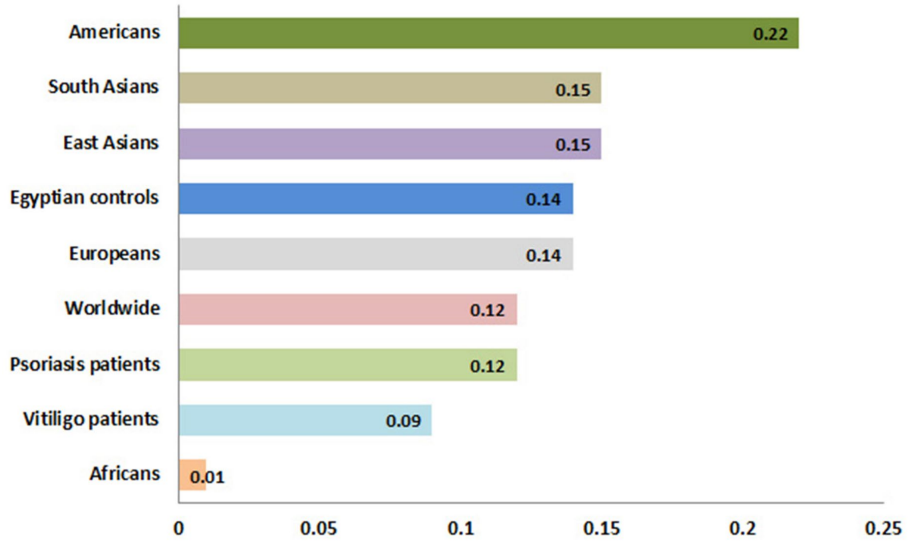

FIGURE 7
Allele frequencies for *MBL2* gene rs1800450 SNP according to the 1,000 Genomes Project. This diagram was constructed based on Ensembl, https://www.ensembl.org/index.html (accessed on 17 July 2023).

TABLE 4 Genetic association models for the *MBL2* gene rs1800450 SNP with psoriasis risk (total *n* = 399, controls = 300, psoriasis patients = 99, adjusted by age and sex).

Model	Genotype	Controls	Psoriasis	OR (95% CI)	Value of *p*
Codominant	A/A	222 (74%)	76 (76.8%)	1	0.7
	A/B	70 (23.3%)	22 (22.2%)	0.99 (0.55–1.79)	
	B/B	8 (2.7%)	1 (1%)	0.42 (0.05–3.80)	
Dominant	A/A	222 (74%)	76 (76.8%)	1	0.83
	A/B – B/B	78 (26%)	23 (23.2%)	0.94 (0.53–1.66)	
Recessive	A/A – A/B	292 (97.3%)	98 (99%)	1	0.4
	B/B	8 (2.7%)	1 (1%)	0.42 (0.05–3.79)	
Overdominant	A/A – B/B	230 (76.7%)	77 (77.8%)	1	0.97
	A/B	70 (23.3%)	22 (22.2%)	1.01 (0.56–1.82)	
Log-additive				0.89 (0.54–1.50)	0.67

CI, confidence interval; OR, Odds ratio; *p*, a value of *p* for comparison between the studied groups.

TABLE 5 Genetic association models for the *MBL2* gene rs1800450 SNP with vitiligo risk (total *n* = 390, controls = 300, vitiligo patients = 90, adjusted by age and sex).

Model	Genotype	Controls	Vitiligo	OR (95% CI)	Value of *p*
Codominant	A/A	222 (74%)	75 (83.3%)	1	0.2
	A/B	70 (23.3%)	14 (15.6%)	0.62 (0.33–1.16)	
	B/B	8 (2.7%)	1 (1.1%)	0.38 (0.05–3.14)	
Dominant	A/A	222 (74%)	75 (83.3%)	1	0.084
	A/B – B/B	78 (26%)	15 (16.7%)	0.59 (0.32–1.09)	
Recessive	A/A – A/B	292 (97.3%)	89 (98.9%)	1	0.37
	B/B	8 (2.7%)	1 (1.1%)	0.42 (0.05–3.44)	
Overdominant	A/A – B/B	230 (76.7%)	76 (84.4%)	1	0.14
	A/B	70 (23.3%)	14 (15.6%)	0.63 (0.33–1.19)	
Log-additive				0.62 (0.35–1.08)	0.074

CI, confidence interval; OR, Odds ratio; *p*, a value of *p* for comparison between the studied groups.

TABLE 6 Correlation matrix between psoriasis patients' demographic data, clinicopathological features, and the *MBL2* gene rs1800450 SNP genotyping.

		Age	Sex	Smoking	FH	Duration	Disease onset	PASI	Severity	Disease	SNP
Age	*r*	1	−0.024	0.032	−0.247*	0.366**	0.890**	0.157	0.154	0.310**	−0.061
	Value of *p*	-	0.628	0.524	0.014	0	0	0.12	0.128	0	0.22
Sex	*r*	−0.024	1	−0.034	0.216*	0.026	−0.366**	−0.064	−0.036	0.042	0.039
	Value of *p*	0.628	-	0.495	0.032	0.801	0	0.53	0.727	0.401	0.438
Smoking	*r*	0.032	−0.034	1	−0.084	0.013	0.114	0.107	0.121	0.025	−0.025
	Value of *p*	0.524	0.495	-	0.406	0.899	0.262	0.292	0.234	0.617	0.618
FH	*r*	−0.247*	0.216*	−0.084	1	−0.032	−0.230*	−0.011	−0.014	-	0.032
	Value of *p*	0.014	0.032	0.406	-	0.755	0.022	0.911	0.888	-	0.756
Duration	*r*	0.366**	0.026	0.013	−0.032	1	−0.039	0.381**	0.297**	-	−0.041
	Value of *p*	0	0.801	0.899	0.755	-	0.698	0	0.003	-	0.685
Disease onset	*r*	0.890**	−0.366**	0.114	−0.230*	−0.039	1	0.001	0.032	-	−0.079
	Value of *p*	0	0	0.262	0.022	0.698	-	0.99	0.755	-	0.439
PASI	*r*	0.157	−0.064	0.107	−0.011	0.381**	0.001	1	0.930**	-	0.063
	Value of *p*	0.12	0.53	0.292	0.911	0	0.99	.	0	-	0.535
Severity	*r*	0.154	−0.036	0.121	−0.014	0.297**	0.032	0.930**	1	-	0.053
	Value of *p*	0.128	0.727	0.234	0.888	0.003	0.755	0	-	-	0.602
Disease	*r*	0.310**	0.042	0.025	-	-	-	-	-	1	−0.034
	Value of *p*	0	0.401	0.617	-	-	-	-	-	-	0.497
SNP	*r*	−0.061	0.039	−0.025	0.032	−0.041	−0.079	0.063	0.053	−0.034	1
	Value of *p*	0.22	0.438	0.618	0.756	0.685	0.439	0.535	0.602	0.497	-

The correlation coefficient (*r*) represents the value for Spearman's correlation analysis and its value of *p*s. Shaded boxes enclose values that are statistically significant at either $p < 0.05$ (*) or $p < 0.01$ (**). FH, family history; PASI, Psoriasis Area and Severity Index; SNP, single nucleotide polymorphism.

3.11 Correlation analysis of the studied variant, psoriasis patients' demographic data, and clinicopathological features

The inter-relationships between rs1800450 SNP genotyping, psoriasis cases' demographic data, and their clinicopathological features were analyzed and shown in Table 6. A direct and significant correlation was found between age and duration ($r = 0.366$; $p < 0.001$), disease onset ($r = 0.890$; $p < 0.001$), and disease ($r = 0.310$; $p < 0.001$). Sex had a direct and significant correlation with family history ($r = 0.890$; $p < 0.05$). Moreover, duration had a direct and significant correlation with age ($r = 0.366$; $p < 0.001$), PASI ($r = 0.381$; $p < 0.001$), and severity ($r = 0.297$; $p = 0.003$). Furthermore, PASI showed a direct and significant correlation with duration ($r = 0.381$; $p < 0.001$) as well as with severity and vice versa ($r = 0.930$; $p < 0.001$).

3.12 Correlation analysis of the studied variant, vitiligo patients' demographic data, and clinicopathological features

The analysis of interrelationships between rs1800450 SNP and vitiligo cases' demographic data and their clinicopathological features was performed, as shown in Table 7. Age showed a direct and significant correlation with VASI ($r=0.271$; $p=0.01$), duration ($r=0.438$; $p<0.001$), and disease onset ($r=0.812$; $p<0.001$). VASI showed a direct and significant correlation with age ($r=0.271$; $p=0.01$) as well as with duration ($r=0.447$; $p<0.001$) and vice versa.

4 Discussion

Genetics were found to have primary roles in both psoriasis and vitiligo disorders (7, 33–36). SNPs signify the most prevalent type of genetic variation (37), with massive attention attracted to the missense SNP subtype that may result in pathogenic consequences (38). Despite the importance of MBL in the immune system and the discovered links between MBL2 polymorphisms and autoimmune diseases (39), only a few studies to our knowledge have addressed the association between these variants with vitiligo and psoriasis disorders.

By applying the *in silico* methods, the SNP was predicted to affect protein function and protein stability. Furthermore, a protein's secondary structure plays crucial functions in both the structure and folding of the protein, demonstrating the relevance of this structure (40); therefore, secondary structure analysis was performed using the SOPMA server. The secondary structure analysis detected no change regarding the alignment of the selected position. Moreover, PTMs could influence protein functions and have the potential to impact many aspects of protein biology, such as stability, cellular localization, and interaction with co-factors (41). Therefore, PTM analysis was performed, which revealed a loss of SUMOylation in position 56 with this mutation. In addition, gene–gene interaction examination revealed the 20 genes with the strongest connections to the *MBL2* gene, which might be affected by *MBL2* variants. The significance of researching gene–gene interactions while examining disease–gene associations was demonstrated by the verified presence of interaction between various genetic loci (42). Furthermore, mutations that cause diseases show a tendency to cause instability in protein–protein interactions (43). Consequently, the interaction pattern of MBL in conjunction with other proteins was predicted using the STRING database showing the 10 most prominent interactors.

Our experimental analysis revealed genotype frequencies of 76.3, 21.7, and 2% for AA, AB, and BB, respectively, which were similar to previously conducted studies on Egyptian populations. A 2018 study conducted by Hammad et al. showed genotype frequencies of 79.7, 18.6, and 1.7% with AA, AB, and BB, respectively (44). Moreover, genotype frequencies in another Egyptian study by Gomaa et al. exposed genotype frequencies of 64.6, 28.8, and 6.7%, respectively (12). Although the heterozygous allele displayed a high frequency in these studies along with our

TABLE 7 Correlation matrix between vitiligo patients' demographic data, clinicopathological features, and the *MBL2* gene rs1800450 SNP genotyping.

		Age	Sex	Smoking	FH	VASI	VIDA	Duration	Disease onset	Disease	SNP
Age	*r*	1	0.096	0.039	0.05	0.271**	−0.325**	0.438**	0.812**	0.061	−0.04
	Value of *p*	-	0.058	0.437	0.642	0.01	0.002	0	0	0.233	0.43
Sex	*r*	0.096	1	−0.014	−0.094	0.041	−0.152	0.123	0.135	−0.091	−0.04
	Value of *p*	0.058	-	0.787	0.379	0.699	0.152	0.248	0.205	0.073	0.425
Smoking	*r*	0.039	−0.014	1	0.121	0.021	−0.083	−0.013	0.097	0.092	−0.012
	Value of *p*	0.437	0.787	-	0.255	0.843	0.438	0.907	0.365	0.071	0.818
FH	*r*	0.05	−0.094	0.121	1	0.029	0.058	−0.058	0.039	-	−0.065
	Value of *p*	0.642	0.379	0.255	-	0.788	0.587	0.587	0.713	-	0.542
VASI	*r*	0.271**	0.041	0.021	0.029	1	−0.254*	0.447**	0.068	-	−0.024
	Value of *p*	0.01	0.699	0.843	0.788	-	0.016	0	0.523	-	0.822
VIDA	*r*	−0.325**	−0.152	−0.083	0.058	−0.254*	1	−0.490**	−0.143	-	−0.023
	Value of *p*	0.002	0.152	0.438	0.587	0.016	-	0	0.178	-	0.828
Duration	*r*	0.438**	0.123	−0.013	−0.058	0.447**	−0.490**	1	−0.063	-	0.078
	Value of *p*	0	0.248	0.907	0.587	0	0	-	0.557	-	0.463
Disease onset	*r*	0.812**	0.135	0.097	0.039	0.068	−0.143	−0.063	1	-	−0.045
	Value of *p*	0	0.205	0.365	0.713	0.523	0.178	0.557	.	-	0.675
Disease	*r*	0.061	−0.091	0.092	-	-	-	-	-	1	−0.094
	Value of *p*	0.233	0.073	0.071	-	-	-	-	-	-	0.065
SNP	*r*	−0.04	−0.04	−0.012	−0.065	−0.024	−0.023	0.078	−0.045	−0.094	1
	p-value	0.43	0.425	0.818	0.542	0.822	0.828	0.463	0.675	0.065	-

The correlation coefficient (*r*) represents the value for Spearman's correlation analysis and its value of *p*s. Shaded boxes enclose values that are statistically significant at either $p<0.05$ (*) or $p<0.01$ (**). FH, family history; SNP, single nucleotide polymorphism; VASI, Vitiligo Area and Severity Index; VIDA, Vitiligo Disease Activity Score.

study, other previous studies showed even more elevated frequencies with the AB allele (9, 45, 46). With only limited exceptions, these remarkably elevated frequencies of this variant allele and other MBL deficiency variant alleles were detected worldwide. It is noteworthy that the allele B has nearly substituted the allele A in specific Indian populations in South America (47). This distinct pattern has raised inquiries about MBL significance and whether these mutations are associated with some kind of biological benefits (48). Different models were introduced in pursuit of the clarification of this phenomenon, resulting in an unsettled debate (49, 50). One explanation assumed a gained protection with these variants due to the associating low MBL levels that could be advantageous in case of tissue damage resulting from a strong inflammatory response (51). This postulation gave explanation to some previous findings that favored the clinical outcome with genotypes of low MBL in case of some disorders; these genotypes were associated with less severe autoimmune manifestations in primary Sjögren's syndrome (10); moreover, tuberculosis protection was correlated with heterozygosity in MBL alleles, which is responsible for low levels of MBL (52). Furthermore, similar protective benefit was found for these genotypes against visceral leishmaniasis (53). In addition, the elevated MBL levels were identified to confer an augmented risk of ischemic heart disease along with myocardial infarction in rheumatoid arthritis (54, 55). On the other hand, another explanation excluded the existence of selective pressure and displayed the absence of statistical support for this pressure (56, 57), further advocating the redundancy of MBL in human immunity as a result (57).

Subsequently, the association between codon 54 polymorphism and psoriasis risk was investigated. Nevertheless, no significant association was revealed with all genetic association models. Moreover, no significant correlation was discovered between rs1800450 SNP and the clinicopathological features of psoriasis. The significance of MBL and MBL deficiency in disease pathogenesis has been a source of ambiguity and queries as studies indicate that low or elevated MBL levels show a damaging, protecting, or no effect on susceptibility to diseases (51), which could also be the case here, as recent studies have revealed the important role of MBL in psoriasis pathogenesis. One study found that MBL may aggravate psoriatic skin inflammation by assisting the infiltration of the neutrophils in psoriatic lesions (58). Another study revealed that psoriasis patients displayed higher MBL levels than control subjects and that the level of MBL correlated positively with PASI score and psoriasis severity, and found that MBL assisted the differentiating and infiltrating processes of plasmacytoid dendritic cells in initial skin lesions of psoriasis, leading to the aggravation of disease severity (59). Therefore, the reason for the lack of protective effect with MBL-deficient genotypes needs to be explained. One suggested explanation for the neutrality of MBL deficiency in previous studies was by assuming the redundancy of MBL and the existence of adequate substituting immune mechanisms (9, 60), as other pathways or molecules have demonstrated that they may make up for the deficiency of MBL (57, 61, 62).

The scarcity of previous related studies addressing *MBL2* variants and psoriasis was noticed. The previously conducted study in the Turkish population included 50 psoriasis vulgaris patients along with 53 control subjects and found a significant association between rs1800450 SNP and psoriasis (20), which differs from our findings and with the potential protective role of MBL as well. However, previous analyses related to the impacts of MBL2 variants have pointed out the role of underpowered studies in inconsistent results (63).

In addition, the analysis was extended to examine the relationship between codon 54 polymorphism and vitiligo risk as well. However, the analysis also did not reveal a significant association with all genetic association models. Furthermore, no significant correlation was uncovered between this SNP and clinicopathological features of vitiligo. To our knowledge, only a limited number of previous studies dealt with this topic. Our result complies with the findings of both Karkucak et al. and Dwivedi et al., who investigated this issue; Karkucak et al. conducted their study on a Turkish population comprising 101 vitiligo cases and 101 control subjects, reporting the absence of the association of vitiligo risk with rs1800450 SNP (18). Moreover, Dwivedi et al. conducted their study on an Indian population comprising 92 cases affected by generalized vitiligo and 94 control subjects, revealing the absence of the association of this disease with rs1800450 SNP and other *MBL2* SNPs (17). However, one more Turkish study, including 40 vitiligo patients and 50 healthy subjects, was conducted that suggested a role for rs1800450 in vitiligo susceptibility (19). Complying with the findings of the two studies with a larger sample size gave robustness to our findings. These findings also suggested the notion of redundancy regarding MBL. Finally, the current study had its limitations as well. The patient sample size included in this preliminary study was relatively small. As a result, there is a need for further studies to be conducted on the Egyptian population with a larger patient sample size that encompasses diverse geographic areas. Furthermore, despite investigating the selected variant's relationship with two autoimmune skin disorders through a combination of experimental and computational approaches, there is still a need for more forthcoming studies that analyze several genetic factors and combine them with environmental factors to understand the full picture regarding these complex diseases.

In conclusion, the *in silico* analysis forecasted the consequences of the presence of the selected mutation on the protein's function, stability, PTMs, and secondary structures. Moreover, our analysis emphasized the same frequency pattern of the codon 54 variant among the Egyptian population in comparison with other populations and highlighted the unsettled ambiguity regarding this pattern. The studied variant showed no association with the risk of psoriasis and vitiligo as well. Furthermore, this mutation did not significantly correlate with the clinicopathological data for both diseases. These findings supported the redundancy of MBL and the worth of the compensatory mechanisms.

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 the Research Ethics Committee at the Faculty of Medicine, Suez Canal University, Egypt. 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 or their next of kin.

Author contributions

MB: Conceptualization, Investigation, Methodology, Writing – original draft. NT: Investigation, Writing – review & editing. RE: Funding acquisition, Writing – review & editing. DN: Funding acquisition, Writing – review & editing. DA: Funding acquisition, Writing – review & editing. EF: Funding acquisition, Writing – review & editing. WE: Writing – review & editing. HA: Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Writing – original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this research through large group research project under grant number RGP2/17/44. This research was also supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R155), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Acknowledgments

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through large group Research Project under grant number RGP2/17/44. The authors extend their appreciation also to Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R155), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Conflict of interest

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

Publisher's note

References

1. Sharquie KE, Salman HA, Yaseen AK. Psoriasis and vitiligo are close relatives. *Clin Cosmet Investig Dermatol*. (2017) 10:341–5. doi: 10.2147/CCID.S142819

2. Rendon A, Schäkel K. Psoriasis pathogenesis and treatment. *Int J Mol Sci*. (2019) 20:1475. doi: 10.3390/ijms20061475

3. Tokuyama M, Mabuchi T. New treatment addressing the pathogenesis of psoriasis. *Int J Mol Sci*. (2020) 21:7488. doi: 10.3390/ijms21207488

4. Kamiya K, Kishimoto M, Sugai J, Komine M, Ohtsuki M. Risk factors for the development of psoriasis. *Int J Mol Sci*. (2019) 20:4347. doi: 10.3390/ijms20184347

5. Feng Y, Lu Y. Advances in vitiligo: update on therapeutic targets. *Front Immunol*. (2022) 13:986918. doi: 10.3389/fimmu.2022.986918

6. Spritz RA, Santorico SA. The genetic basis of vitiligo. *J Invest Dermatol*. (2021) 141:265–73. doi: 10.1016/j.jid.2020.06.004

7. Marchioro HZ, Silva de Castro CC, Fava VM, Sakiyama PH, Dellatorre G, Miot HA. Update on the pathogenesis of vitiligo. *An Bras Dermatol*. (2022) 97:478–90. doi: 10.1016/j.abd.2021.09.008

8. Kalia N, Singh J, Kaur M. The ambiguous role of mannose-binding lectin (MBL) in human immunity. *Open Med (Warsaw, Poland)*. (2021) 16:299–310. doi: 10.1515/med-2021-0239

9. Behairy MY, Abdelrahman AA, Abdallah HY, Ibrahim EE-DA, Hashem HR, Sayed AA, et al. Role of MBL2 polymorphisms in Sepsis and survival: a pilot study and in silico analysis. *Diagnostics*. (2022) 12:460. doi: 10.3390/diagnostics12020460

10. Ramos-Casals M, Brito-Zerón P, Soria N, Nardi N, Vargas A, Muñoz S, et al. Mannose-binding lectin-low genotypes are associated with milder systemic and immunological disease expression in primary Sjögren's syndrome. *Rheumatology (Oxford)*. (2009) 48:65–9. doi: 10.1093/rheumatology/ken411

11. Jacobson S, Larsson P, Åberg A-M, Johansson G, Winsö O, Söderberg S. Levels of mannose-binding lectin (MBL) associates with sepsis-related in-hospital mortality in women. *J Inflamm (Lond)*. (2020) 17:28. doi: 10.1186/s12950-020-00257-1

12. Gomaa MH, Ali SS, Fattouh AM, Hamza HS, Badr MM. MBL2 gene polymorphism rs1800450 and rheumatic fever with and without rheumatic heart disease: an Egyptian pilot study. *Pediatr Rheumatol Online J*. (2018) 16:24. doi: 10.1186/s12969-018-0245-x

13. Speletas M, Dadouli K, Syrakouli A, Gatselis N, Germanidis G, Mouchtouri VA, et al. MBL deficiency-causing B allele (rs1800450) as a risk factor for severe COVID-19. *Immunobiology*. (2021) 226:152136. doi: 10.1016/j.imbio.2021.152136

14. Tiyo BT, Vendramini ECL, de Souza VH, Colli CM, Alves HV, Sell AM, et al. Association of MBL2 exon 1 polymorphisms with multibacillary leprosy. *Front Immunol*. (2020) 11:1927. doi: 10.3389/fimmu.2020.01927

15. Zhang C, Zhu J, Li S-L, Wang H, Zhu Q-X. The association of mannose-binding lectin genetic polymorphisms with the risk of rheumatoid arthritis: a meta-analysis. *J Recept Signal Transduct Res*. (2015) 35:357–62. doi: 10.3109/10799893.2014.975247

16. Lee YH, Lee H-S, Choi SJ, Ji JD, Song GG. The association between the mannose-binding lectin codon 54 polymorphism and systemic lupus erythematosus: a meta-analysis update. *Mol Biol Rep*. (2012) 39:5569–74. doi: 10.1007/s11033-011-1361-6

17. Dwivedi M, Gupta K, Gulla KC, Laddha NC, Hajela K, Begum R. Lack of genetic association of promoter and structural variants of mannan-binding lectin (MBL2) gene with susceptibility to generalized vitiligo. *Br J Dermatol*. (2009) 161:63–9. doi: 10.1111/j.1365-2133.2009.09140.x

18. Karkucak M, Solak B, Turan H, Uslu E, Yakut T, Aliagaoglu C, et al. MBL2 gene polymorphism and risk of vitiligo in Turkish patients. *Int J Hum Genet*. (2015) 15:93–6. doi: 10.1080/09723757.2015.11886257

19. Onay H, Pehlivan M, Alper S, Ozkinay F, Pehlivan S. Might there be a link between mannose binding lectin and vitiligo? *Eur J Dermatol*. (2007) 17:146–8. doi: 10.1684/ejd.2007.0128

20. Turan H, Karkucak M, Yakut T, Ozsahin M, Gurlevik Z, Yanik ME, et al. Does MBL2 codon 54 polymorphism play a role in the pathogenesis of psoriasis? *Int J Dermatol*. (2014) 53:34–8. doi: 10.1111/j.1365-4632.2012.5657.x

21. Behairy MY, Abdelrahman AA, Toraih EA, Ibrahim EE-DA, Azab MM, Sayed AA, et al. Investigation of TLR2 and TLR4 polymorphisms and Sepsis susceptibility: computational and experimental approaches. *Int J Mol Sci*. (2022) 23:10982. doi: 10.3390/ijms231810982

22. Behairy MY, Abdelrahman A, Abdallah HY, Ibrahim EE-DA, Sayed AA, Azab MM. In silico analysis of missense variants of the C1qA gene related to infection and autoimmune diseases. *J Taibah Univ Med Sci*. (2022) 17:1074–82. doi: 10.1016/j.jtumed.2022.04.014

23. Choi BG, Hong JY, Hong JR, Hur MS, Kim SM, Lee YW, et al. The IL17F His161Arg polymorphism, a potential risk locus for psoriasis, increases serum levels of interleukin-17F in an Asian population. *Sci Rep*. (2019) 9:18921. doi: 10.1038/s41598-019-55062-5

24. Zhang Y, Liu J, Wang C, Liu J, Lu W. Toll-like receptors gene polymorphisms in autoimmune disease. *Front Immunol*. (2021) 12:672346. doi: 10.3389/fimmu.2021.672346

25. Pejaver V, Urresti J, Lugo-Martinez J, Pagel KA, Lin GN, Nam H-J, et al. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nat Commun*. (2020) 11:5918. doi: 10.1038/s41467-020-19669-x

26. Cheng J, Randall A, Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins*. (2006) 62:1125–32. doi: 10.1002/prot.20810

27. Geourjon C, Deléage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci*. (1995) 11:681–4. doi: 10.1093/bioinformatics/11.6.681

28. Wang D, Liang Y, Xu D. Capsule network for protein post-translational modification site prediction. *Bioinformatics*. (2019) 35:2386–94. doi: 10.1093/bioinformatics/bty977

29. Wang D, Liu D, Yuchi J, He F, Jiang Y, Cai S, et al. MusiteDeep: a deep-learning based webserver for protein post-translational modification site prediction and visualization. *Nucleic Acids Res*. (2020) 48:W140–w146. doi: 10.1093/nar/gkaa275

30. Wang D, Zeng S, Xu C, Qiu W, Liang Y, Joshi T, et al. MusiteDeep: a deep-learning framework for general and kinase-specific phosphorylation site prediction. *Bioinformatics*. (2017) 33:3909–16. doi: 10.1093/bioinformatics/btx496

31. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res*. (2010) 38:W214–20. doi: 10.1093/nar/gkq537

32. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res*. (2023) 51:D638–46. doi: 10.1093/nar/gkac1000

33. Abdallah HY, Abdelhamid NR, Mohammed EA, AbdElWahab NY, Tawfik NZ, Gomaa AHA, et al. Investigating melanogenesis-related microRNAs as disease biomarkers in vitiligo. *Sci Rep*. (2022) 12:13526. doi: 10.1038/s41598-022-17770-3

34. Abdallah HY, Faisal S, Tawfik NZ, Soliman NH, Kishk RM, Ellawindy A. Expression signature of immune-related MicroRNAs in autoimmune skin disease: psoriasis and vitiligo insights. *Mol Diagn Ther*. (2023) 27:405–23. doi: 10.1007/s40291-023-00646-1

35. Raharja A, Mahil SK, Barker JN. Psoriasis: a brief overview. *Clin Med*. (2021) 21:170–3. doi: 10.7861/clinmed.2021-0257

36. Tawfik NZ, Abdallah HY, Hassan R, Hosny A, Ghanem DE, Adel A, et al. PSORS1 locus genotyping profile in psoriasis: a pilot case-control study. *Diagnostics (Basel, Switzerland)*. (2022) 12:1035. doi: 10.3390/diagnostics12051035

37. Behairy MY, Soltan MA, Eldeen MA, Abdulhakim JA, Alnoman MM, Abdel-Daim MM, et al. HBD-2 variants and SARS-CoV-2: new insights into inter-individual susceptibility. *Front Immunol*. (2022) 13:1008463. doi: 10.3389/fimmu.2022.1008463

38. Behairy MY, Soltan MA, Adam MS, Refaat AM, Ezz EM, Albogami S, et al. Computational analysis of deleterious SNPs in NRAS to assess their potential correlation with carcinogenesis. *Front Genet*. (2022) 13:872845. doi: 10.3389/fgene.2022.872845

39. Araujo J, Segat L, Guimarães RL, Brandão LAC, Souza PER, Santos S, et al. Mannose binding lectin gene polymorphisms and associated auto-immune diseases in type 1 diabetes Brazilian patients. *Clin Immunol*. (2009) 131:254–9. doi: 10.1016/j.clim.2008.12.010

40. Ji Y-Y, Li Y-Q. The role of secondary structure in protein structure selection. *Eur Phys J E Soft Matter*. (2010) 32:103–7. doi: 10.1140/epje/i2010-10591-5

41. Brunmeir R, Xu F. Functional regulation of PPARs through post-translational modifications. *Int J Mol Sci*. (2018) 19:1738. doi: 10.3390/ijms19061738

42. Cordell HJ. Detecting gene-gene interactions that underlie human diseases. *Nat Rev Genet*. (2009) 10:392–404. doi: 10.1038/nrg2579

43. Teng S, Madej T, Panchenko A, Alexov E. Modeling effects of human single nucleotide polymorphisms on protein-protein interactions. *Biophys J*. (2009) 96:2178–88. doi: 10.1016/j.bpj.2008.12.3904

44. Hammad NM, El Badawy NE, Nasr AM, Ghramh HA, Al Kady LM. Mannose-binding lectin gene polymorphism and its association with susceptibility to recurrent vulvovaginal candidiasis. *Biomed Res Int*. (2018) 2018:7648152–8. doi: 10.1155/2018/7648152

45. Badawy M, Mosallam DS, Saber D, Madani H. Use of mannose-binding lectin gene polymorphisms and the serum MBL level for the early detection of neonatal Sepsis. *J Pediatr Genet*. (2018) 7:150–7. doi: 10.1055/s-0038-1675801

46. Nasr M, Marie A, Boghdadi G, Elsaid R, Salah E. Role of mannose binding lectin in response to candida antigen immunotherapy of warts. *J Dermatolog Treat*. (2021) 32:376–80. doi: 10.1080/09546634.2019.1662365

47. Garred P, Larsen F, Seyfarth J, Fujita R, Madsen HO. Mannose-binding lectin and its genetic variants. *Genes Immun*. (2006) 7:85–94. doi: 10.1038/sj.gene.6364283

48. Turner MW. Mannose-binding lectin (MBL) in health and disease. *Immunobiology*. (1998) 199:327–39. doi: 10.1016/S0171-2985(98)80037-5

49. Eisen DP, Osthoff M. If there is an evolutionary selection pressure for the high frequency of MBL2 polymorphisms, what is it? *Clin Exp Immunol*. (2014) 176:165–71. doi: 10.1111/cei.12241

50. Tereshchenko SY, Smolnikova MV, Freidin MB. Mannose-binding lectin gene polymorphisms in the East Siberia and Russian Arctic populations. *Immunogenetics*. (2020) 72:347–54. doi: 10.1007/s00251-020-01175-5

51. Heitzeneder S, Seidel M, Förster-Waldl E, Heitger A. Mannan-binding lectin deficiency – good news, bad news, doesn't matter? *Clin Immunol*. (2012) 143:22–38. doi: 10.1016/j.clim.2011.11.002

52. Søborg C, Madsen HO, Andersen AB, Lillebaek T, Kok-Jensen A, Garred P. Mannose-binding lectin polymorphisms in clinical tuberculosis. *J Infect Dis*. (2003) 188:777–82. doi: 10.1086/377183

53. Alonso DP, Ferreira AFB, Ribolla PEM, Santos IKFM, Cruz MSP, Carvalho FA, et al. Genotypes of the mannan-binding lectin gene and susceptibility to visceral leishmaniasis and clinical complications. *J Infect Dis*. (2007) 195:1212–7. doi: 10.1086/512683

54. Troelsen LN, Garred P, Jacobsen S. Mortality and predictors of mortality in rheumatoid arthritis--a role for mannose-binding lectin? *J Rheumatol*. (2010) 37:536–43. doi: 10.3899/jrheum.090812

55. Troelsen LN, Garred P, Madsen HO, Jacobsen S. Genetically determined high serum levels of mannose-binding lectin and agalactosyl IgG are associated with ischemic heart disease in rheumatoid arthritis. *Arthritis Rheum*. (2007) 56:21–9. doi: 10.1002/art.22302

56. Boldt ABW, Messias-Reason IJ, Meyer D, Schrago CG, Lang F, Lell B, et al. Phylogenetic nomenclature and evolution of mannose-binding lectin (MBL2) haplotypes. *BMC Genet*. (2010) 11:38. doi: 10.1186/1471-2156-11-38

57. Verdu P, Barreiro LB, Patin E, Gessain A, Cassar O, Kidd JR, et al. Evolutionary insights into the high worldwide prevalence of MBL2 deficiency alleles. *Hum Mol Genet*. (2006) 15:2650–8. doi: 10.1093/hmg/ddl193

58. Zeng J, Chen X, Lei K, Wang D, Lin L, Wang Y, et al. Mannan-binding lectin promotes keratinocyte to produce CXCL1 and enhances neutrophil infiltration at the early stages of psoriasis. *Exp Dermatol*. (2019) 28:1017–24. doi: 10.1111/exd.13995

59. Zeng J, Wang D, Luo J, Li L, Lin L, Li J, et al. Mannan-binding lectin exacerbates the severity of psoriasis by promoting plasmacytoid dendritic cell differentiation via the signal transducer and activator of transcription 3-interferon regulatory factor 8 axis. *J Dermatol*. (2022) 49:496–507. doi: 10.1111/1346-8138.16323

60. Casanova J-L, Abel L. Human genetics of infectious diseases: unique insights into immunological redundancy. *Semin Immunol*. (2018) 36:1–12. doi: 10.1016/j.smim.2017.12.008

61. Lynch NJ, Roscher S, Hartung T, Morath S, Matsushita M, Maennel DN, et al. L-ficolin specifically binds to lipoteichoic acid, a cell wall constituent of gram-positive bacteria, and activates the lectin pathway of complement. *J Immunol*. (2004) 172:1198–202. doi: 10.4049/jimmunol.172.2.1198

62. Roos A, Garred P, Wildenberg ME, Lynch NJ, Munoz JR, Zuiverloon TCM, et al. Antibody-mediated activation of the classical pathway of complement may compensate for mannose-binding lectin deficiency. *Eur J Immunol*. (2004) 34:2589–98. doi: 10.1002/eji.200324401

63. Mills TC, Chapman S, Hutton P, Gordon AC, Bion J, Chiche J-D, et al. Variants in the mannose-binding lectin gene MBL2 do not associate with Sepsis susceptibility or survival in a large European cohort. *Clin Infect Dis*. (2015) 61:695–703. doi: 10.1093/cid/civ378

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 16 April 2024
DOI 10.3389/fmed.2024.1366691

OPEN ACCESS

EDITED BY
J. Francis Borgio,
Imam Abdulrahman Bin Faisal University, Saudi Arabia

REVIEWED BY
Hanqing Liu,
Renmin Hospital of Wuhan University, China
Jeehee Yoon,
Chonnam National University Bitgoeul Hospital, Republic of Korea
Anoop Kumar,
Delhi Pharmaceutical Sciences and Research University, India
Weimin Zhong,
Xiamen Fifth Hospital, China

*CORRESPONDENCE
Jing Wang
✉ 1040110291@qq.com
Li Chen
✉ chenl_hxey@scu.edu.cn

RECEIVED 07 January 2024
ACCEPTED 07 February 2024
PUBLISHED 16 April 2024

CITATION
Shi H, He Y, Dan S, Yang L, Wang J, Chen L and Chen Z (2024) Endocrine system-related adverse events associated with PD-1/PD-L1 inhibitors: data mining from the FDA adverse event reporting system. *Front. Med.* 11:1366691. doi: 10.3389/fmed.2024.1366691

Endocrine system-related adverse events associated with PD-1/PD-L1 inhibitors: data mining from the FDA adverse event reporting system

Hongxia Shi[1], Yunhua He[1], Siyuan Dan[1], Lin Yang[1], Jing Wang[2]*, Li Chen[3,4]* and Zelian Chen[1]

[1]Department of Pharmacy, West China Hospital, Sichuan University, Chengdu, China, [2]Department of Pharmacy, Sichuan Mianyang 404 Hospital, Mianyang, China, [3]Department of Pharmacy/Evidence-Based Pharmacy Center, West China Second University Hospital, Sichuan University, Chengdu, China, [4]Department of Pharmacology, Faculty of Medicine, University of the Basque Country UPV/EHU, Leioa, Spain

Background: Various immune checkpoint inhibitors, such as programmed cell death protein-1 (PD-1) and its ligand (PD-L1), have been approved for use, but they have side effects on the endocrine glands.

Methods: Adverse event reports related to PD-1/PD-L1 inhibitors from the FDA Adverse Event Reporting System (FAERS) from the first quarter of 2019 to the first quarter of 2023 were extracted, and the reported Odds ratio methods (ROR method) and comprehensive standard methods (MHRA methods) were used for data mining and analysis.

Results: A total of 5,322 reports (accounts for 6.68% of the total reports)of AEs in endocrine system were collected, including 1852 of pabolizumab (34.80%), 2,326 of navuliumab (43.71%), 54 of cimipriliumab (1.01%), 800 of atilizumab (15.03%), 222 of duvariumab (4.17%) and 68 of averumab (1.28%). Endocrine system-related AEs were mainly present in men (excluding those treated with pembrolizumab) aged ≥65 years. The ratio of AEs components in the endocrine system for the six drugs was approximately 3–8%. The main endocrine glands involved in AEs were the thyroid (pembrolizumab), pituitary and adrenal (nivolumab), adrenal (cemiplimab, atezolizumab, and avelumab), and thyroid (durvalumab). Most patients experienced AEs between 30 and 365 (mean, 117) days,the median time was 61d. AEs resulted in prolonged hospitalization in >40% and death in >10% of cases after administration of pembrolizumab, nivolumab, or durvalumab.

Conclusion: Men aged ≥65 years should be concerned about endocrine-related AEs. There was a lengthy interval between the use of PD-1/PD-L1 inhibitors and endocrine system-related AEs, but the outcome was serious. Special attention should be given to endocrine system-related AEs when using pembrolizumab, nivolumab, or durvalumab.

KEYWORDS

PD-1/PD-L1 inhibitor, endocrine system, adverse events, FAERS, data mining

1 Introduction

The programmed death-1 and programmed cell death-ligand 1 (PD-1/PD-L1) were major immune checkpoint inhibitors. The PD-1/PD-L1 derived drugs were specifically recognizing and blocking immunosuppressive molecules to achieve anti-tumor response, namely enhancing anti-tumor immune response, inhibiting immune evasion, inducing tumor cell death, It's called immunotherapy for tumors (1–3). The immunotherapy is another important therapy strategy after surgery, chemotherapy and radiotherapy, which has been widely applied in the treatment of melanoma, lung, lymphoma, kidney, endometrial and other tumors (2, 4–6). However, PD-1/PD-L1 inhibitor will enhance over-activated immune cells response to normal cells, resulting in immune-related adverse events (irAEs) in organs or tissues.

Therefore, while benefiting from treatment, cancer patients will also be troubled by irAEs, such as gastrointestinal toxicity, skin toxicity, endocrine toxicity, immune-associated pneumonia, etc. (2), involving multiple systems. Treatment can trigger autoimmune reactions in various ways (e.g., increasing the level of autoantibodies (1)) and then involve multiple glands (e.g., pituitary, thyroid, and adrenal) to cause the corresponding functional disorders. Recent studies have shown that endocrine toxicity is irreversible in 50% of cases (7) and can be life-threatening if not identified and treated appropriately (8–10). The disproportional reporting is most usually employed in adverse drug events signal monitoring, which containing reporting odds ratio (ROR), comprehensive standard (MHRA) and proportional reporting ratio (PRR) methods. Up to date, there are numerous studies have been applied these methods in drug safety investigation (11–13).

Therefore, the proportional imbalance method was adopted in this study, we conducted data mining through the US Food and Drug Administration (FDA) Adverse Event Reporting System (FAERS) database. We focused on AEs reported after the use of PD-1/PD-L1 inhibitors in the endocrine system. We concentrated on the risks and characteristics of AEs caused by these drugs and provided references for further prevention and management.

2 Methods

2.1 Data sources

FAERS is an open database of anonymous patient health and treatment information that contains information on adverse event and medication error reports submitted to FDA. We used data from the FAERS database. The present study did not involve therapeutic interventions or the collection of human samples and, as such, was exempt from approval from an institutional review board approval. There are seven tables: patient demographic and administrative information, medication information, adverse drug reaction information, information on reporting sources, start and end dates of drug therapy, indications for use/diagnosis, and case deletions.

There are numerous types of PD-1/PD-L1 inhibitors. We included only single-agent preparations and excluded varieties available only on the market in China. As a result, we included six drugs for analyses. Pembrolizumab is a PD-1 inhibitor. Nivolumab is a PD-1 inhibitor. Cemiplimab is a PD-L1 inhibitor of atezolizumab, durvalumab, and avelumab.

2.2 Data processing

Cemiplimab was launched in the USA relatively recently (September 2018) compared with the other drugs. Hence, the time range of data extraction in the present study was the first quarter of 2019 to the first quarter of 2023 (17 quarters in total). The search terms (including the generic name and trade name of the drug) we used were "pembrolizumab/Keytruda," "nivolumab/opdivo/opdualag," "cemiplimab/libtayo," "atezolizumab/tecentriq," "durvalumab/imfinzi," and "avelumab/bavencio." The FAERS database is updated each quarter, so published reports will inevitably be duplicated. Hence, reprocessing was done using MySQL 8.0,[1] as recommended by the FDA. If the CASEID and FDA_DT were identical, then the latest PRIMARYID was selected. If the CASEID and FDA_DT were identical, then the DELETE report was selected from the DELETE table. The FAERS database is encoded using the Medical Dictionary for Regulatory Activities (MedDRA) of the International Council for Organizations of Medical Sciences. Therefore, systematic organ classification (SOC) and the preferred term (PT) in the latest edition of the MedDRA Glossary of Adverse Drug Reactions (MedDRA 25.1) were used in the present study. MedDRA 25.1 standardizes and minimizes international terminology of terms to describe AEs (14, 15). According to PUBMED age grouping standard (16), the entire population was categorized into three distinct age groups: minors (age < 18 years), adults (18 ≤ age < 65 years), and senior citizens (≥65 years).

2.3 Data analyses

Data (e.g., target number of AEs reports and background number of AEs occurrences of the primary suspected drug) were obtained. Potential AE signals were screened based on a four-cell table (14, 17) of the proportional imbalance method (See Supplementary Table S1). Adopt the ROR and MHRA method, the ROR, proportional reporting ratio (PRR), and X^2 equivalents are calculated, respectively. To avoid false-positive signals, the calculated corresponding values should reach the set threshold and be defined as the PT of valid signals (See Supplementary Table S2) (18–20). The larger the calculated value, the stronger the signal, indicating that the target drug is more likely to be associated with the target AEs, but this does not mean that there is a causal relationship between the two (21). All statistical analyses were undertaken using Prism 8 (GraphPad, La Jolla, CA, United States), SPSS29 and Excel® (Microsoft, Redmond, WA, United States).

3 Results

3.1 Primary characteristics of AEs reported in the endocrine system

As of the first quarter of 2023, the FAERS database collected 79,700 AEs reports. Among them, there were 22,918, 34,267,

1 https://dev.mysql.com/doc/refman/8.2/en/

1,239, 13,862, 6,014, and 1,300 AEs reports involving pembrolizumab, nivolumab, cemiplimab, atezolizumab, durvalumab, and avelumab as the primary suspected drug, respectively. Among them, there were 5,322 reports related to the endocrine system (accounting for 6.68% of the total reports), with the aforementioned six drugs accounting for 34.80, 43.71, 1.01, 15.03, 4.17 and 1.28%, respectively. Patients who suffered PD-1/PD-L1 inhibitor-associated AEs were predominantly male (except for those who had AEs after taking pembrolizumab) and most of the population is over aged >65 years. The occupations at the top of the list were physician and consumers. The countries with the largest number of AEs reports by drug were mostly Japan. The specific number and proportion of reports are shown in Table 1.

3.2 Proportion of AEs in the endocrine system

After SOC classification of excavated effective signals, we found no significant difference in the ratio of AEs components associated with PD-1/PD-L1 inhibitors in the endocrine system (approximately 3–8%). The specific numbers of cases and composition ratios are shown in Figure 1.

3.3 AEs signals and correlation with the endocrine system

A total of 131 AEs signals were detected in the endocrine system: 37 for pembrolizumab, 43 for nivolumab, six for cemiplimab, 25 for atezolizumab, 13 for durvalumab, and seven for avelumab. Endocrine system-related AEs had a strong correlation with pembrolizumab, including immune-mediated hypothyroidism ($X^2 = 27{,}216.81$, ROR = 312.66), adrenocorticotropin deficiency (10,269.46, 105.25), and hypothyroidism (6307.58, 14.73). AEs closely associated with nivolumab were pituitary inflammation ($X^2 = 14{,}605.84$, ROR = 82.57), adrenal insufficiency (6549.80, 19.03), and hypopituitarism (5874.78, 48.99). Adrenal dysfunction was the main factor in AEs attributed to a strong correlation between the use of cemiplimab and atezolizumab ($X^2 = 497.78$, ROR = 25.10; $X^2 = 4582.73$, ROR = 23.79). The AEs with a strong correlation with durvalumab use was silent thyroiditis ($X^2 = 508.62$, ROR = 94.22). The AEs with strong association of avilumab was adrenal disease (X^2: 252.79, ROR: 44.68). See Figure 2 for details.

TABLE 1 Clinical characteristics of reported cases of AEs in endocrine system (*n*, %).

Characteristic	Pembrolizumab	Nivolumab	Cemiplimab	Atezolizumab	Durvalumab	Avelumab
Number of cases	1852	2,326	54	800	222	68
Sex						
Male	860 (46.44)	1,425 (61.26)	15 (27.78)	383 (47.88)	130 (58.56)	41 (60.29)
Female	940 (50.76)	778 (33.45)	3 (5.56)	265 (33.13)	76 (34.23)	26 (38.24)
Unknown	52 (2.81)	123 (5.29)	36 (66.67)	152 (19.00)	16 (7.21)	1 (1.47)
P	0.009	<0.001	0.002	<0.001	<0.001	0.010
Age						
<18 years	2 (0.11)a	5 (0.21)a	0 (0.00)b	1 (0.13)a	0 (0.00)a	0 (0.00)a
18 ≤ years <65	584 (31.53)a	929 (39.94)a	3 (5.56)b	247 (30.88)a	80 (36.04)b	27 (39.71)b
18 ≤ years <65	939 (50.70)a	1,125 (48.37)a	10 (18.52)a	369 (46.13)a	103 (46.40)a	35 (51.47)a
Unknown	327 (17.66)	267 (11.48)	41 (75.93)	183 (22.88)	39 (17.57)	6 (8.82)
p	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Reporter occupation						
Physician	1,128 (60.91)a	1,492 (64.14)a	44 (81.48)a	665 (83.13)a	158 (71.17)a	63 (92.65)a
Consumer	518 (27.97)a	250 (10.75)a	6 (11.11) a	18 (2.25)a	26 (11.71)a	5 (7.35) b
Other	4 (0.22)a	2 (0.09%)a	0 (0.00)a	0 (0.00)a	0 (0.00)a	0 (0.00)a
Unknown	202 (10.91)	582 (25.02)	4 (7.41)	117 (14.63)	38 (17.12)	0 (0.00)
p	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Reporting countries (top 5)	Japan (897, 48.43)	Japan (824, 35.43)	USA (19, 35.19)	Japan (381, 47.63)	Japan (104, 46.85)	Japan (25, 36.76)
	USA (452, 24.41)	USA (535, 23.00)	Australia (8,14.81)	USA (94, 12.13)	USA (48, 21.62)	France (9, 13.24)
	France (78, 4.21)	Germany (233, 10.02)	France (7, 12.96)	France (43, 5.38)	France (12, 5.38)	USA (9, 13.24)
	Germany (45, 2.43)	France (228, 9.80)	Korea (5, 9.26)	China (32, 4.00)	Canada (11, 4.95)	Finland (3, 4.41)
	Italy (32, 1.73)	Canada (48, 2.06)	Japan (4, 7.41)	Germany (27, 3.38)	China (10, 4.50)	Britain (3, 4.41)

The categorical variables were tested by X^2, and $p < 0.05$ was statistically significant. When the three groups of data were pairwise compared, $p < 0.0167$, there was a statistical difference, $a < 0.001$, $b > 0.0167$. Unknown groups are not compared.

3.4 Time of occurrence of AEs in the endocrine system

The onset time of AEs in the endocrine system was more distributed between 30–365 days, the median time was 61d, the median onset time of AEs in the endocrine system was 42d for pembrolizumab, 63 days for nivolumab, 161 days for cemiplimab, 73.5 days for atezolizumab, and 42 days for durvalumab. Avelumab was 56 days. Time of adverse endocrine system events with the use of six drugs. See Figure 3 for details.

3.5 PD-1/PD-L1 inhibitors produce AEs in the endocrine system

In addition to unknown serious medical events, the most prevalent outcome of endocrine system AEs due to the use of PD-1/PD-L1 inhibitors was prolonged hospital stay (44.18–57.78%). The other most prevalent outcomes were death, life-threatening injury, and disability. Death due to taking pembrolizumab, nivolumab, or durvalumab accounted for >10% of cases. See Figures 3, 4 for details.

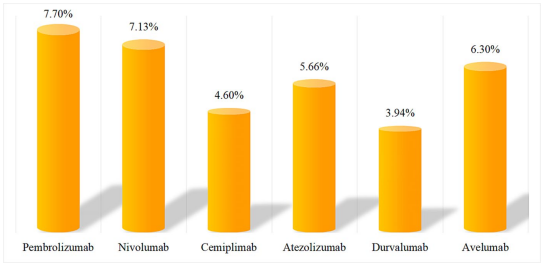

FIGURE 1
Component ratio of AEs in endocrine system.

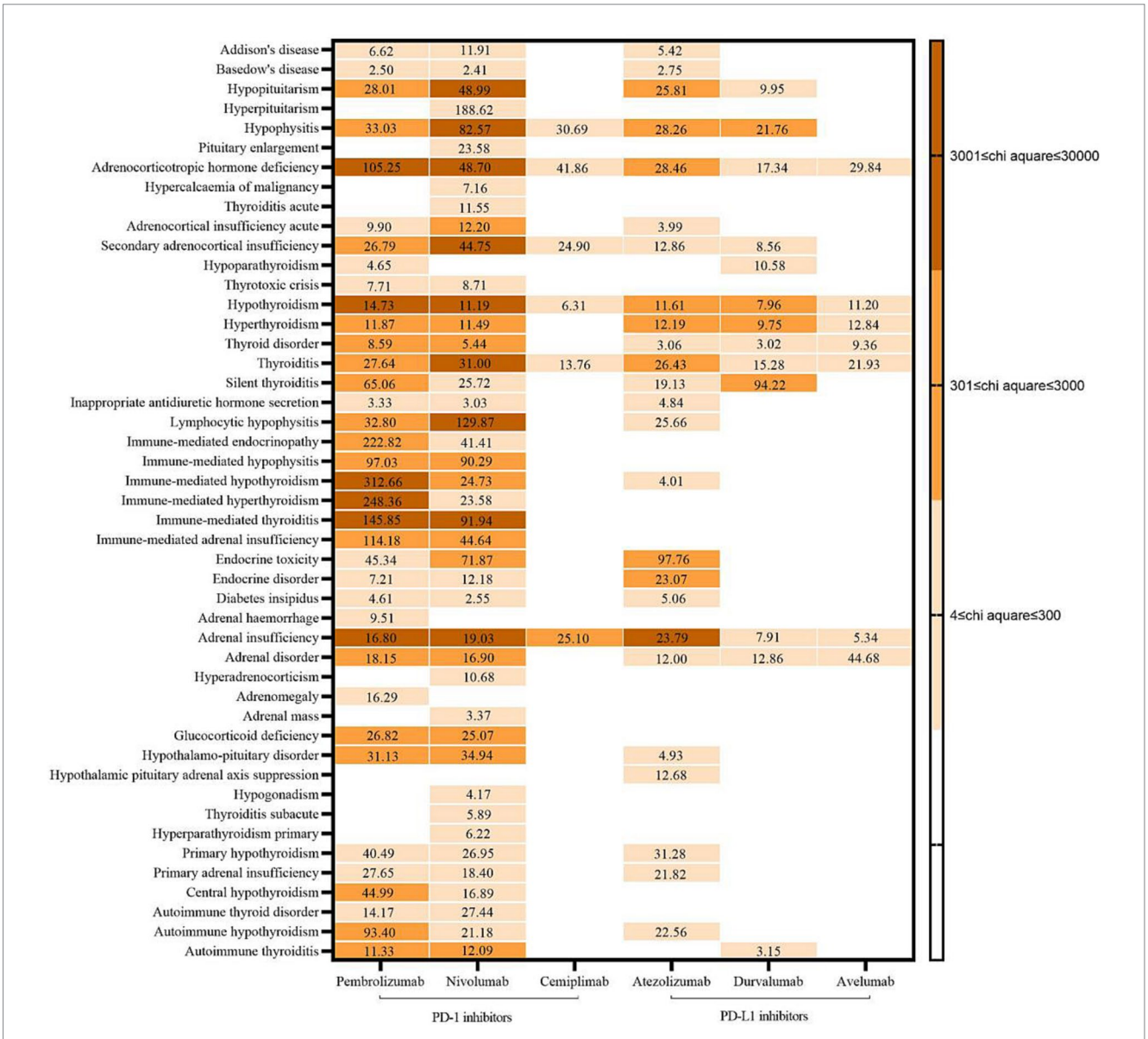

FIGURE 2
Signals and correlation of adverse events in endocrine system.

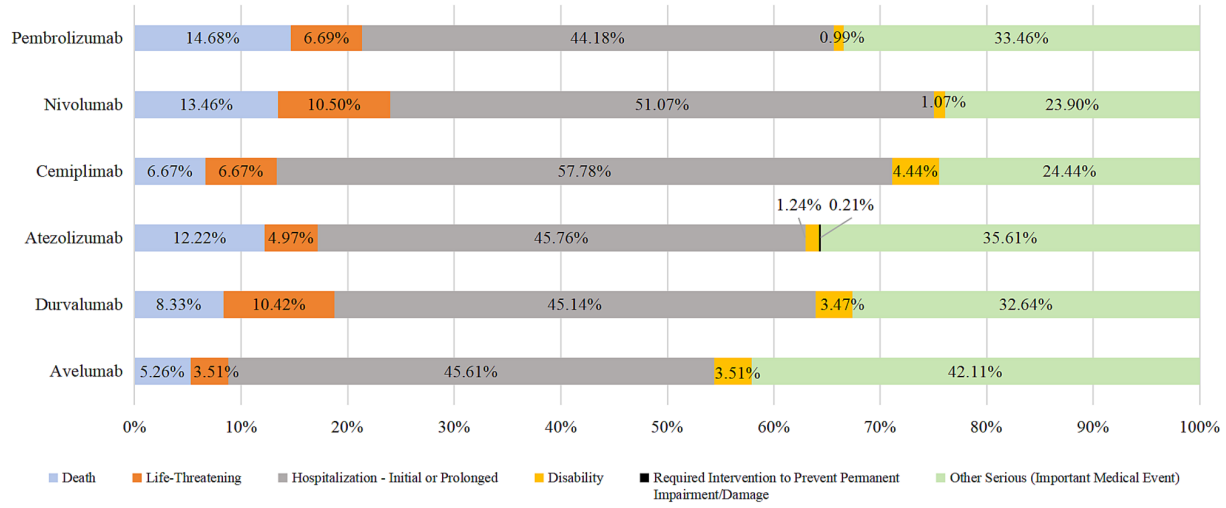

FIGURE 3
Time distribution of AEs in endocrine system.

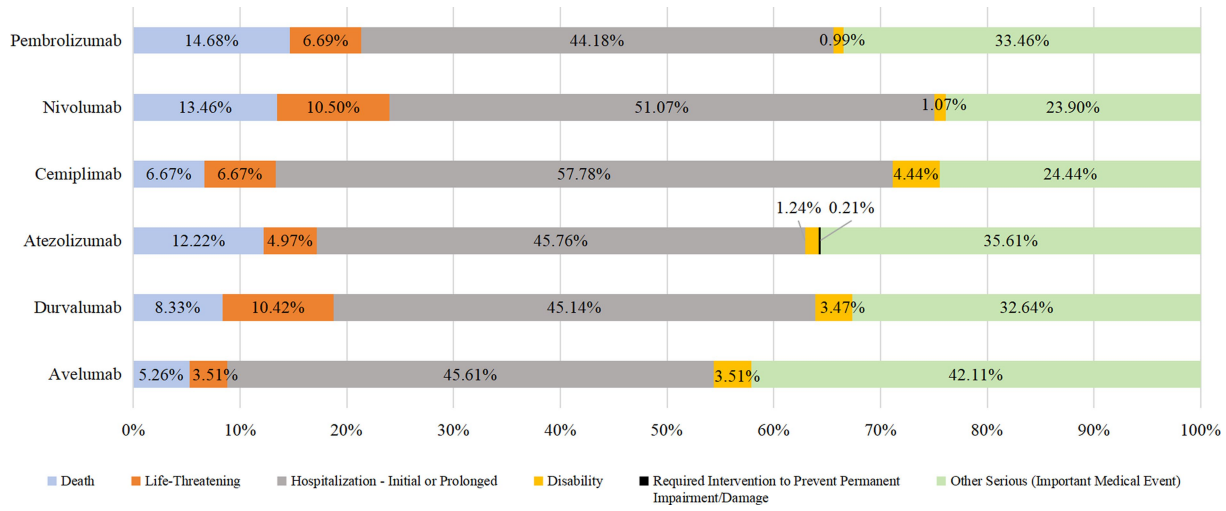

FIGURE 4
Outcome distribution of AEs in endocrine system.

4 Discussion

4.1 Basic characteristics of reported cases of PD-1/PD-L1 inhibitor-associated endocrine system AEs

We found that, except for pembrolizumab, five PD-1/PD-L1 inhibitors associated with AEs occurred mainly in men, a result which is consistent with those of other studies (22, 23). Our finding is probably related to the completeness of the data. Compared with the data for the other five drugs, the data size for pembrolizumab was relatively large. The proportion of cases in which the sex of the patient was not known was relatively low (2.81%). Results can vary if data on sex are missing. All six PD-1/PD-L1 inhibitors were used for patients aged >65 years, which may have been related to the age when the disease was diagnosed (24, 25), however, elderly men over 65 years of age should be especially aware of the occurrence of relevant AEs. Most of the reporters are medical personnel, indicating that the AEs reported to the database has strong reliability. The countries that reported the most AEs were Japan and the USA. This finding suggests that other countries may be paying less attention to AEs in the endocrine system, but this may also be related to the number of people taking these drugs in such countries.

4.2 AEs distribution and correlation in endocrine system associated PD-1/PD-L1 inhibitors

The findings of this study, the number of AEs signals associated with the endocrine system was higher for PD-1 inhibitors than for PD-L1 inhibitors and the degree of correlation was also larger,

especially pembrolizumab and nivolumab. The above findings are similar to the results of previous relevant studies (26), and this kind of ADR should be paid attention to when using related drugs. The results also suggest that, which should be considered (especially for pembrolizumab and nivolumab). The main organs involved in AEs associated with pembrolizumab use were the thyroid gland and adrenal glands, whereas they were the pituitary gland and adrenal glands for nivolumab, the adrenal glands for cemiplimab, the adrenal glands for atezolizumab, and the thyroid gland for avelumab. According to the *Clinical Application Guide of Immune Checkpoint Inhibitors for Gynecological Tumors* (2), the common types of endocrine toxicity caused by ICIs are dysfunction in the thyroid gland and acute pituitary inflammation, these data are consistent with the results of our study. As mentioned above, endocrine glands contain rich blood flow and may trigger autoimmune reactions through various ways such as the activation of autoimmune cells, thus involving multiple glands (1). However, the mechanism of adverse reactions in different glands may be different. In addition to immune-mediated mechanisms (3), thyroid injury may also be associated with the upregulation of PD-1 receptors in the thyroid gland (1). Furthermore, elevated levels of serum IL-1β, IL-2, and GM-CSF at baseline, as well as decreased levels of serum IL-8, G-CSF, and MCP-1 at an early stage are correlated with thyroid dysfunction (27, 28). Pituitary is also associated with humoral immunity, usually involving the anterior pituitary, which results in permanent dysfunction of one or more pituitary endocrine axes (29, 30). Combined with the results of our study, the use of PD-1/PD-L1 inhibitors seems to result in additional abnormalities in the functions of the thyroid gland and pituitary gland. These data suggest that monitoring abnormalities in the function of these two organs is important. In the event of a serious acute reaction, immunotherapy should be stopped promptly, and the corresponding drug treatment and symptomatic treatment should be applied (1).

In addition, the above guidelines mentioned that adrenal diseases were rare endocrine toxicity, but the results of this study found that the use of PD-1/PD-L1 inhibitors had a large association with adrenal diseases, which is worth noting. The mechanism of this glandular disease may be related to the infiltration of CD4 + t cells, especially Th1 and Th17 cells (3). Therefore, in clinical practice, medical personnel should still pay attention to the occurrence of such AEs. If patients have abnormal adrenal function and indicators, physicians and clinical pharmacists should judge whether it is caused by such drugs according to the baseline assessment, so as to correctly handle adverse drug reaction (ADR), it also provides reference for whether to adjust the anti-tumor therapy regimen in the future.

4.3 Occurrence time and outcome of AEs in endocrine system associated with PD-1/PD-L1 inhibitors

Our data suggested that AEs in the endocrine system occurred between 30 and 365 (mean, 117) days after the use of PD-1/PD-L1 inhibitors, the median time was 61d, which was consistent with Viola suggestion (31). The above results indicate that this type of AEs occurs slowly, and long-term follow-up monitoring is needed for patients using this type of drugs.

Hospitalization/prolonged hospitalization was the most prevalent outcome of AEs, followed by death, life-threatening illness, and disability. According to the Common Terminology Criteria for Adverse Events (CTCAE), patients who reach grade 3 or above will be hospitalized for intravenous hormone therapy (2), especially if there is a corresponding emergency/crisis, Additional treatments, such as anti-infection therapy, blood purification, and ventilatory support, are needed (1), if the treatment is not timely and incorrect, it can endanger life or even death (33), indicating the severity of the AEs, it is necessary to do a baseline assessment before the use of this type of drug in clinical practice, and close monitoring during use to identify ADR as early as possible and timely intervention to reduce or even avoid the occurrence of adverse outcomes. It is worth noting that the death outcome of pembrolizumab, nivolumab and durvalumab accounted for more than 10%. When patients use the above three drugs, clinical pharmacists should set them as key monitoring objects, closely monitor the corresponding indicators and changes in symptoms and signs, pay attention to the suitability of medication, and cooperate with the clinic to improve the prognosis.

5 Research limitations

First, the FAERS database has a large amount of data, there are also a lot of missing data information such as gender, age and adverse event occurrence time, especially the time is not accurate enough. Secondly, the accuracy and professionalism of the information of "adverse events" in the database need to be improved. Reporters from different occupational backgrounds may use different descriptions of endocrine toxicity and endocrine disorders, which may lead to deviations in the equivalence calculation of ROR and PRR for a single PT. Despite these limitations, spontaneous reporting may be the best way to collect more AEs that might otherwise be overlooked (34). Third, the AEs signal detected in this study indicates that the drug is statistically correlated with the AEs, but it does not mean that there is a causal link in biology, and further clinical trials are needed to explore (35, 36). In addition, sensitivity analysis cannot be performed in the current proportion imbalance method, and the impact of combined drug use on the outcome is difficult to predict,other research methods can be explored to evaluate the impact in the future. Fourth, due to a lack of information on the total number of people used, the incidence of specific adverse events cannot be calculated (37), therefore, the intensity of the association between drugs and adverse events was measured.

6 Conclusion

Men aged ≥65 years should be concerned about endocrine-related AEs. The use of different PD-1/PD-L1 inhibitors mainly involves interaction with the endocrine glands, so physicians should be careful when prescribing drugs for patients with associated underlying diseases. There was a lengthy interval between the use of PD-1/PD-L1 inhibitors and endocrine system-related AEs, but the outcome was serious. Therefore, long-term, meticulous monitoring and appropriate treatment are necessary. Special attention should be given to endocrine system-related AEs when using pembrolizumab, nivolumab, or durvalumab.

Data availability statement

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

Ethics statement

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

Author contributions

HS: Conceptualization, Investigation, Validation, Writing – original draft. YH: Visualization, Writing – original draft. SD: Data curation, Writing – original draft. LY: Data curation, Writing – original draft. JW: Data curation, Formal analysis, Writing – review & editing. LC: Methodology, Supervision, Writing – review & editing. ZC: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was financially supported by the National Natural Science Foundation of China (82073921) and the National Key Clinical Specialties Construction Program.

Acknowledgments

We sincerely thank everyone who participated in this study.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2024.1366691/full#supplementary-material

References

1. Yongzhong W, Qilan W, Danlan P, Danlan P, Bing C, Danbo B, et al. Consensus of Chinese experts on emergency management of major endocrine adverse reactions of immune checkpoint inhibitors. *J Chongqing Med Univ*. (2023) 48:1–12. doi: 10.13406/j.cnki.cyxb.002758 (in Chinese)

2. Beihua K, Jihong L, Aijun Y, Yun Z, Xiaoping L, Qinglei G, et al. Clinical application guide of immunological checkpoint inhibitors for gynecological tumors. *J Integr Cancer Ther Electr*. (2023) 9:67–98. doi: 10.13283/j.cnki.xdfckjz.2023.05.001 (in Chinese)

3. Okiyama N, Tanaka R. Immune-related adverse events in various organs caused by immune checkpoint inhibitors. *Allergol Int*. (2022) 71:169–78. doi: 10.1016/j.alit.2022.01.001

4. Xiaoli L, Zhenyan LI, Lili W. Progress and challenge of radiotherapy combined with immunotherapy for esophageal cancer. [J]. *J Oncol* (2023) 29:1–7. (in Chinese)

5. Dongli Z, Shen C, Weichuan Z, et al. Efficacy and safety of programmed death factor-1 / programmed death factor-1 ligand inhibitors in treatment of renal cell carcinoma. *Chin Gen Pract* (2023) 1–9. (in Chinese)

6. Chenxi Z, Yi X, Yingying H, Dandan L, Yawei L, Wei Z, et al. Effect of smoking on the efficacy of immune checkpoint inhibitors in patients with driver-negative non-small cell lung cancer. *Chin J Clin Pharmacol*. (2023) 39:1397–401. doi: 10.13699/j.cnki.1001-6821.2023.10.007. (in Chinese)

7. Ferrari SM, Fallahi P, Elia G, et al. Autoimmune endocrine dysfunctions associated with cancer immunotherapies. *Int J Mol Sci*. (2019) 20:2560. doi: 10.3390/ijms20102560

8. Torino F, Barnabei A, Paragliola RM, Marchetti P, Salvatori R, Corsello SM. Endocrine side-effects of anti-cancer drugs: mAbs and pituitary dysfunction: clinical evidence and pathogenic hypotheses. *Eur J Endocrinol*. (2013) 169:R153–64. doi: 10.1530/EJE-13-0434

9. Hughes J, Vudattu N, Sznol M, Gettinger S, Kluger H, Lupsa B, et al. Precipitation of autoimmune diabetes with anti-PD-1 immunotherapy [J]. *Diabetes Care*. (2015) 38:e55–7. doi: 10.2337/dc14-2349

10. Orlov S, Salari F, Kashat L, Walfish PG. Induction of painless thyroiditis in patients receiving programmed death 1 receptor immunotherapy for metastatic malignancies. *J Clin Endocrinol Metab*. (2015) 100:1738–41. doi: 10.1210/jc.2014-4560

11. Jain D, Sharma G, Kumar A. Adverse effects of proton pump inhibitors (PPIs) on the renal system using data mining algorithms (DMAs). *Expert Opin Drug Saf*. (2023) 22:741–52. doi: 10.1080/14740338.2023.2189698

12. Faiza J, Anoop K. Identification of signal of clindamycin associated renal failure acute: a disproportionality analysis. *Curr Drug Saf*. (2023) 19:123–8. doi: 10.2174/1574886318666230228142856

13. Shu Y, He X, Liu Y, Wu P, Zhang Q. A real-world disproportionality analysis of Olaparib: data Mining of the Public Version of FDA adverse event reporting system. *Clin Epidemiol*. (2022) 14:789–802. doi: 10.2147/CLEP.S365513

14. Sakaeda T, Tamon A, Kadoyama K, Okuno Y. Data mining of the public version of the FDA adverse event reporting system. *Int J Med Sci*. (2013) 10:796–803. doi: 10.7150/ijms.6048

15. Omar NE, Fahmy Soliman AI, Eshra M, Saeed T, Hamad A, Abou-Ali A. Postmarketing safety of anaplastic lymphoma kinase (ALK) inhibitors: an analysis of the FDA adverse event reporting system (FAERS). *ESMO Open*. (2021) 6:100315. doi: 10.1016/j.esmoop.2021.100315

16. National Library of Medicine. *Age Groups*. [EB/OL] (1998–2024–01–30]. Available at: https://www.ncbi.nlm.nih.gov/mesh/68009273

17. Lin L, Jiaying Z, Li C, Enwei L, et al. Adverse event signal mining of tocilizumab based on FAERS database in the United States. *Chin J Pharm*. (2021) 32:1874–9. (in Chinese)

18. Chen JJ, Huo XC, Wang SX, Wang F, Zhao Q. Data mining for adverse drug reaction signals of daptomycin based on real-world data: a disproportionality analysis of the US Food and Drug Administration adverse event reporting system [J]. *Int J Clin Pharm*. (2022) 44:1351–60. doi: 10.1007/s11096-022-01472-x

19. Sakaeda T, Kadoyama K, Okuno Y. Adverse event profiles of platinum agents: data mining of the public version of the FDA adverse event reporting system, AERS, and reproducibility of clinical observations. *Int J Med Sci*. (2011) 8:487–91. doi: 10.7150/ijms.8.487

20. Ziyan J, Chen C, Yufan D, Bing W, Ting X, et al. Study on signal mining of benzodiazepine-related dementia events based on FAERS. *Herald Med*. (2021) 40:1356–60. (in Chinese)

21. Zhou Y, Chen M, Liu L, Chen Z. Difference in gastrointestinal risk associated with use of GLP-1 receptor agonists: a real-world pharmacovigilance study. *Diabetes Metab Syndr Obes*. (2022) 15:155–63. doi: 10.2147/DMSO.S348025

22. Ling L, Anqi L, Junxian Y. Data analysis of pneumonia induced by PD-1/PD-L1 inhibitors based on FAERS. *Chin J Hosp Pharm*. (2021) 41:1288–92. doi: 10.13286/j.1001-5213.2021.13.02. (in Chinese)

23. Bai S, Tian T, Pacheco JM, Tachihara M, Hu P, Zhang J. Immune-related adverse event profile of combination treatment of PD-(L)1 checkpoint inhibitors and bevacizumab in non-small cell lung cancer patients: data from the FDA adverse event reporting system. *Transl Lung Cancer Res*. (2021) 10:2614–24. doi: 10.21037/tlcr-21-464

24. National Health Commission of the People's Republic of China Medical Administration. Guidelines for diagnosis and treatment of esophageal cancer. *Chin J Dig Surg*. (2022) 21:1247–68. (in Chinese)

25. National Comprehensive Cancer Network. *Hepatocellular Carcinoma (Version 1.2023)*. [EB/OL]. (2023-09-14)[2024-01-30]. Available at: https://www.nccnchina.org.cn/guide/detail/443

26. Sonpavde GP, Petros G, Yushun L. Immune-related adverse events with PD-1 versus PD-L1 inhibitors: a meta-analysis of 8730 patients from clinical trials [J]. *Future Oncol*. (2021):2545–58.

27. Kurimoto C, Inaba H, Ariyasu H, Iwakura H, Ueda Y, Uraki S, et al. Predictive and sensitive biomarkers for thyroid dysfunctions during treatment with immune-checkpoint inhibitors. *Cancer Sci*. (2020) 111:1468–77. doi: 10.1111/cas.14363

28. Chennamadhavuni A, Abushahin L, Jin N, Presley CJ, Manne A. Risk factors and biomarkers for immune-related adverse events: a practical guide to identifying high-risk patients and Rechallenging immune checkpoint inhibitors&13. *Front Immunol*. (2022) 13:779691. doi: 10.3389/fimmu.2022.779691

29. Chang LS, Barroso-Sousa R, Tolaney SM, Hodi FS, Kaiser UB, Min L. Endocrine toxicity of cancer immunotherapy targeting immune checkpoints. *Endocr Rev*. (2019) 40:17–65. doi: 10.1210/er.2018-00006

30. Scott ES, Long GV, Guminski A, Clifton-Bligh RJ, Menzies AM, Tsang VH. The spectrum, incidence, kinetics and management of endocrinopathies with immune checkpoint inhibitors for metastatic melanoma. *Eur J Endocrinol*. (2018) 178:173–80. doi: 10.1530/EJE-17-0810

31. Trevisani V, Iughetti L, Lucaccioni L, Predieri B. Endocrine immune-related adverse effects of immune-checkpoint inhibitors. *Expert Rev Endocrinol Metab*. (2023) 18:441–51. doi: 10.1080/17446651.2023.2256841

32. de Filette J, Jansen Y, Schreuer M, Everaert H, Velkeniers B, Neyns B, et al. Incidence of thyroid-related adverse events in melanoma patients treated with Pembrolizumab. *J Clin Endocrinol Metab*. (2016) 101:4431–9. doi: 10.1210/jc.2016-2300

33. Iglesias P. Cancer immunotherapy-induced endocrinopathies: clinical behavior and therapeutic approach. *Eur J Intern Med*. (2018) 47:6–13. doi: 10.1016/j.ejim.2017.08.019

34. Yang Z, Lv Y, Yu M, Mei M, Xiang L, Zhao S, et al. GLP-1 receptor agonist-associated tumor adverse events: a real-world study from 2004 to 2021 based on FAERS. *Front Pharmacol*. (2022) 13:925377. doi: 10.3389/fphar.2022.925377

35. Jiaying Z, Xxiangdan C, Li C, et al. Application of ratio imbalance method in signal mining of adverse reactions of nivolumab. *Oncol Pharm*. (2019) 9:798–803. (in Chinese)

36. Xiaowen H, Liu L, Yiwen R, et al. Adverse event signal mining of platinum drugs based on FAERS. *Oncol Pharm*. (2020) 10:608–16. (in Chinese)

37. Ahdi HS, Wichelmann TA, Pandravada S, Ehrenpreis ED. Medication-induced osteonecrosis of the jaw: a review of cases from the Food and Drug Administration adverse event reporting system (FAERS) [J]. *BMC Pharmacol Toxicol*. (2023) 24:15.

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 13 June 2024
DOI 10.3389/fmed.2024.1381031

OPEN ACCESS

EDITED BY
Miriana d'Alessandro,
University of Siena, Italy

REVIEWED BY
Ourania Papaioannou,
General University Hospital of Patras, Greece
Jaya Talreja,
Wayne State University, United States

*CORRESPONDENCE
Peter J. van der Spek
✉ p.vanderspek@erasmusmc.nl

RECEIVED 02 February 2024
ACCEPTED 23 May 2024
PUBLISHED 13 June 2024

CITATION
van Wijck RTA, Sharma HS, Swagemakers SMA, Dik WA, IJspeert H, Dalm VASH, van Daele PLA, van Hagen PM and van der Spek PJ (2024) Bioinformatic meta-analysis reveals novel differentially expressed genes and pathways in sarcoidosis.
Front. Med. 11:1381031.
doi: 10.3389/fmed.2024.1381031

Bioinformatic meta-analysis reveals novel differentially expressed genes and pathways in sarcoidosis

Rogier T. A. van Wijck[1], Hari S. Sharma[1], Sigrid M. A. Swagemakers[1], Willem A. Dik[2], Hanna IJspeert[2], Virgil A. S. H. Dalm[2,3], Paul L. A. van Daele[2,3], P. Martin van Hagen[2,3] and Peter J. van der Spek[1]*

[1]Department of Pathology & Clinical Bioinformatics, Erasmus MC University Medical Center, Rotterdam, Netherlands, [2]Laboratory Medical Immunology, Department of Immunology, Erasmus MC University Medical Center, Rotterdam, Netherlands, [3]Department of Internal Medicine, Division of Allergy & Clinical Immunology, Erasmus MC University Medical Center, Rotterdam, Netherlands

Introduction: Sarcoidosis is a multi-system inflammatory disease of unknown origin with heterogeneous clinical manifestations varying from a single organ non-caseating granuloma site to chronic systemic inflammation and fibrosis. Gene expression studies have suggested several genes and pathways implicated in the pathogenesis of sarcoidosis, however, due to differences in study design and variable statistical approaches, results were frequently not reproducible or concordant. Therefore, meta-analysis of sarcoidosis gene-expression datasets is of great importance to robustly establish differentially expressed genes and signalling pathways.

Methods: We performed meta-analysis on 22 published gene-expression studies on sarcoidosis. Datasets were analysed systematically using same statistical cut-offs. Differentially expressed genes were identified by pooling of *p*-values using Edgington's method and analysed for pathways using Ingenuity Pathway Analysis software.

Results: A consistent and significant signature of novel and well-known genes was identified, those collectively implicated both type I and type II interferon mediated signalling pathways in sarcoidosis. *In silico* functional analysis showed consistent downregulation of eukaryotic initiation factor 2 signalling, whereas cytokines like interferons and transcription factor STAT1 were upregulated. Furthermore, we analysed affected tissues to detect differentially expressed genes likely to be involved in granuloma biology. This revealed that matrix metallopeptidase 12 was exclusively upregulated in affected tissues, suggesting a crucial role in disease pathogenesis.

Discussion: Our analysis provides a concise gene signature in sarcoidosis and expands our knowledge about the pathogenesis. Our results are of importance to improve current diagnostic approaches and monitoring strategies as well as in the development of targeted therapeutics.

KEYWORDS

sarcoidosis, granuloma, gene expression, meta-analysis, IFN-JAK-STAT signalling

Background

Sarcoidosis is an inflammatory disorder characterized by the formation of non-caseating epithelioid granulomas in various organs. However, the aetiology and pathogenesis of sarcoidosis are not fully understood. The lungs and hilar lymph nodes are most often affected, but almost all organs can be involved (1). The clinical presentation of sarcoidosis is therefore heterogeneous and can range from small benign skin lesions to chronic systemic inflammation. This variability in disease presentation makes it challenging to diagnose sarcoidosis. This diagnosis is mainly based on clinical and radiographic presentation, pathological evidence of non-caseating granulomas and the exclusion of other granulomatous diseases (2, 3). Even though, the lesions resolve spontaneously in a large portion of patients, irreversible tissue damage, like pulmonary fibrosis occurs in up to 20% of cases leading to increased morbidity and mortality (4, 5). More knowledge about the molecular mechanisms involved in the pathophysiology of sarcoidosis is warranted to develop better and more adequate monitoring strategies as well as treatment options.

There is a consensus that complex interactions between genetic and environmental triggers culminate into an aberrant immune response to unidentified antigens including infectious agents (6). Moreover, it is proposed that sarcoidosis encompasses both autoinflammatory and autoimmune features (7). Particularly the similarities of sarcoidosis with Blau syndrome (early-onset sarcoidosis), caused by mutations in nucleotide-binding oligomerization domain containing 2 (*NOD2*), supports an autoinflammatory hypothesis. Meanwhile, the association with *HLA-DRB1* genotypes provides more evidence in the direction of autoimmunity (8). In patients with sarcoidosis, an altered T helper 1 (Th1) immune response is observed partly through activation of signal transducer and activator of transcription 1 (*STAT1*) and production of interferons (IFNs) (9, 10). This has led to the targeting of the Janus kinase (JAK)-STAT signalling pathway in sarcoidosis using inhibitors like baricitinib (11) and tofacitinib (12) for patients with refractory symptoms.

Gene expression studies, also referred to as transcriptomic studies, have been very promising and widely used to identify disease-associated differentially expressed genes (DEGs). Such studies can provide candidate targets for therapy as well as disease biomarkers. However, a concern about transcriptomic studies is their reproducibility and generalizability mainly due to differences in study design, data analysis strategies and limited sample size. With the increasing awareness of open data, more and more datasets are becoming available allowing to identify specific disease associated genes and pathways suitable for therapeutic intervention. Systematic meta-analysis of transcriptomic data for sarcoidosis provides a powerful tool to identify robust gene signatures. Therefore, we systematically analysed sarcoidosis transcriptome by performing meta-analysis on 22 gene expression datasets obtained from various tissues, bronchoalveolar lavage fluid (BALF) and peripheral blood comparing sarcoidosis patients with healthy controls taking into account both the blood and target tissue samples.

Methods

Dataset acquisition

The genome expression omnibus (GEO) database (13) was queried for expression profiling by array or high-throughput sequencing using the following string: "Sarcoidosis [All Fields] AND GSE [All Fields]." Datasets containing human RNA expression were selected and further explored with original papers for study design. We excluded single-cell RNA sequencing experiments as well as those with unclear study design or sample annotation. Only datasets with more than four sarcoidosis patients and healthy controls were included in this comparative study. Raw data from these selected studies were downloaded from the GEO database and further processed.

Dataset preparation and processing

Normalization of Affymetrix and Illumina BeadChip array data was performed with robust microarray average (RMA) within the R package *affy* (14) and with *neqc* within the R package *limma* (15), respectively. Quantile normalization on the gProcessedSignal and subsequent log-transformation, was used for Agilent datasets. RNAseq count data was normalized within the R package *DESeq2* (16). Principal component analysis was performed to assess batch effects and if that was suspected, the *ComBat* function within the *SVA* R package (17) was used for batch correction of the gene expression dataset. Differential expression was calculated for every dataset using the *limma* and *DESeq2* R packages for array and sequencing data, respectively. If a dataset contained multiple cell-types or tissues, it was analysed separately based on the cell-types. A paired analysis was performed on datasets containing multiple samples from a single individual. We did not adjust for confounding factors like age, gender or ethnicity, due to the scarcity of data in the datasets, whereas we aimed at analysing each dataset systematically and uniformly.

Analysis of datasets for shared genes and signalling pathways

First, we investigated the sarcoidosis datasets for commonly DEGs. In this analysis, we did not differentiate datasets based on cell type or tissue. A list of DEGs was obtained per dataset by setting the significance level of the adjusted p-value (p_{adj}) to less than 0.05 and subsequently annotated the acquired gene lists with HUGO gene symbols. Because the threshold put on the log fold change (FC) depends on the gene and experimental context (18), we did not use the log FC to determine DEGs as a standardized threshold would be too lenient for some datasets and too stringent for others. Ingenuity Pathway Analysis (IPA) (19) was used for functional core analysis of all acquired gene lists. A comparative analysis was performed to investigate the pathways and upstream regulators involved across datasets. Second, the gene lists of DEGs were analysed for overlapping genes. After the individual differential expression analyses, a meta-analysis was performed to test the robustness of our findings. The p-values from the individual analyses were combined using Edgington's method (20) within the R package *metap* (21). The calculated $p_{meta\text{-}analysis}$ was subsequently Bonferroni corrected to adjust for multiple testing. Pattern of differential expression was investigated for each gene in the individual datasets through the log FC, where a positive and negative log FC were categorized as upregulation and downregulation of the gene, respectively. If the pattern differed in more than three datasets, the gene was not considered consistently differentially expressed and was excluded from further analysis. Thus an acquired gene list was loaded in IPA to build an integrated gene

network. Furthermore, to examine these genes in the context of JAK–STAT pathway, IFN signalling was investigated in a separate analysis. These genes were loaded in a dataset (GSE110549) generated by the Immunological Genome Project exploring *in vivo* effects of IFN-α and IFN-γ stimulation on murine macrophages (22).

Identification of tissue-specific genes and pathways

Furthermore, we investigated whether there are cell type or tissue-specific gene expression profiles in sarcoidosis to determine biological mechanisms of granulomatous inflammation and find specific targets for therapy. Because the majority of datasets contained blood samples, we made a distinction between blood and other cell types. We investigated the acquired gene lists for genes exclusively differentially expressed in the affected sarcoidosis tissues. Finally, identified genes that were exclusively up- or downregulated in tissues were uploaded in IPA to investigate possible tissue-specific pathways, upstream regulators and gene networks.

Results

Search and analysis of GEO database

A systematic search was conducted in the GEO database for sarcoidosis specific datasets. Up until January 2024, our search strategy retrieved 80 datasets worldwide, of which 22 datasets (23–41) were selected and subjected to further analysis after rigorous exclusion criteria (Table 1 and Figure 1). In these 22 datasets, a total of 461 sarcoidosis and 497 healthy control samples across multiple tissues and cell types were analysed. Thirteen datasets used blood derived cells for their analysis, whereas 8 datasets were derived from lung, BALF, nasal brush, lymph node and skin tissue. One dataset contained both blood and BALF samples, which were analysed as separate datasets. Besides diagnostic parameters, clinical details were often not included or reported for individual samples. According to the original manuscripts, 11 out of 22 datasets contained at least one sample that used immunosuppressive therapy for sarcoidosis. Analysis of individual datasets for DEGs ($p_{adj} < 0.05$) revealed altered expression of a large number of genes. Thus derived lists of DEGs per dataset ranged from zero to thousands of genes, including two out of 22 datasets that did not show any DEGs (Figure 1).

Genes and signalling pathways associated with sarcoidosis

In order to allow cross-platform comparison of the data sets, we annotated the lists of DEGs using the platform identifier with HUGO gene nomenclature. Highly consistent DEGs were identified: 30 genes were differentially expressed in at least 13 datasets, of which 20 showed very consistent pattern of differential expression (if the pattern of differential expression differed in more than three datasets, the gene was not considered as a DEG) across the datasets (Table 2). These genes were differentially expressed in both blood-derived datasets as well as in datasets derived from affected tissues. Guanlytate-binding protein 1 (*GBP1*), *STAT1* and tryptophayl-tRNA synthetase 1 (*WARS1*) showed most consistent differential expression in 16 out of 22 datasets and expression levels were found predominantly upregulated in sarcoidosis. Differential expression of genes high-affinity gamma FC receptor I (*FCGR1A* also referred as CD64), *GBP2*, and vesicle-associated membrane protein 5 (*VAMP5*) was observed in 15 datasets. Interestingly, *STAT1*, *GBP1*, *GBP2* and *GBP5* were upregulated in all datasets (Table 2).

Next, we integrated all the datasets to perform meta-analysis on the *p*-values to identify additional consistent DEGs as some genes could not be investigated on all transcriptomic platforms due to their varied design. Pooling of *p*-values with Edgington's method and subsequent Bonferroni correction resulted in 36 significantly expressed genes. Of these 36 genes, 12 displayed variable pattern of up or down regulation and hence, they were excluded from further analysis (Table 3). To investigate the relationship between the remaining 24 genes, a connectivity plot was generated in IPA (Figure 2). Remarkably, more than half (14 out of the 24) of the genes were well connected in the connectivity network. *STAT1*, IL-12, and IFN signalling pathways appeared centrally positioned within the network and were predicted to be activated. Also insulin was positioned in the network. Both type I IFN (IFN-α and IFN-β) and type II IFN (IFN-γ) were predicted to be upregulated within the connectivity network. Therefore, we explored this further in a murine dataset specifically on IFN signalling (GSE110549) where we verified that half of the genes identified in the meta-analysis were upregulated upon IFN stimulation, advocating for a pivotal role of altered IFN signalling in sarcoidosis (Supplementary Figure S1).

In silico functional analysis for the signalling pathways

All individual gene lists of DEGs were loaded into IPA for functional pathway analyses. Eukaryoitc initation factor 2 (EIF2) signalling pathway was consistently predicted to be downregulated across datasets in sarcoidosis patients as compared to healthy controls. On the other hand, pro-inflammatory pathways like IFN-signalling, neuro-inflammation and hypercytokinaemia/hyperchemokinaemia were predicted to be activated in sarcoidosis patients (Figure 3A). This was further indicated by the upstream regulators found within IPA analysis. Pro-inflammatory transcription factors and cytokines (i.e., *STAT1*, TNF, IL-6 and IL-1β) were consistently predicted to be activated (Figure 3B). Also type I and type II IFNs were consistently predicted to be upregulated, strongly suggesting activation of the IFN-STAT1 pathway. Only a few upstream regulators were consistently downregulated after stringent filtering within IPA. Inhibitors like, PD98059 U0126 and LY294002 (Figure 3B) target the MAPK/ERK and PI3K/AKT/mTOR pathways, suggesting that these chemical compounds could inhibit these activated signalling pathways in sarcoidosis. These pathways often run in parallel and converge to regulate important cellular processes.

Identification of tissue-specific DEGs

Finally, we investigated tissue-specific gene signatures in sarcoidosis. We hypothesized that DEGs in affected tissues may play

TABLE 1 Included datasets and their characteristics.

GSE Accession No	Organism	Technique	Array/Sequencing technology	Tissue	Sarcoidosis (n=)	Control (n=)	Treated samples (at least one)	Reference
GSE83456	Human	Array	Illumina HumanHT-12 V4.0	Whole blood	49	61	Unknown	(24)
GSE37912	Human	Array	Affymetrix Human Exon 1.0 ST	PBMC	39	35	Yes	(37)
GSE19314	Human	Array	Affymetrix Human Genome U133 Plus 2.0	PBMC	38	20	Yes	(31)
GSE42832	Human	Array	lllumina HumanHT-12 V4.0	PBMC subsets	30	30	Yes	(25)
GSE32887	Human	Array	Affymetrix Human Genome U133 Plus 2.0	Skin	26	5	Yes	(30)
GSE73394	Human	Array	Affymetrix Human Gene 1.0 ST	BALF	26	20	Unknown	(34)
GSE42826	Human	Array	Illumina HumanHT-12 V4.0	Whole blood	25	52	Yes	(25)
GSE42830	Human	Array	Illumina HumanHT-12 V4.0	Whole blood	25	38	Yes	(25)
GSE34608	Human	Array	Agilent-014850 Whole Human 4x44K G4112F	PBMC	18	18	Unknown	(33)
GSE75023	Human	Array	Affymetrix Human Genome U133A	BALF	15	12	Yes	(28)
GSE119136	Human	Array	Affymetrix Human Gene 1.0 ST	Nasal brushings	14	12	Yes	(29)
GSE18781	Human	Array	Affymetrix Human Genome U133 Plus 2.0	Whole blood	12	25	Yes	(39)
GSE1907	Human	Array	Affymetrix Human Genome U95A	PBMC	12	12	Unknown	(38)
GSE42825	Human	Array	Illumina HumanHT-12 V4.0	Whole blood	11	23	Yes	(25)
GSE105149	Human	Array	Affymetrix Human Genome U133 Plus 2.0	Lacrimal gland	8	7	Yes	(36)
GSE56998	Human	Array	Affymetrix Human Exon 1.0 ST Array	CD4 T cells	8	8	Unknown	(23)
GSE16538	Human	Array	Affymetrix Human Genome U133 Plus 2.0	Lung	6	6	No	(27)
GSE174659	Human	Sequencing	Illumina HiSeq 4,000	Whole blood & BALF	62	76	Unknown	(32)
GSE155644	Human	Sequencing	Illumina MiSeq	Whole blood	14	14	No	(41)
GSE174443	Human	Sequencing	Ion Torrent Proton	Lymph node	10	7	Unknown	(35)
GSE148036	Human	Sequencing	Illumina HiSeq 3,000	Lung	5	5	No	(26)
GSE192829	Human	Sequencing	Illumina NextSeq 500	PBMC	8	11	No	(40)

10.3389/fmed.2024.1381031

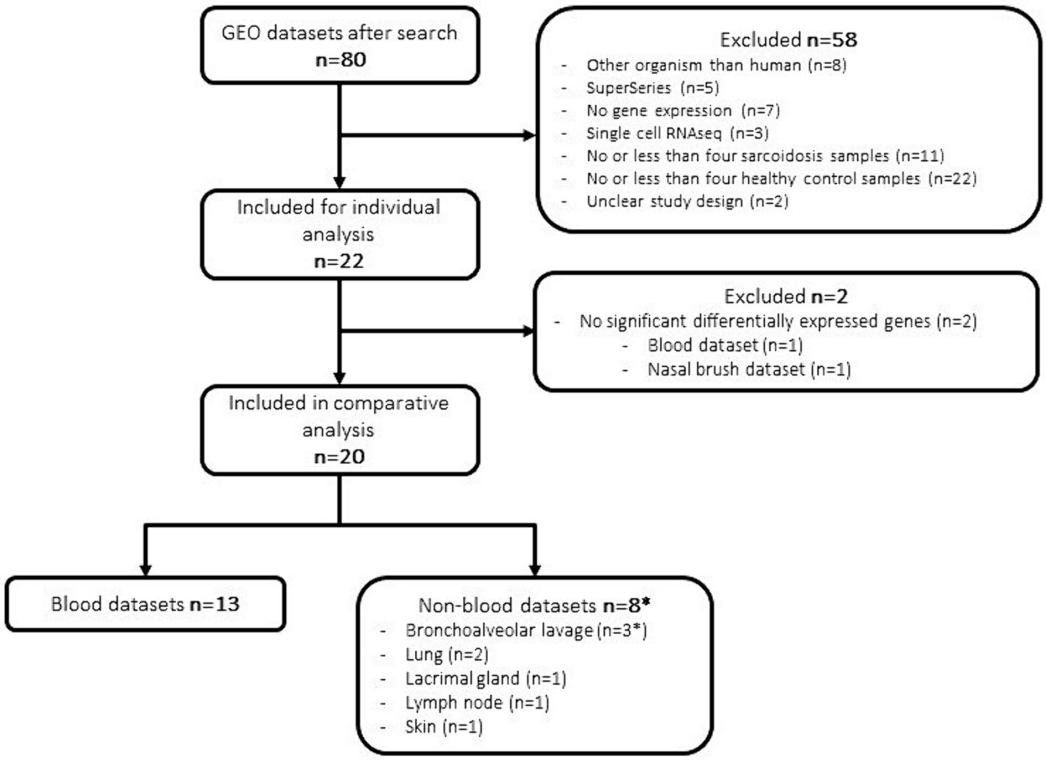

FIGURE 1
Flowchart of dataset selection and analysis. *One dataset contained both blood and BALF samples.

a crucial role in granuloma formation and their maintenance. Nine datasets were retrieved from tissues, of which 8 showed DEGs. A disintegrin and metalloprotease like decysin 1 (*ADAMDEC1*) and WD repeat and SOCS box-containing protein 1 (*WSB1*) were differentially expressed (upregulated) in seven datasets, however, these were also differentially expressed in a few blood datasets. The most consistent tissue-specific DEG was matrix metallopeptidase 12 (*MMP12*), which was highly upregulated in 6 tissue-derived datasets (two BALF, two lung, one lacrimal gland and one lymph node derived dataset) and in no blood-derived dataset. In five of the tissue-specific datasets C-X-C motif chemokine receptor 6 (*CXCR6*) and syntrophin beta 2 (*SNTB2*), whereas in four datasets colony stimulating factor 2 (*CSF2*), fatty acid desaturase 1 (*FADS1*), interleukin 18 binding protein (*IL18BP*), acyl-CoA synthetase family member 2 (*ACSF2*), cystatin B (*CSTB*), C-C motif chemokine ligand 4 (*CCL4*), adenosine deaminase (*ADA*), *JAK3*, malic enzyme 1 (*ME1*), muscle RAS Oncogene Homolog (*MRAS*) and RAS guanyl releasing protein 3 (*RASGRP3*) were differentially expressed. Most of these genes are involved in the activation of processes such as leukocyte recruitment and migration and form a concise network around PI3K and MAPK/ERK signalling (Figure 4).

Discussion

Using public datasets in the GEO database, we studied 22 sarcoidosis datasets and performed a meta-analysis to identify genes and pathways those are common across these studies and differentiate sarcoidosis patients from healthy controls. We found that there are clear differences in gene expression profiles for *GBP1*, *STAT1* and *WARS1* among others between sarcoidosis patients and healthy controls. Furthermore, the integration of datasets provided a comprehensive view for certain genes (*STAT1, WARS1, GBP1, VAMP5*, and *PSTPIP2*) being consistently expressed in the majority of datasets advocating for their role in the pathogenesis of sarcoidosis. These genes could potentially be used to develop meaningful genomic-derived biomarkers for sarcoidosis. In analogy to several other transcriptomic studies, we identified both type I and type II IFN signalling as important pathways (42), but in this meta-analysis also poorly studied pathways became apparent in the context of sarcoidosis such as EIF2 signalling and neuro-inflammation. Finally, we explored datasets with samples from sarcoidosis-affected tissues and identified tissue-specific DEGs (*MMP12*, *CXCR6*, and *SNTB2*) those likely to play important respective roles, specifically in granulomatous inflammation. Hence, it can be proposed that these specific genes and their respective translated proteins could be targeted for precise therapy of sarcoidosis lesions.

Whether a gene is considered differentially expressed in transcriptomic studies revolves around the chosen cut-offs in (adjusted) *p*-value and fold change. The results between transcriptomic studies often differ due to differences in study design, analysis strategies and sample size. Therefore, to circumvent these issues, a meta-analysis on several datasets of particular disease, becomes an important tool for analysis and inferences increasing the strength of such studies to establish true signals (43). We assessed the differential expression pattern by integrating these datasets based on the *p*-values thus creating a robust analysis to identify consistently DEGs. In this study, to the best of our knowledge, the largest systematic meta-analysis of transcriptomic data in sarcoidosis is being presented.

TABLE 2 Differentially expressed genes in at least 13 datasets.

Gene symbol	Frequency differentially expressed	BAL (*n* = 3)	Blood (*n* = 13)	Lung (*n* = 2)	Lacrimal gland (*n* = 1)	Lymph node (*n* = 1)	Nasal brush (*n* = 1)	Skin (*n* = 1)	Number of datasets Upregulated/ Downregulated	General directionality in sarcoidosis patients
GBP1	16	2	10	1	1	1	0	0	21/0	Upregulated
STAT1	16	2	10	1	1	1	0	0	21/0	Upregulated
WARS1	16	2	11	0	1	1	0	0	20/1	Upregulated
FCGR1A	15	2	10	0	1	1	0	0	17/3	Upregulated
GBP2	15	1	10	1	1	1	0	0	21/0	Upregulated
VAMP5	15	2	10	1	1	0	0	0	20/1	Upregulated
GBP4	14	1	9	1	1	1	0	0	18/1	Upregulated
GBP5	14	1	10	0	1	1	0	0	19/0	Upregulated
LAP3	14	1	9	1	1	1	0	0	18/2	Upregulated
PSTPIP2	14	2	9	0	1	1	0	0	19/1	Upregulated
TAP1	14	2	9	1	1	0	0	0	19/2	Upregulated
ABLIM1	13	2	9	0	1	0	0	0	3/18	Downregulated
ANKRD22	13	1	9	0	1	1	0	0	18/1	Upregulated
C2	13	1	8	1	1	1	0	0	18/3	Upregulated
GADD45B	13	1	10	0	1	0	0	0	18/3	Upregulated
HNRNPDL	13	1	10	1	0	1	0	0	3/18	Downregulated
IRF1	13	2	8	0	1	1	0	0	19/2	Upregulated
MAFB	13	2	8	0	1	1	0	0	19/1	Upregulated
STX11	13	1	9	0	1	1	0	0	20/1	Upregulated
TYMP	13	2	9	0	1	1	0	0	19/2	Upregulated

List of genes that were differentially expressed in at least twelve datasets comparing sarcoidosis samples to healthy controls and which showed consistent directionality of differential expression across the datasets. Information by tissue type is incorporated in this table.

10.3389/fmed.2024.1381031

TABLE 3 Gene list of the significant ($p_{adj} < 0.05$) genes in the meta-analysis.

Gene symbol	Adjusted *p*-value	Number of datasets Upregulated/Downregulated
STAT1	9.27E-07	21/0
WARS1	1.26E-06	20/1
GBP1	1.79E-05	21/0
VAMP5	5.35E-05	20/1
PSTPIP2	7.20E-05	19/1
ANKRD22	9.75E-05	18/1
GBP5	0.000169	19/0
FCGR1A	0.000452	17/3
GBP2	0.000827	21/0
FCGR1B	0.001204	13/2
STX11	0.001292	20/1
PSME2	0.001619	20/1
UBE2L6	0.002172	20/1
GBP4	0.002705	18/1
PLEK	0.003109	19/2
CALHM6	0.005246	17/0
DHRS9	0.005825	18/2
IL15	0.014057	19/2
SOD2	0.0171	20/0
IRF1	0.0262	19/2
GLUL	0.027109	18/3
LAP3	0.03706	18/2
WDFY1	0.043271	19/0
FYB1	0.049856	19/2

Adjusted *p*-value is Bonferroni corrected. The direction of the differentially expressed gene is determined by investigating the directionality of fold change in the individual datasets. A gene was not considered significant if the directionality differed in more than three datasets from the main directionality.

Sarcoidosis research so far has predominantly focused on the IFN-STAT1 pathway linked to T cells and macrophages. In fact, there have been numerous reports on the development and exacerbation of sarcoidosis after IFN therapy (44, 45). Our study highlights the importance of this pathway as several of the most consistently DEGs are associated with IFN signalling. Furthermore, in IPA analysis JAK1/2 was found in the connectivity network, strongly suggesting the rationale to target the JAK-STAT pathway with JAK inhibitors in sarcoidosis patients. Interestingly, we identified multiple members of GBP gene family to be consistently differentially expressed (*GBP1, GBP2, GBP4 and GBP5*). These genes are a group of seven IFN-inducible GTPases implicated in the host defence against intracellular pathogens by targeting and inducing lysis of pathogen-containing vacuoles (46). Differential expression of GBPs as found in our study, is not only attributed to the pulmonary sarcoidosis, but aberrant expression of *GBP1* has also been demonstrated earlier in acute respiratory distress syndrome (47).

One of the most consistently differentially expressed genes was *FCGR1A* (CD64), which was differentially expressed in 15 out of 22 datasets. This gene is strongly induced by IFN-γ and plays a central role in antibody-dependent cytotoxicity and FCγ receptor-mediated phagocytosis (48). Phagosome and phagocytosis has been reported to be upregulated in monocytes of sarcoidosis patients (49). In proteomic studies, FCγ receptor-mediated phagocytosis is upregulated in sarcoidosis (50, 51). Several genes including *FCGR1A*, ubiquitin conjugating enzyme E2 L6 (*UBE2L6*) and *VAMP5*, found statistically significant in our meta-analysis have also been described in the context of tuberculosis (TB) (52). In both sarcoidosis and TB, granulomas are the hallmark, but the granulomas in sarcoidosis are non-caseating, whereas the granulomas in TB frequently contain a necrotic core. Most likely these diseases, despite their differences, may share common inflammatory pathways and mechanisms corroborated by overlapping gene expression profiles (25, 31, 33).

Our analysis revealed novel genes, such as *WARS1* and *VAMP5* those were never implicated in sarcoidosis. *WARS1* is an essential enzyme called tryptophayl-tRNA synthetase 1 that charges tryptophane to its cognate tRNA and also plays a role in the innate immune system. *WARS1* is upregulated upon infection and can act as a ligand of toll-like receptor (TLR) 2 and TLR4. This leads to secretion of cytokines and activation of various immune pathways (53). *VAMP5* is part of the SNARE protein family, which is involved in vesicle fusion and recycling (54). *VAMP5* is involved in intracellular transport including exocytosis, endocytosis and recycling of endosomes (55). These processes are closely related to autophagic pathways, which have been implicated in the pathogenesis of sarcoidosis (56). Identification of these genes adds to the knowledge about the genetics and

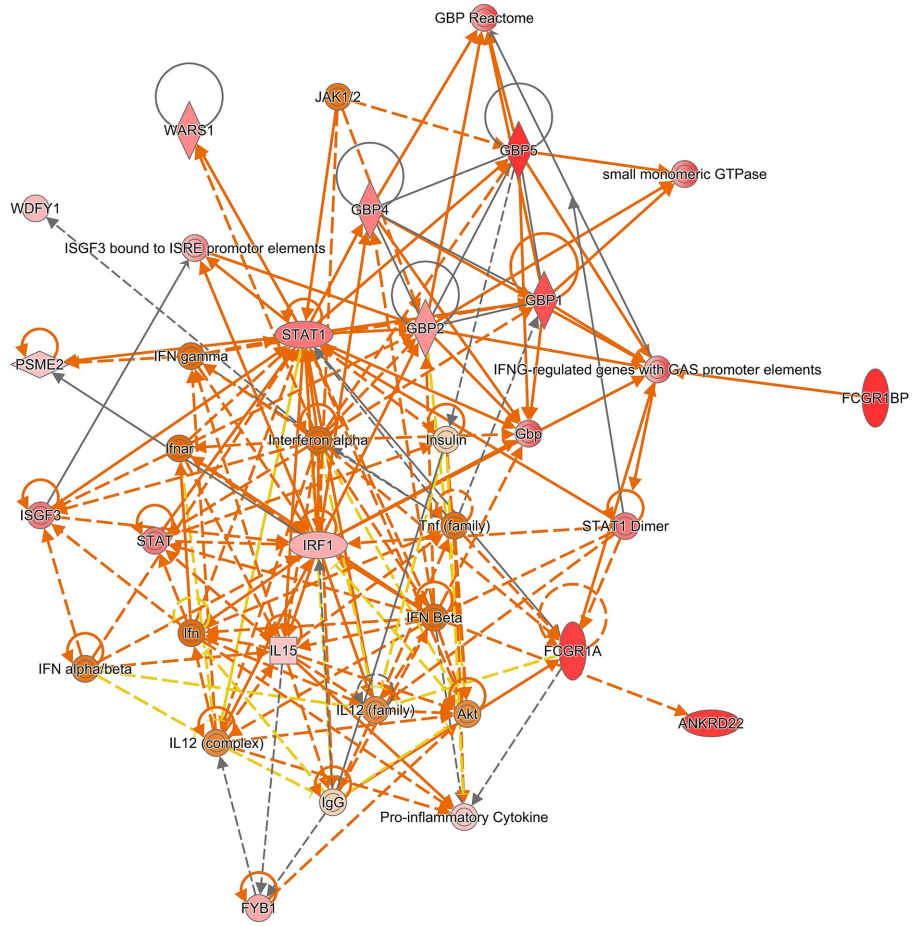

FIGURE 2
Connectivity network generated on 16 of the 25 significant genes from the meta-analysis. Centrally positioned in the network are *STAT1* and interferon alpha, which suggests these play an important role in sarcoidosis. Red and green molecules are up- and downregulated in the genelist respectively, whilst orange and blue are predicted to be up- and downregulated based on the input genelist by the molecule activity predictor (MAP) function within IPA.

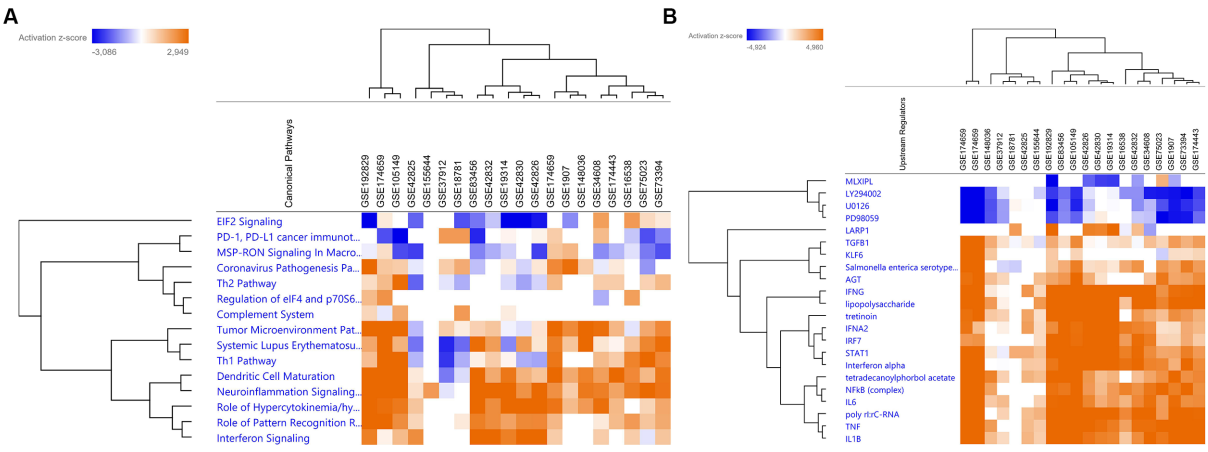

FIGURE 3
In silico functional analysis of the individual differentially expressed gene lists. **(A)** Heatmap of the comparison analysis within IPA that shows the canonical pathways involved in each individual dataset. Across the datasets, similar canonical pathways were predicted to be involved. **(B)** Heatmap that shows the upstream regulators predicted to be involved across the functional analyses of the individual datasets.

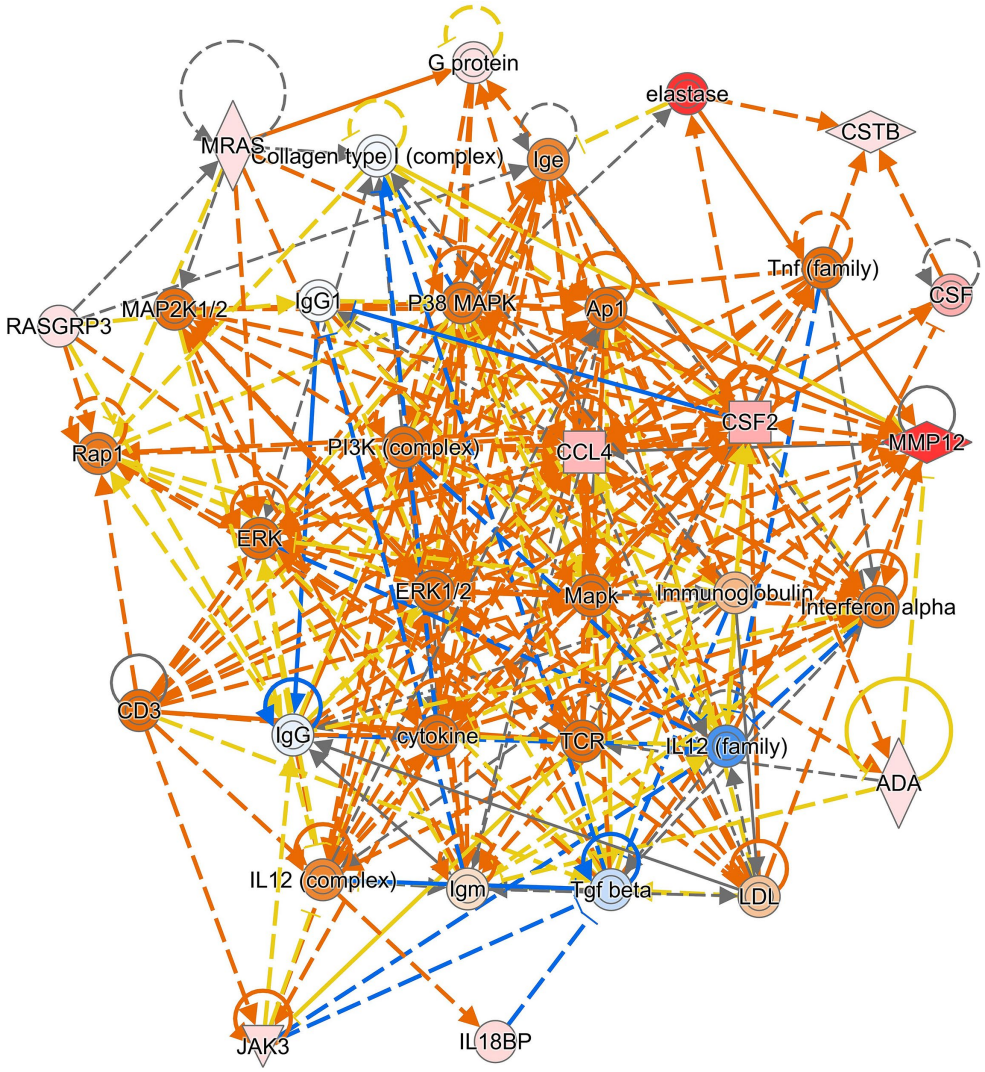

FIGURE 4
Connectivity network generated on 9 of the 14 genes with high tissue specificity. The PI3K-pathway and MAPK/ERK signalling are centrally positioned in the network. Red and green molecules are up- and downregulated in the genelist respectively, whilst orange and blue are predicted to be up- and downregulated based on the input genelist by the molecule activity predictor (MAP) function within IPA.

pathogenesis of sarcoidosis and opens avenues for further research into these proteins.

IPA analysis of datasets analysed in this study predicted EIF2 signalling to be downregulated. The EIF2 signalling cascade is involved in autophagy, protein translation and cell survival as well as the mammalian target of rapamycin (mTOR) pathway (57, 58). Recently, the mTOR pathway gained much attention in sarcoidosis (59–61), after the finding that constitutive activity of mTORC1 causes formation of granulomas (62). In this regard, Gupta and colleagues successfully treated a patient with pulmonary sarcoidosis with mTOR inhibitor sirolimus (63). Moreover, our tissue-specific analysis showed PI3K in the connectivity network, suggesting that the PI3K/mTOR pathway is involved only in affected tissues. Another interesting finding in IPA analysis was the upregulation of neuro-inflammation signalling pathway. Small fiber neuropathy is observed in about 30% of patients with systemic sarcoidosis (64), in which circulating inflammatory and neurotoxic cytokines may be involved (65). Recently, the upregulation of the neuro-inflammatory response was found by another group studying the sarcoidosis transcriptome and proteome (66). This interesting study is a meta-analysis as well, however, their strategy differed significantly from ours as they performed meta-analysis on the common pathways rather than DEGs. Moreover, recently two other meta-analyses on the transcriptome of sarcoidosis have been published (67, 68). These studies used only 11 and 13 studies respectively, excluding many relevant studies in their analyses. Therefore, our study is the largest systematic meta-analysis to identify robust DEGs which can be of value as biomarkers for sarcoidosis.

Interestingly, insulin was another hub in the connectivity network generated by IPA. Sarcoidosis patients are at increased risk for developing type 2 diabetes (T2D) (69). Much is unknown about this association, however, chronic inflammation and increased secretion of cytokines might predispose sarcoidosis patients to develop T2D (70). The IFN-γ/STAT1 upregulation, as found in our study, could be linked to insulin resistance in adipocytes through multiple mechanisms, including downregulation of the insulin receptor and

glucose transporter type 4 (71). Additionally, *STAT1* expression in white adipose tissue is elevated in prediabetic patients and STAT1 levels are positively correlated with plasma glucose (72). Together, these data suggest an important role for IFN-γ/STAT1 signalling in T2D observed in sarcoidosis patients, that is further exacerbated by steroid treatment and warrants attention from clinicians.

Finally, we looked for a tissue-specific gene signature in sarcoidosis and found *MMP12* upregulated in most tissue-specific datasets. *MMP12* is an elastase enzyme predominately produced by M2 tissue macrophages (73), those aggregate and form the characteristic multinucleated giant cells as seen in granulomas (74). Involvement of *MMP12* in sarcoidosis specially in granuloma progression has been reported previously (27, 75). In our study, *MMP12* was only differentially expressed at tissue sites, which is highly suggestive for its crucial role in granuloma formation. *CXCR6* was also tissue-specific in our analysis, and this gene has been found to be expressed in Th1 cells surrounding the central core of sarcoidosis granulomas (76). Therefore, tissue-specific DEGs like *MMP12* and *CXCR6*, or potentially its ligand *CXCL16*, could be interesting therapeutic targets for sarcoidosis lesions.

Limitations of the study

Among 22 datasets included in this meta-analysis, we were limited with respect to sample size, variable tissues and cell types as well as different technological platforms. Additionally, sarcoidosis patients can differ substantially in clinical presentation, disease progression and treatment response. Studies have shown distinct gene expression profiles between self-limiting sarcoidosis and progressive sarcoidosis (37, 77). Unfortunately, clinical information such as age, sex, ethnicity, treatment regimen, and disease activity were poorly reported by most studies. To appropriately circumvent these limitations, we opted for an iterative and systematic approach, through which we observed highly consistent DEGs and pathways in sarcoidosis despite this variability among studies. Furthermore, we were not able to adjust for other confounding factors such as lymphopenia, which is often observed in sarcoidosis patients and associated with disease activity (78). Attention should be paid to treatment regime and disease activity. Eleven of the 22 included datasets contained at least one sample that used immunosuppressive therapy, whereas three datasets excluded patients who were on immunosuppression. We did not observe major differences in DEGs between these datasets, but it is known that immunosuppression can alter gene expression in disease (79). Appropriate studies with disease endotyping are needed to study and identify potential biomarkers stratifying sarcoidosis subgroups leading to a precision-medicine approach (80). Finally, only eight of the 22 studies investigated here were derived from target tissues, highlighting the need for more studies investigating tissue-specific signatures to gain more insight in the genes and pathways involved in granuloma formation. Whether the described genes in this study represent activation or perpetuation of the disease needs further exploration.

Conclusion

In this meta-analysis study, 22 sarcoidosis gene expression datasets were systematically and uniformly assessed to identify DEGs and their signalling pathways. Integration of the results from individual datasets revealed a number of novel candidate genes (i.e., *GBPs*, *VAMP5* and *WARS1*) and pathways in addition to previously described DEGs in sarcoidosis. Meta-analysis identified a robust and compact gene signature that points towards altered IFN-JAK-STAT1 signalling in sarcoidosis. Our findings add to the emerging evidence to employ JAK inhibitors as a targeted treatment in sarcoidosis patients. More strikingly, the DEGs found in our meta-analysis can further be explored to develop genomic-derived biomarkers for sarcoidosis. We found tissue-specific signature of genes like *MMP12*, *CXCR6*, and *SNTB2* suggesting their pathways are likely to be involved in granuloma formation and progression and could eventually be potential therapeutic targets for sarcoidosis. Clinical manifestation still remains a challenge with respect to disease activity and progression, which warrants the need for further transcriptomic studies with endotyping investigating pulmonary phenotypes and immune responses.

Data availability statement

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

Ethics statement

Ethical approval was not required for the studies involving humans because the current study uses existing publicly available data from gene expression studies, which obtained consent and ethical approval. No new data was generated in this study. Ethical approval was therefore not required. The studies were conducted in accordance with the local legislation and institutional requirements. The human samples used in this study were acquired from published publicly available datasets. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

RW: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. HS: Writing – review & editing, Writing – original draft, Investigation. SS: Writing – review & editing, Writing – original draft, Investigation, Formal analysis, Data curation. WD: Writing – review & editing, Writing – original draft, Investigation. HI: Writing – review & editing, Writing – original draft, Investigation. VD: Writing – review & editing, Writing – original draft, Investigation. PD: Writing – review & editing, Writing – original draft, Investigation. PH: Writing – review & editing, Writing – original draft, Investigation, Conceptualization. PS: Writing – review & editing, Writing – original draft, Investigation, Formal analysis, Data curation, Conceptualization.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The current study was supported by the European Union's Horizon 2020 research and innovation program under grant agreement no. 779295 (ImmunAID—Immunome project consortium for AutoInflammatory Disorders).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2024.1381031/full#supplementary-material

References

1. Drent M, Crouser ED, Grunewald J. Challenges of sarcoidosis and its management. *N Engl J Med*. (2021) 385:1018–32. doi: 10.1056/NEJMra2101555

2. Hunninghake GW, Costabel U, Ando M, Baughman R, Cordier JF, du Bois R, et al. ATS/ERS/WASOG statement on sarcoidosis. American Thoracic Society/European Respiratory Society/world Association of Sarcoidosis and other granulomatous disorders. *Sarcoidosis Vasc Diffuse Lung Dis*. (1999) 16:149–73.

3. Crouser ED, Maier LA, Wilson KC, Bonham CA, Morgenthau AS, Patterson KC, et al. Diagnosis and detection of sarcoidosis. An official American Thoracic Society clinical practice guideline. *Am J Respir Crit Care Med*. (2020) 201:e26–51. doi: 10.1164/rccm.202002-0251ST

4. Iannuzzi MC, Rybicki BA, Teirstein AS. Sarcoidosis. *N Engl J Med*. (2007) 357:2153–65. doi: 10.1056/NEJMra071714

5. Gupta R, Judson MA, Baughman RP. Management of Advanced Pulmonary Sarcoidosis. *Am J Respir Crit Care Med*. (2022) 205:495–506. doi: 10.1164/rccm.202106-1366CI

6. Chen ES, Moller DR. Etiologies of sarcoidosis. *Clin Rev Allergy Immunol*. (2015) 49:6–18. doi: 10.1007/s12016-015-8481-z

7. McGonagle D, McDermott MF. A proposed classification of the immunological diseases. *PLoS Med*. (2006) 3:e297. doi: 10.1371/journal.pmed.0030297

8. Levin AM, Adrianto I, Datta I, Iannuzzi MC, Trudeau S, Li J, et al. Association of HLA-DRB1 with sarcoidosis susceptibility and progression in African Americans. *Am J Respir Cell Mol Biol*. (2015) 53:206–16. doi: 10.1165/rcmb.2014-0227OC

9. Prasse A, Georges CG, Biller H, Hamm H, Matthys H, Luttmann W, et al. Th1 cytokine pattern in sarcoidosis is expressed by bronchoalveolar CD4+ and CD8+ T cells. *Clin Exp Immunol*. (2000) 122:241–8. doi: 10.1046/j.1365-2249.2000.01365.x

10. Grunewald J, Grutters JC, Arkema EV, Saketkoo LA, Moller DR, Muller-Quernheim J. Sarcoidosis. *Nat Rev Dis Primers*. (2019) 5:45. doi: 10.1038/s41572-019-0096-x

11. Scheinberg M, Maluf F, Wagner J. Steroid-resistant sarcoidosis treated with baricitinib. *Ann Rheum Dis*. (2020) 79:1259–60. doi: 10.1136/annrheumdis-2020-217271

12. Friedman MA, Le B, Stevens J, Desmarais J, Seifer D, Ogle K, et al. Tofacitinib as a steroid-sparing therapy in pulmonary sarcoidosis, an open-label prospective proof-of-concept study. *Lung*. (2021) 199:147–53. doi: 10.1007/s00408-021-00436-8

13. Edgar R, Domrachev M, Lash AE. Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. (2002) 30:207–10. doi: 10.1093/nar/30.1.207

14. Gautier L, Cope L, Bolstad BM, Irizarry RA. Affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*. (2004) 20:307–15. doi: 10.1093/bioinformatics/btg405

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

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

17. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. (2012) 28:882–3. doi: 10.1093/bioinformatics/bts034

18. McCarthy DJ, Smyth GK. Testing significance relative to a fold-change threshold is a TREAT. *Bioinformatics*. (2009) 25:765–71. doi: 10.1093/bioinformatics/btp053

19. Kramer A, Green J, Pollard J, Tugendreich S. Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics*. (2014) 30:523–30. doi: 10.1093/bioinformatics/btt703

20. Edgington ES. An additive method for combining probability values from independent experiments. *J Psychol*. (1972) 80:351–63. doi: 10.1080/00223980.1972.9924813

21. Dewey M. *Metap: meta-analysis of significance values. R package version*, vol. 1 (2021). 5 p.

22. Mostafavi S, Yoshida H, Moodley D, LeBoite H, Rothamel K, Raj T, et al. Parsing the interferon transcriptional Network and its disease associations. *Cell*. (2016) 164:564–78. doi: 10.1016/j.cell.2015.12.032

23. Bignon A, Regent A, Klipfel L, Desnoyer A, de la Grange P, Martinez V, et al. DUSP4-mediated accelerated T-cell senescence in idiopathic CD4 lymphopenia. *Blood*. (2015) 125:2507–18. doi: 10.1182/blood-2014-08-598565

24. Blankley S, Graham CM, Turner J, Berry MP, Bloom CI, Xu Z, et al. The transcriptional signature of active tuberculosis reflects symptom status in extra-pulmonary and pulmonary tuberculosis. *PLoS One*. (2016) 11:e0162220. doi: 10.1371/journal.pone.0162220

25. Bloom CI, Graham CM, Berry MP, Rozakeas F, Redford PS, Wang Y, et al. Transcriptional blood signatures distinguish pulmonary tuberculosis, pulmonary sarcoidosis, pneumonias and lung cancers. *PLoS One*. (2013) 8:e70630. doi: 10.1371/journal.pone.0070630

26. Chai Q, Lu Z, Liu Z, Zhong Y, Zhang F, Qiu C, et al. Lung gene expression signatures suggest pathogenic links and molecular markers for pulmonary tuberculosis, adenocarcinoma and sarcoidosis. *Commun Biol*. (2020) 3:604. doi: 10.1038/s42003-020-01318-0

27. Crouser ED, Culver DA, Knox KS, Julian MW, Shao G, Abraham S, et al. Gene expression profiling identifies MMP-12 and ADAMDEC1 as potential pathogenic mediators of pulmonary sarcoidosis. *Am J Respir Crit Care Med*. (2009) 179:929–38. doi: 10.1164/rccm.200803-490OC

28. Gharib SA, Malur A, Huizar I, Barna BP, Kavuru MS, Schnapp LM, et al. Sarcoidosis activates diverse transcriptional programs in bronchoalveolar lavage cells. *Respir Res*. (2016) 17:93. doi: 10.1186/s12931-016-0411-y

29. Grayson PC, Steiling K, Platt M, Berman JS, Zhang X, Xiao J, et al. Defining the nasal transcriptome in granulomatosis with polyangiitis (Wegener's). *Arthritis Rheumatol*. (2015) 67:2233–9. doi: 10.1002/art.39185

30. Judson MA, Marchell RM, Mascelli M, Piantone A, Barnathan ES, Petty KJ, et al. Molecular profiling and gene expression analysis in cutaneous sarcoidosis: the role of interleukin-12, interleukin-23, and the T-helper 17 pathway. *J Am Acad Dermatol*. (2012) 66:901–910.e2. doi: 10.1016/j.jaad.2011.06.017

31. Koth LL, Solberg OD, Peng JC, Bhakta NR, Nguyen CP, Woodruff PG. Sarcoidosis blood transcriptome reflects lung inflammation and overlaps with tuberculosis. *Am J Respir Crit Care Med*. (2011) 184:1153–63. doi: 10.1164/rccm.201106-1143OC

32. Lepzien R, Liu S, Czarnewski P, Nie M, Osterberg B, Baharom F, et al. Monocytes in sarcoidosis are potent tumour necrosis factor producers and predict disease outcome. *Eur Respir J*. (2021) 58:2003468. doi: 10.1183/13993003.03468-2020

33. Maertzdorf J, Weiner J, Mollenkopf HJ, Network TB, Bauer T, Prasse A, et al. Common patterns and disease-related signatures in tuberculosis and sarcoidosis. *Proc Natl Acad Sci USA*. (2012) 109:7853–8. doi: 10.1073/pnas.1121072109

34. Prasse A, Binder H, Schupp JC, Kayser G, Bargagli E, Jaeger B, et al. BAL cell gene expression is indicative of outcome and airway basal cell involvement in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*. (2019) 199:622–30. doi: 10.1164/rccm.201712-2551OC

35. Reichmann MT, Tezera LB, Vallejo AF, Vukmirovic M, Xiao R, Reynolds J, et al. Integrated transcriptomic analysis of human tuberculosis granulomas and a biomimetic model identifies therapeutic targets. *J Clin Invest*. (2021) 131:148136. doi: 10.1172/JCI148136

36. Rosenbaum JT, Choi D, Harrington CA, Wilson DJ, Grossniklaus HE, Sibley CH, et al. Gene expression profiling and heterogeneity of nonspecific orbital inflammation affecting the lacrimal gland. *JAMA Ophthalmol*. (2017) 135:1156–62. doi: 10.1001/jamaophthalmol.2017.3458

37. Zhou T, Zhang W, Sweiss NJ, Chen ES, Moller DR, Knox KS, et al. Peripheral blood gene expression as a novel genomic biomarker in complicated sarcoidosis. *PLoS One*. (2012) 7:e44818. doi: 10.1371/journal.pone.0044818

38. Rutherford RM, Staedtler F, Kehren J, Chibout SD, Joos L, Tamm M, et al. Functional genomics and prognosis in sarcoidosis--the critical role of antigen presentation. *Sarcoidosis Vasc Diffuse Lung Dis*. (2004) 21:10–8.

39. Sharma SM, Choi D, Planck SR, Harrington CA, Austin CR, Lewis JA, et al. Insights in to the pathogenesis of axial spondyloarthropathy based on gene expression profiles. *Arthritis Res Ther*. (2009) 11:R168. doi: 10.1186/ar2855

40. Yoshioka K, Sato H, Kawasaki T, Ishii D, Imamoto T, Abe M, et al. Transcriptome analysis of peripheral blood mononuclear cells in pulmonary sarcoidosis. *Front Med*. (2022) 9:822094. doi: 10.3389/fmed.2022.822094

41. Ascoli C, Schott CA, Huang Y, Turturice BA, Wang W, Ecanow N, et al. Altered transcription factor targeting is associated with differential peripheral blood mononuclear cell proportions in sarcoidosis. *Front Immunol*. (2022) 13:848759. doi: 10.3389/fimmu.2022.848759

42. Schupp JC, Vukmirovic M, Kaminski N, Prasse A. Transcriptome profiles in sarcoidosis and their potential role in disease prediction. *Curr Opin Pulm Med*. (2017) 23:487–92. doi: 10.1097/MCP.0000000000000403

43. Cahan P, Rovegno F, Mooney D, Newman JC, St Laurent G, McCaffrey TA. Meta-analysis of microarray results: challenges, opportunities, and recommendations for standardization. *Gene*. (2007) 401:12–8. doi: 10.1016/j.gene.2007.06.016

44. Chakravarty SD, Harris ME, Schreiner AM, Crow MK. Sarcoidosis triggered by interferon-Beta treatment of multiple sclerosis: a case report and focused literature review. *Semin Arthritis Rheum*. (2012) 42:206–12. doi: 10.1016/j.semarthrit.2012.03.008

45. Kato A, Ishihara M, Mizuki N. Interferon-induced sarcoidosis with uveitis as the initial symptom: a case report and review of the literature. *J Med Case Rep*. (2021) 15:568. doi: 10.1186/s13256-021-03181-x

46. Ngo CC, Man SM. Mechanisms and functions of guanylate-binding proteins and related interferon-inducible GTPases: roles in intracellular lysis of pathogens. *Cell Microbiol*. (2017) 19:e12791. doi: 10.1111/cmi.12791

47. Kong SL, Chui P, Lim B, Salto-Tellez M. Elucidating the molecular physiopathology of acute respiratory distress syndrome in severe acute respiratory syndrome patients. *Virus Res*. (2009) 145:260–9. doi: 10.1016/j.virusres.2009.07.014

48. Dubaniewicz A, Typiak M, Wybieralska M, Szadurska M, Nowakowski S, Staniewicz-Panasik A, et al. Changed phagocytic activity and pattern of Fcgamma and complement receptors on blood monocytes in sarcoidosis. *Hum Immunol*. (2012) 73:788–94. doi: 10.1016/j.humimm.2012.05.005

49. Talreja J, Farshi P, Alazizi A, Luca F, Pique-Regi R, Samavati L. RNA-sequencing identifies novel pathways in sarcoidosis monocytes. *Sci Rep*. (2017) 7:2720. doi: 10.1038/s41598-017-02941-4

50. Kjellin H, Silva E, Branca RM, Eklund A, Jakobsson PJ, Grunewald J, et al. Alterations in the membrane-associated proteome fraction of alveolar macrophages in sarcoidosis. *Sarcoidosis Vasc Diffuse Lung Dis*. (2016) 33:17–28.

51. Silva E, Souchelnytskyi S, Kasuga K, Eklund A, Grunewald J, Wheelock AM. Quantitative intact proteomics investigations of alveolar macrophages in sarcoidosis. *Eur Respir J*. (2013) 41:1331–9. doi: 10.1183/09031936.00178111

52. La Distia NR, Sitompul R, Bakker M, Versnel MA, Swagemakers SMA, van der Spek PJ, et al. Type 1 interferon-inducible gene expression in QuantiFERON gold TB-positive uveitis: a tool to stratify a high versus low risk of active tuberculosis? *PLoS One*. (2018) 13:e0206073. doi: 10.1371/journal.pone.0206073

53. Lee HC, Lee ES, Uddin MB, Kim TH, Kim JH, Chathuranga K, et al. Released Tryptophanyl-tRNA Synthetase stimulates innate immune responses against viral infection. *J Virol*. (2019) 93:e18. doi: 10.1128/JVI.01291-18

54. Hong W. SNAREs and traffic. *Biochim Biophys Acta*. (2005) 1744:120–44. doi: 10.1016/j.bbamcr.2005.03.014

55. Tajika Y, Takahashi M, Khairani AF, Ueno H, Murakami T, Yorifuji H. Vesicular transport system in myotubes: ultrastructural study and signposting with vesicle-associated membrane proteins. *Histochem Cell Biol*. (2014) 141:441–54. doi: 10.1007/s00418-013-1164-z

56. Pacheco Y, Valeyre D, El Jammal T, Vallee M, Chevalier F, Lamartine J, et al. Autophagy and Mitophagy-related pathways at the crossroads of genetic pathways involved in familial sarcoidosis and host-pathogen interactions induced by coronaviruses. *Cells*. (2021) 10:81995. doi: 10.3390/cells10081995

57. Wengrod JC, Wang D, Weiss S, Zhong H, Osman I, Gardner LB. Phosphorylation of eIF2alpha triggered by mTORC1 inhibition and PP6C activation is required for autophagy and is aberrant in PP6C-mutated melanoma. *Sci Signal*. (2015) 8:e899. doi: 10.1126/scisignal.aaa0899

58. Wengrod JC, Gardner LB. Cellular adaptation to nutrient deprivation: crosstalk between the mTORC1 and eIF2alpha signaling pathways and implications for autophagy. *Cell Cycle*. (2015) 14:2571–7. doi: 10.1080/15384101.2015.1056947

59. Crouser ED, Locke LW, Julian MW, Bicer S, Sadee W, White P, et al. Phagosome-regulated mTOR signalling during sarcoidosis granuloma biogenesis. *Eur Respir J*. (2021) 57:2002695. doi: 10.1183/13993003.02695-2020

60. Pacheco Y, Lim CX, Weichhart T, Valeyre D, Bentaher A, Calender A. Sarcoidosis and the mTOR, Rac1, and autophagy triad. *Trends Immunol*. (2020) 41:286–99. doi: 10.1016/j.it.2020.01.007

61. Rastogi R, Jiang Z, Ahmad N, Rosati R, Liu Y, Beuret L, et al. Rapamycin induces mitogen-activated protein (MAP) kinase phosphatase-1 (MKP-1) expression through activation of protein kinase B and mitogen-activated protein kinase kinase pathways. *J Biol Chem*. (2013) 288:33966–77. doi: 10.1074/jbc.M113.492702

62. Linke M, Pham HT, Katholnig K, Schnoller T, Miller A, Demel F, et al. Chronic signaling via the metabolic checkpoint kinase mTORC1 induces macrophage granuloma formation and marks sarcoidosis progression. *Nat Immunol*. (2017) 18:293–302. doi: 10.1038/ni.3655

63. Gupta N, Bleesing JH, McCormack FX. Successful response to treatment with Sirolimus in pulmonary sarcoidosis. *Am J Respir Crit Care Med*. (2020) 202:e119–20. doi: 10.1164/rccm.202004-0914IM

64. Tavee JO, Karwa K, Ahmed Z, Thompson N, Parambil J, Culver DA. Sarcoidosis-associated small fiber neuropathy in a large cohort: clinical aspects and response to IVIG and anti-TNF alpha treatment. *Respir Med*. (2017) 126:135–8. doi: 10.1016/j.rmed.2017.03.011

65. Basantsova NY, Starshinova AA, Dori A, Zinchenko YS, Yablonskiy PK, Shoenfeld Y. Small-fiber neuropathy definition, diagnosis, and treatment. *Neurol Sci*. (2019) 40:1343–50. doi: 10.1007/s10072-019-03871-x

66. Bhargava M, Liao S-Y, Elliott ED, Maier LA, Leach SM. The landscape of transcriptomics and proteomics studies in sarcoidosis. *ERJ Open Res*. (2021) 8:00621–2021. doi: 10.1183/23120541.00621-2021

67. Jiang Y, Jiang D, Costabel U, Dai H, Wang C. A transcriptomics-based meta-analysis identifies a cross-tissue signature for sarcoidosis. *Front Med*. (2022) 9:960266. doi: 10.3389/fmed.2022.960266

68. Duo M, Liu Z, Li P, Wang Y, Zhang Y, Weng S, et al. Integrative bioinformatics analysis to explore a robust diagnostic signature and landscape of immune cell infiltration in sarcoidosis. *Front Med (Lausanne)*. (2022) 9:942177. doi: 10.3389/fmed.2022.942177

69. Entrop JP, Kullberg S, Grunewald J, Eklund A, Brismar K, Arkema EV. Type 2 diabetes risk in sarcoidosis patients untreated and treated with corticosteroids. *ERJ Open Res*. (2021) 7:00028–2021. doi: 10.1183/23120541.00028-2021

70. Donath MY, Shoelson SE. Type 2 diabetes as an inflammatory disease. *Nat Rev Immunol*. (2011) 11:98–107. doi: 10.1038/nri2925

71. McGillicuddy FC, Chiquoine EH, Hinkle CC, Kim RJ, Shah R, Roche HM, et al. Interferon gamma attenuates insulin signaling, lipid storage, and differentiation in human adipocytes via activation of the JAK/STAT pathway. *J Biol Chem*. (2009) 284:31936–44. doi: 10.1074/jbc.M109.061655

72. Cox AR, Chernis N, Bader DA, Saha PK, Masschelin PM, Felix JB, et al. STAT1 dissociates adipose tissue inflammation from insulin sensitivity in obesity. *Diabetes*. (2020) 69:2630–41. doi: 10.2337/db20-0384

73. Nakagomi D, Suzuki K, Meguro K, Hosokawa J, Tamachi T, Takatori H, et al. Matrix metalloproteinase 12 is produced by M2 macrophages and plays important roles in the development of contact hypersensitivity. *J Allergy Clin Immunol*. (2015) 135:1397–400. doi: 10.1016/j.jaci.2014.10.055

74. Locke LW, Crouser ED, White P, Julian MW, Caceres EG, Papp AC, et al. IL-13-regulated macrophage polarization during granuloma formation in an in vitro human sarcoidosis model. *Am J Respir Cell Mol Biol*. (2019) 60:84–95. doi: 10.1165/rcmb.2018-0053OC

75. Mohan A, Neequaye N, Malur A, Soliman E, McPeek M, Leffler N, et al. Matrix Metalloproteinase-12 is required for granuloma progression. *Front Immunol*. (2020) 11:553949. doi: 10.3389/fimmu.2020.553949

76. Agostini C, Cabrelle A, Calabrese F, Bortoli M, Scquizzato E, Carraro S, et al. Role for CXCR6 and its ligand CXCL16 in the pathogenesis of T-cell alveolitis in sarcoidosis. *Am J Respir Crit Care Med*. (2005) 172:1290–8. doi: 10.1164/rccm.200501-142OC

77. Lockstone HE, Sanderson S, Kulakova N, Baban D, Leonard A, Kok WL, et al. Gene set analysis of lung samples provides insight into pathogenesis of progressive, fibrotic pulmonary sarcoidosis. *Am J Respir Crit Care Med*. (2010) 181:1367–75. doi: 10.1164/rccm.200912-1855OC

78. Sweiss NJ, Salloum R, Gandhi S, Alegre ML, Sawaqed R, Badaracco M, et al. Significant CD4, CD8, and CD19 lymphopenia in peripheral blood of sarcoidosis patients correlates with severe disease manifestations. *PLoS One*. (2010) 5:e9088. doi: 10.1371/journal.pone.0009088

79. Northcott M, Gearing LJ, Bonin J, Koelmeyer R, Hoi A, Hertzog PJ, et al. Immunosuppressant exposure confounds gene expression analysis in systemic lupus erythematosus. *Front Immunol*. (2022) 13:964263. doi: 10.3389/fimmu.2022.964263

80. Karampitsakos T, Juan-Guardela BM, Tzouvelekis A, Herazo-Maya JD. Precision medicine advances in idiopathic pulmonary fibrosis. *EBioMedicine*. (2023) 95:104766. doi: 10.1016/j.ebiom.2023.104766

10.3389/fmed.2024.1381031

Glossary

ADAMDEC1	a disintegrin and metalloprotease like decysin 1
ACSF2	Acyl-CoA synthetase family member 2
ADA	adenosine deaminase
BALF	bronchoalveolar lavage fluid
CCL4	C-C motif chemokine ligand 4
CSF2	colony stimulating factor 2
CSTB	Cystatin B
CXCR6	C-X-C motif chemokine receptor 6
DEG	differentially expressed gene
EIF2	Eukaryoitc initation factor 2
FADS1	fatty acid desaturase 1
FC	fold change
FCGR1A	high-affinity gamma FC receptor I
GEO	genome expression omnibus
GBP	guanylate-binding protein
IFN	interferon
IL18BP	interleukin 18 binding protein
IPA	ingenuity pathway analysis
JAK	Janus kinase
ME1	malic enzyme 1
mTOR	mammalian target of rapamycin
MMP12	matrix metallopeptidase 12
MRAS	muscle RAS oncogene homolog
NOD2	nucleotide-binding oligomerization domain containing 2
RASGRP3	RAS guanyl releasing protein 3
RMA	robust microarray average
STAT1	signal transducer and activator of transcription 1
SNTB2	syntrophin beta 2
WARS1	Tryptophayl-tRNA synthetase 1
T2D	type 2 diabetes
TB	tuberculosis
Th1	T helper 1
TLR	toll-like receptor
UBE2L6	ubiquitin conjugating enzyme E2 L6
VAMP5	vesicle-associated membrane protein 5
WSB1	WD repeat and SOCS box-containing protein 1

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 19 June 2024
DOI 10.3389/fmed.2024.1406149

OPEN ACCESS

EDITED BY
Hari S. Sharma,
Erasmus Medical Center, Netherlands

REVIEWED BY
Anna Ermund,
University of Gothenburg, Sweden
Jian Liang,
Guangzhou University of Chinese Medicine, China

*CORRESPONDENCE
Yujiao Zheng
✉ luoyuorz@163.com
Xuemei Zhou
✉ zhouxmei0868@sina.com

†These authors share first authorship

RECEIVED 24 March 2024
ACCEPTED 07 June 2024
PUBLISHED 19 June 2024

CITATION
Hu F, Xiong L, Li Z, Li L, Wang L, Wang X, Zhou X and Zheng Y (2024) Deciphering the shared mechanisms of Gegen Qinlian Decoction in treating type 2 diabetes and ulcerative colitis via bioinformatics and machine learning.
Front. Med. 11:1406149.
doi: 10.3389/fmed.2024.1406149

Deciphering the shared mechanisms of Gegen Qinlian Decoction in treating type 2 diabetes and ulcerative colitis via bioinformatics and machine learning

Faquan Hu†, Liyuan Xiong†, Zhengpin Li, Lingxiu Li, Li Wang, Xinheng Wang, Xuemei Zhou* and Yujiao Zheng*

College of Traditional Chinese Medicine, Anhui University of Chinese Medicine, Hefei, China

Background: Although previous clinical studies and animal experiments have demonstrated the efficacy of Gegen Qinlian Decoction (GQD) in treating Type 2 Diabetes Mellitus (T2DM) and Ulcerative Colitis (UC), the underlying mechanisms of its therapeutic effects remain elusive.

Purpose: This study aims to investigate the shared pathogenic mechanisms between T2DM and UC and elucidate the mechanisms through which GQD modulates these diseases using bioinformatics approaches.

Methods: Data for this study were sourced from the Gene Expression Omnibus (GEO) database. Targets of GQD were identified using PharmMapper and SwissTargetPrediction, while targets associated with T2DM and UC were compiled from the DrugBank, GeneCards, Therapeutic Target Database (TTD), DisGeNET databases, and differentially expressed genes (DEGs). Our analysis encompassed six approaches: weighted gene co-expression network analysis (WGCNA), immune infiltration analysis, single-cell sequencing analysis, machine learning, DEG analysis, and network pharmacology.

Results: Through GO and KEGG analysis of weighted gene co-expression network analysis (WGCNA) modular genes and DEGs intersection, we found that the co-morbidity between T2DM and UC is primarily associated with immune-inflammatory pathways, including IL-17, TNF, chemokine, and toll-like receptor signaling pathways. Immune infiltration analysis supported these findings. Three distinct machine learning studies identified IGFBP3 as a biomarker for GQD in treating T2DM, while BACE2, EPHB4, and EPHA2 emerged as biomarkers for GQD in UC treatment. Network pharmacology revealed that GQD treatment for T2DM and UC mainly targets immune-inflammatory pathways like Toll-like receptor, IL-17, TNF, MAPK, and PI3K-Akt signaling pathways.

Conclusion: This study provides insights into the shared pathogenesis of T2DM and UC and clarifies the regulatory mechanisms of GQD on these conditions. It also proposes novel targets and therapeutic strategies for individuals suffering from T2DM and UC.

KEYWORDS

Gegen Qinlian Decoction, type 2 diabetes, ulcerative colitis, bioinformatics, network pharmacology, traditional Chinese medicine

1 Introduction

Emerging research posits that chronic tissue inflammation is a central player in the pathogenesis of Type 2 Diabetes Mellitus (T2DM), characterized by a state of low-grade inflammation (1). The disturbance in gut mucosal ecology in individuals with T2DM, combined with the active migration of intestinal flora to mesenteric adipose tissue (MAT) and the bloodstream, results in a continuous influx of inflammatory antigens (2). Ulcerative colitis (UC) is a chronic, nonspecific inflammatory condition characterized by extensive mucosal inflammation in the colon (3), typically arising from an imbalance between the gut flora and the immune system (4). Although T2DM and UC share common features, such as disruptions in gut microbiota and intestinal mucosa inflammation, the precise mechanisms underlying their co-occurrence remain unclear.

Currently, there is no cure for Type 2 diabetes. Despite the recent successful development of numerous antidiabetic drugs, single-target treatments are increasingly seen as inadequate due to individual variations, diverse pathogenesis, and issues related to drug and body resistance (5). Similarly, ulcerative colitis carries a heightened risk of adverse events, treatment resistance, and loss of response over time, highlighting the limitations of current therapies (6). Thus, multi-target drugs offer greater potential advantages over single-target drugs, underscoring the need to continually identify new targets to develop effective and safe therapies. Traditional Chinese medicine formulations are characterized by their multi-component approach, targeting multiple pathways and targets simultaneously.

The venerable Chinese herbal prescription, Gegen Qinlian Decoction (GQD), traces its origins to the era of the Eastern Han Dynasty. This formulation, consisting of four vital herbs—*Radix puerariae*, *Radix scutellariae, Rhizoma coptidis*, and *Glycyrrhizae Radix,* represents a traditional prescription deeply rooted in the principles of Traditional Chinese Medicine (TCM), specifically tailored for addressing intestinal damp-heat syndrome. A large-scale randomized controlled study (RCT) has demonstrated GQD's efficacy in significantly lowering HbA1c and fasting blood glucose (FBG) levels, offering relief in cases of T2DM (7). Animal experiments suggest that GQD may mitigate systemic and local inflammation by promoting the enrichment of butyrate-producing intestinal flora, thereby ameliorating clinical manifestations associated with T2DM (8). Meta-analyses have shown that GQD effectively alleviates symptoms in individuals with UC, resulting in decreased Ulcerative Colitis Endoscopic Index of Severity (UCEIS) scores and maintaining a low recurrence rate, all while exhibiting minimal adverse events (9). Furthermore, additional animal studies have elucidated the mechanisms underlying GQD's therapeutic effects in alleviating ulcerative colitis, including the reduction of inflammation and oxidative stress, inhibition of the IL-6/JAK2/STAT3 signaling pathway, restoration of the balance between Treg and Th17 cells in colonic tissues, and enhancement of intestinal barrier function (10, 11).

As bioinformatics advances and the widespread adoption of gene chips continue, their integration into the biomedical domain has become indispensable. The analysis of microarray data emerges as a transformative tool, offering fresh insights into the shared etiological underpinnings of both T2DM and UC. In this investigation, a comprehensive strategy merging bioinformatics and machine learning was employed, drawing upon datasets from the GEO database to unravel the intertwined comorbid mechanisms associated with T2DM and UC. Furthermore, our study delved into network pharmacology, shedding light on the intricate mechanisms governing the utilization of GQD across diverse diseases sharing a common treatment modality (Figure 1).

2 Methods

2.1 Datasets

We queried the GEO database[1] to retrieve gene expression profiles of individuals diagnosed with Type 2 Diabetes Mellitus (T2DM) and Ulcerative Colitis (UC), using search terms such as "type 2 diabetes" and "ulcerative colitis." For the subsequent phase of our investigation, we selected the following GEO datasets: GSE3365, GSE48958, GSE75214, GSE231993, GSE20966, GSE25724, GSE29221, and GSE220939 (Table 1).

2.2 Construction of weighted gene co-expression networks

We leveraged Weighted Gene Co-Expression Network Analysis (WGCNA) to pinpoint clusters of closely linked genes. Applying WGCNA, we scrutinized the differential genes within the GEO datasets GSE20966 and GSE75214, unveiling co-expression modules and pivotal genes intricately linked to both T2DM and UC. In the dataset about T2DM, the outlier GSM524165 was omitted, and subsequent parameter selection involved $R2 = 0.85$ and $\beta = 9$. In the case of the UC dataset, parameters $R2 = 0.85$ and $\beta = 8$ were applied. The matrices undergo sequential transformation and derivation, ultimately yielding the Topological Overlap Matrix (TOM). Hierarchical clustering, with a minimum module size set at 30, is employed to identify modules. Subsequently, feature genes are computed, and hierarchical clustering is applied to the modules. Ultimately, we identified the common genes within the top three significantly ranked modules associated with both T2DM and UC. Subsequently, we subjected these intersecting genes to a thorough GO enrichment analysis.

2.3 Acquisition of differentially expressed genes (DEGs)

The study analyzed the DEGs using the GSE25724 and GSE48958 datasets. The initial gene expression data undergoes thorough cleaning, with the Robust Multi-array Average (RMA) method employed to equalize sample differences. Subsequently, gene name normalization is necessary to eliminate empty columns and duplicate values, ensuring the acquisition of normalized expression data. Our analysis utilized the R limma package, employing a filtering criterion for DEGs set at $|\log FC| \geq 1$ and a p-value threshold of <0.05. We proceeded to identify the common genes between the two sets of DEGs and subjected this intersection to KEGG pathway enrichment analysis.

1 https://www.ncbi.nlm.nih.gov/geo/

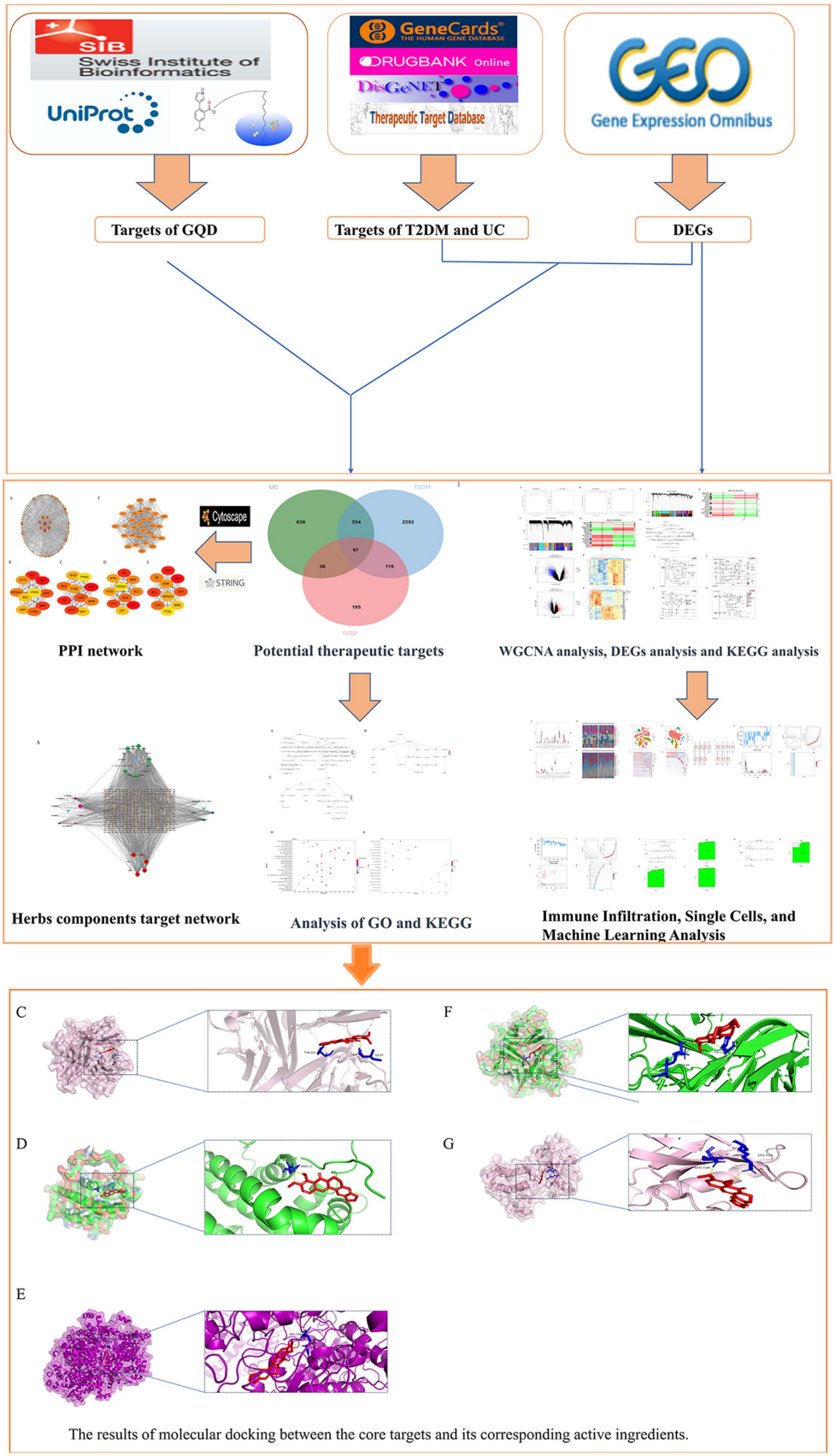

FIGURE 1
Workflow diagram illustrating the research strategy, encompassing five main components: database preparation, exploration of co-morbidity mechanisms in T2DM and UC, biomarker prediction for GQD treatment, and network pharmacology along with molecular docking analyses.

TABLE 1 Data sets and their characteristics.

Dataset	Database	Platform	Sample
GSE3365	GEO	GPL96	26 cases of UC and 42 controls
GSE48958	GEO	GPL6244	13 cases of UC and 8 controls
GSE75214	GEO	GPL6244	97 cases of UC and 11 controls
GSE231993	GEO	GPL18573	4 cases of UC and 4 controls
GSE20966	GEO	GPL1352	10 cases of T2DM and 10 controls
GSE25724	GEO	GPL96	6 cases of T2DM and 7 controls
GSE29221	GEO	GPL6947	3 cases of T2DM and 3 controls
GSE220939	GEO	GPL11154	16 cases of T2DM and 6 controls

2.4 Immune infiltration analysis

To precisely evaluate the composition of immune cells in UC compared to T2DM, we conducted calculations utilizing the CIBERSORT algorithm. We utilized the authentic CIBERSORT gene signature file that delineates 22 immune cell subtypes for analyzing the T2DM versus UC dataset. A significance level of $p < 0.05$ denotes a meaningful difference. The datasets utilized were GSE3365 and GSE29221.

2.5 Network pharmacology analysis

2.5.1 Acquisition of GQD active ingredients and target proteins

UPLC-Q-TOF/MS analysis revealed a comprehensive profile of 130 active chemical components, out of which 37 components met the criteria of Oral Bioavailability (OB) $\geq$ 30% and Drug-Likeness (DL) $\geq$ 0.18 (12). Retrieve the 2D structure of the compound from the PubChem database.[2] Chemical targets were then predicted using PharmMapper and SwissTargetPrediction (13).

2.5.2 Predictive hub genes for GQD treatment

We employed three machine learning algorithms to systematically screen biomarker targets associated with GQDs for the treatment of T2DM and UC, respectively. LASSO logistic regression selectively assigns coefficients to significant variables by imposing an L1 penalty to remove less relevant ones, thus optimizing the classification model. SVM-RFE analysis, a supervised learning approach, identifies key genes by iteratively eliminating feature vectors derived from SVM. Random forest analysis, rooted in decision trees, evaluates variable importance by scoring each variable (14). The seed was set to 123 for consistency in the analysis. The targets of GQD were compared with the DEGs of T2DM and modules from WGCNA that exhibited a positive correlation with T2DM at $p < 0.05$. The overlapping elements from these three sets were analyzed, and a parallel approach was applied to UC. To enhance diagnostic accuracy and prediction capabilities, diagnostic nomograms were created utilizing hub genes as the foundation.

2.5.3 Common targets of GQD for the treatment of T2DM and UC

Disease-associated targets were queried in the DrugBank,[3] GeneCards,[4] TTD,[5] and DisGeNET[6] databases using the keywords "Type 2 diabetes" and "Ulcerative colitis" (15, 16). Targets occurring in at least two instances across DrugBank, GeneCards, TTD, DisGeNET, and DEGs datasets are identified as disease targets. The ultimate selection comprises overlapping genes from the targets associated with T2DM, UC, and GQD, serving as potential common targets for GQD treatment of both T2DM and UC.

2.5.4 Protein–protein interaction (PPI) network

TSV files of PPI were obtained by uploading potential therapeutic target genes into the STRING database and constructing networks in Cytoscape 3.9.1. We employed the cytoHubba plugin and computed parameters such as Degree Centrality (DC), Betweenness Centrality (BC), Closeness Centrality (CC), and Maximal Clique Centrality (MCC). Central genes were identified through two methods: firstly, by calculating the top ten targets ranked by each of the four parameters and then determining the overlap among them; secondly, by utilizing the MCODE plugin for cluster analysis, generating a highly connected sub-network.

2.5.5 The analysis of GO and KEGG

To comprehend the shared physiological mechanisms of GQD for both T2DM and UC, we conducted GO and KEGG enrichment analyses of therapeutic targets using the R language. Significance thresholds were set at $p \leq 0.05$ and $q \leq 0.01$, and the outcomes were visually presented for comprehensive understanding.

2.5.6 Molecular docking

Key genes were selected from the PPI sub-network, and core chemicals were screened from the drug target network map for subsequent molecular docking analysis. The structures of the core active ingredients were sourced from online databases. Protein stereo structures were also retrieved from databases and subjected to dehydration, hydrogenation, and removal of impurity ligands using PYMOL software. Following that, molecular docking analysis was conducted using Autodock, and the results were graphically presented.

2.6 Single-cell sequencing analysis

Seurat objects were initialized by loading gene expression data from the GEO database via the read10X function. Cell curation

2 https://pubchem.ncbi.nlm.nih.gov/

3 https://go.drugbank.com/

4 https://www.genecards.org/

5 https://idrblab.net/ttd/

6 https://www.disgenet.org/

involved preserving those with a gene count between >200 and < 10,000, while filtering out those with mitochondrial and ribosomal gene proportions exceeding 20%. Following this, standardization and normalization procedures were applied for data uniformity. Spatial relationships between clusters were evaluated using the tSNE method, and subsequent cluster annotations were conducted using the celldex package. The reclassification of cell subpopulations was accomplished through the singleR annotation tool, concurrent with referencing the Thermofisher website to identify genes characterizing different immune cell types. Following observation of the expression patterns of these genes within the clustering results, a manual classification of immune cell classes was performed for annotation purposes. Finally, we have successfully visualized the expression patterns of GQD targets at the single-cell level and elucidated the distribution of the seven core targets.

2.7 Statistical analysis

The R packages utilized in this study include WGCNA, GEOquery, reshape2, ggfortify, limma, pheatmap, ggplot2, org.Hs.eg.db, pathview, topGO, and Rgraphviz.

3 Results

3.1 Construction of WGCNA network

WGCNA analysis revealed that in T2DM, higher independence and greater biological significance were observed at $\beta = 9$. Similarly, for UC, the optimal fit was achieved at $\beta = 8$ (Figures 2A,B). When reaching the optimal fit, a hierarchical clustering dendrogram was generated, allowing the classification of similar gene expressions into distinct modules. The expression profiles within each module were then summarized using modular eigengenes (MEs), and correlations between MEs and clinical features were subsequently calculated. In T2DM, a total of 24 modules were identified, with each color denoting a distinct module. Heat maps illustrating module-trait relationships were constructed based on Spearman correlation coefficients to evaluate the association between each module and the disease (Figures 2C,D, and Supplementary Table S1). In the heat map depicting module-trait relationships, cyan signifies a negative correlation, red indicates a positive correlation, and white denotes no correlation. Six modules, namely pale turquoise, turquoise, white, dark gray, pink, and violet, exhibit substantial positive correlations with T2DM, thus qualifying them as T2DM positively correlated modules (pale turquoise module: $r = 0.58$, $p = 0.009$; turquoise module: $r = 0.53$, $p = 0.02$; white module: $r = 0.59$, $p = 0.008$, dark gray module: $r = 0.52$, $p = 0.02$, pink module: $r = 0.67$, $p = 0.002$, violet module: $r = 0.66$, $p = 0.002$). Likewise, in UC, 22 modules were identified, among which lightsteelblue1, black, mediumpurple3, green, darkolivegreen, orange4, plum1, and thistle1 exhibited positive correlations with UC (lightsteelblue1 module: $r = 0.25$, $p = 0.009$; black module: $r = 0.40$, $p = 2e\text{-}05$; mediumpurple3 module: $r = 0.24$, $p = 0.01$, green module: $r = 0.24$, $p = 0.01$, darkolivegreen module: $r = 0.24$, $p = 0.01$, orange4 module: $r = 0.27$, $p = 0.005$, plum1 module: $r = 0.46$, $p = 5e\text{-}07$, thistle1 module: $r = 0.49$, $p = 7e\text{-}08$) (Figures 2E,F, and Supplementary Table S2). Biological process analysis indicates that the interacting genes are primarily engaged in regulating the immune system and activating immune cells, among other functions (Figure 2G).

3.2 Identification of DEGs

With the limma package, we identified 70 genes exhibiting high expression levels and 1,171 genes showing low expression levels associated with T2DM. Similarly, in UC, 236 genes were found to be highly expressed, while 168 genes exhibited low expression levels (Figures 3A–D, and Supplementary Tables S3, S4). The KEGG enrichment analysis of overlapping genes predominantly focused on pathways involving IL-17, TNF, Chemokine, and Toll-like receptor signaling pathways, indicating that the shared mechanism between T2DM and UC may be linked to immunity and inflammation (Figures 3E–H).

3.3 Immune infiltration analysis

The findings indicate a close association between the pathogenesis of T2DM and UC with the immune system (Figures 4A–D). In the T2DM group, there were observed differences in both the T cell population and resting NK cells compared to the normal group ($p < 0.05$). Compared to the normal group, significant differences ($p < 0.05$) were observed in Plasma cells, T cells regulatory (Tregs), NK cells resting, Neutrophils, NK cells activated, Monocytes, and Dendritic cells activated in UC.

3.4 Predictive hub genes for GQD treatment

Potential targets for 37 core chemicals were identified through PharmMapper and SwissTargetPrediction. Afterward, the obtained results underwent the removal of identical values, culminating in a total of 444 targets (Supplementary Tables S5, S6). When employing the SVM-RFE method, we conducted 10-fold cross-validation. In the identification of core targets of GQD for treating T2DM, SVM-RFE achieved the highest accuracy of 95% with 73 features. LASSO identified 10 core targets, while RF identified one target. The intersection of core targets identified by the three models resulted in one core therapeutic target, IGFBP3 (Figures 5A–D). Applying a similar approach to identify core targets for GQD treatment of UC, SVM-RFE identified 9 core targets, LASSO also identified 9 core targets, and RF identified 37 core targets. The final intersection yielded 3 core targets: BACE2, EPHB4, and EPHA2 (Figures 5E–H). Nomograms and ROC curves depict the robust diagnostic potential of pivotal genes for both T2DM and UC (Figures 5I–N).

3.5 Common targets of GQD for the treatment of T2DM and UC

From the DrugBank database, we retrieved 150 targets associated with T2DM and 66 targets associated with UC. Additionally, GeneCard yielded 17,916 targets for T2DM and 5,282 targets for UC, while DisGeNET provided 2,359 targets for T2DM and 1,458 targets

10.3389/fmed.2024.1406149

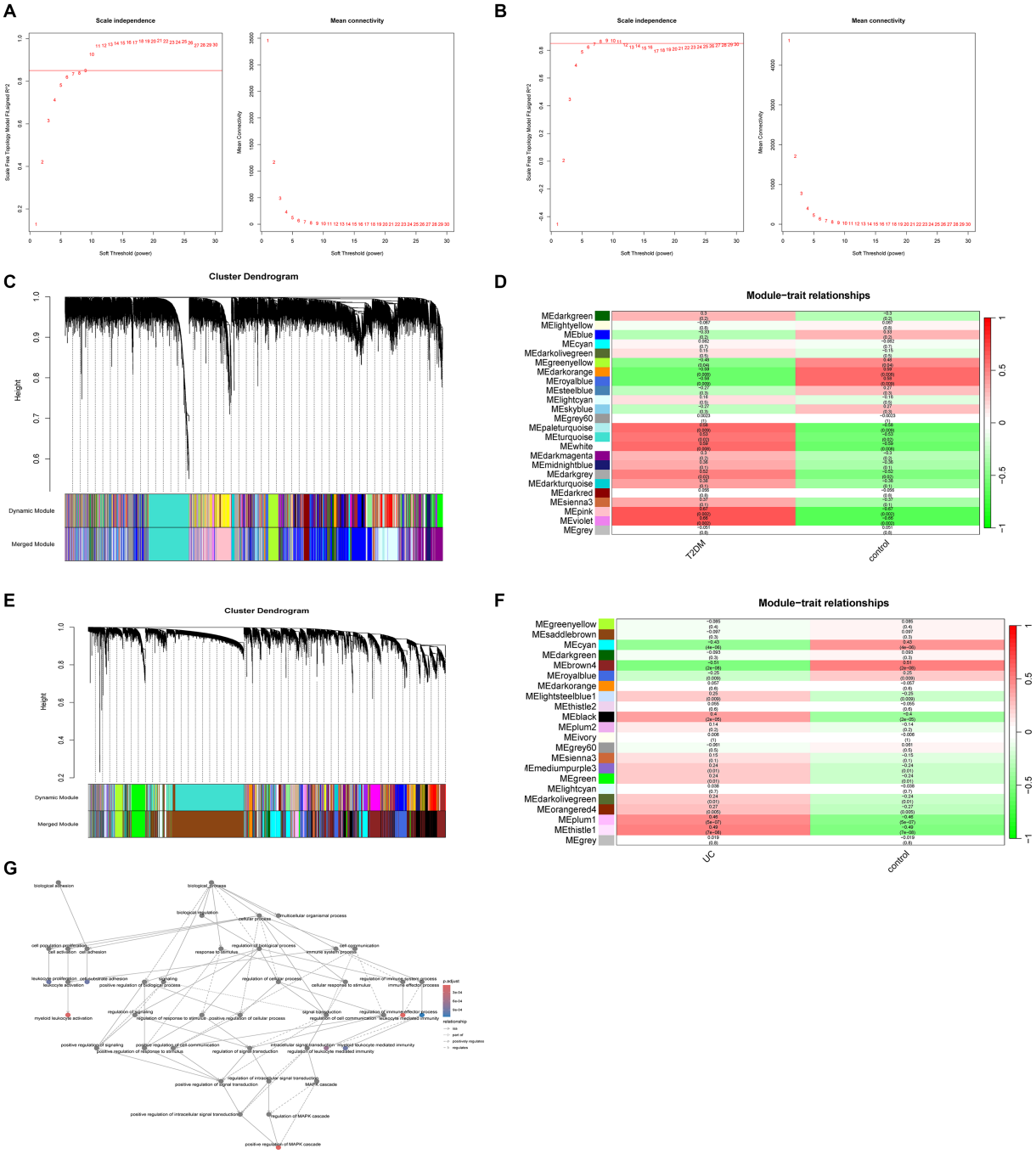

FIGURE 2
Weighted gene co-expression networks. **(A)** Scale independence and average connectivity in GSE20966. **(B)** Scale independence and average connectivity in GSE75214. **(C)** Different modules obtained from GSE20966 are displayed in various colors, aggregating genes of high relevance within each module. **(D)** Correlation analysis between modules and T2DM. **(E)** Different modules obtained from GSE75214 are displayed in various colors, aggregating genes of high relevance within each module. **(F)** Correlation analysis between modules and UC. **(G)** Biological process analysis of T2DM and UC module intercourse genes.

for UC. Furthermore, TTD identified 88 targets for T2DM and 48 targets for UC. After performing the process of removing duplicates, taking targets that appear at least twice in five databases—DrugBank, GeneCards, TTD, DisGeNET, and DEGs—and then intersecting them with the GQD target as a common target for GQD treatment of T2DM and UC. Subsequently, a total of 97 potential common targets were finalized (Supplementary Table S7).

3.6 PPI network

For the identification of common core targets of GQD for the treatment of T2DM and UC, we conducted an in-depth analysis of 97 targets using Cytoscape (Figure 6A). By considering the genes identified in the overlapping sections of the four algorithms, including DC, BC, CC, and NCC, we ultimately identified seven

10.3389/fmed.2024.1406149

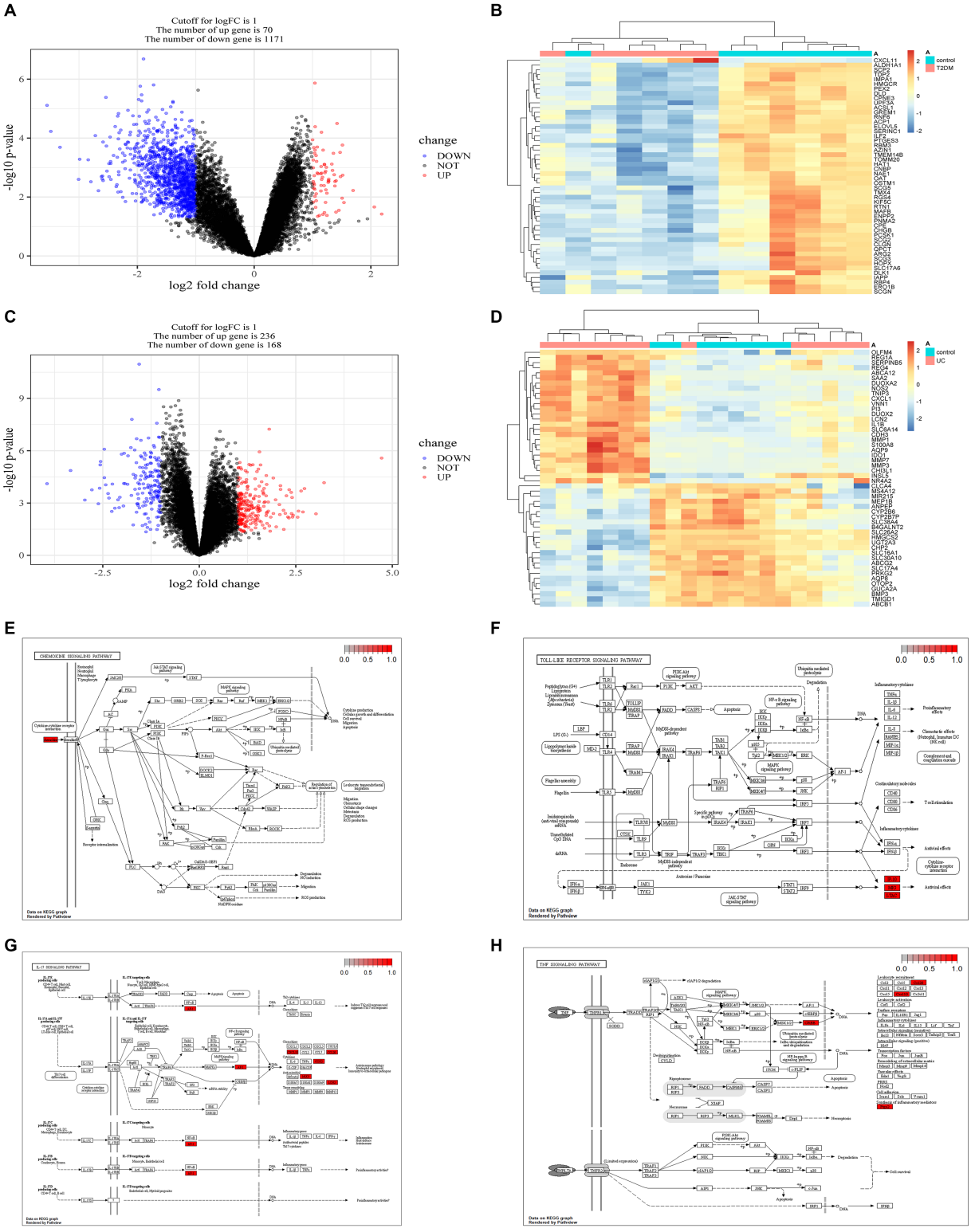

FIGURE 3
Acquisition of DEGs in T2DM and UC and KEGG enrichment analysis of genes intersecting both DEGs. **(A)** Volcano plot depicting the DEGs associated with T2DM (GSE25724). **(B)** Heatmap illustrating the DEGs associated with T2DM (GSE25724). **(C)** Volcano plot showing the DEGs associated with UC (GSE48958). **(D)** Heatmap displaying the DEGs associated with UC (GSE48958). **(E)** Chemokine signaling pathway. **(F)** Toll-like receptor signaling pathway. **(G)** IL-17 signaling pathway. **(H)** TNF signaling pathway. DEGs, Differentially Expressed Genes.

core target proteins of GQD for managing T2DM concomitant with UC (Figures 6B–E). The application of the MCODE plugin in cluster analysis generated highly connected sub-networks. The highest-scoring network comprised a total of 30 targets, among which the seven core targets identified previously were also encompassed (Figure 6F).

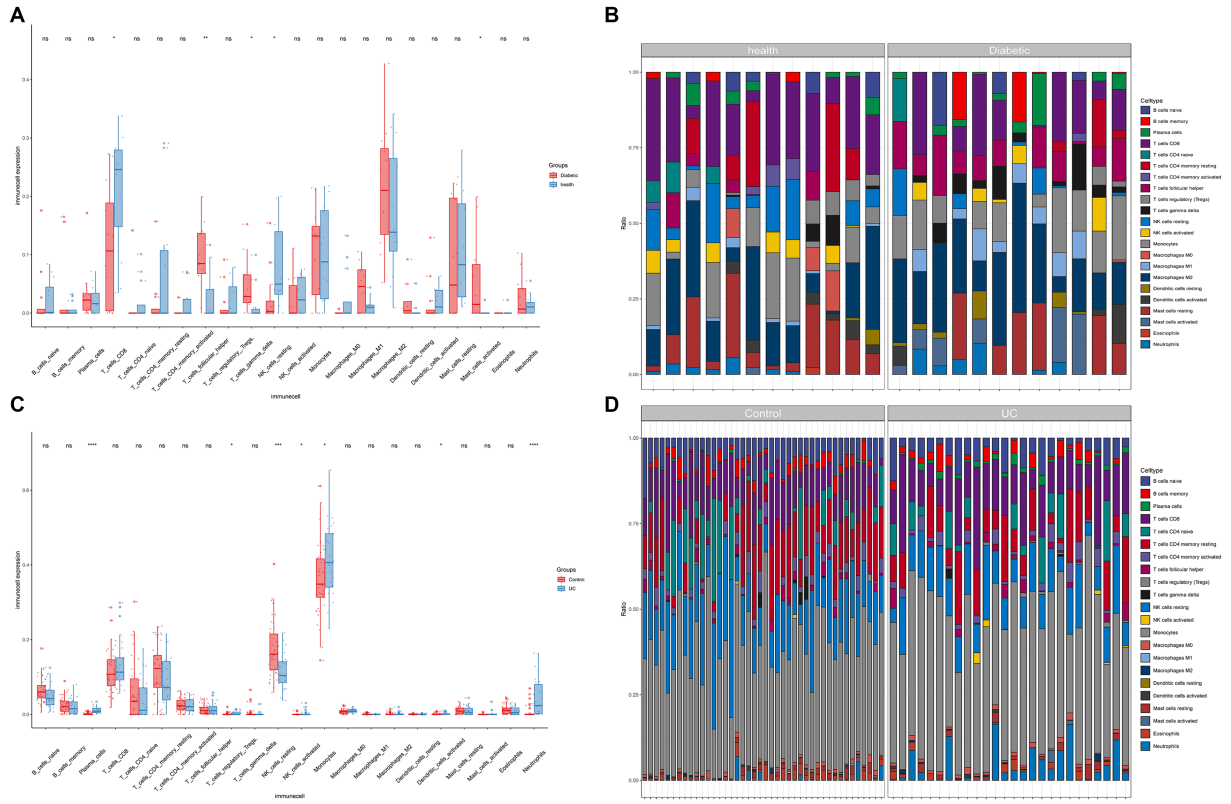

FIGURE 4
Immune infiltration analysis. **(A)** Boxplots for T2DM immune infiltration analysis. **(B)** Bar graph for T2DM immune infiltration analysis. **(C)** Boxplots for UC immune infiltration analysis. **(D)** Bar graph for UC immune infiltration analysis (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

3.7 GO, KEGG enrichment analysis

GO enrichment analysis revealed significant enrichment (p-value ≤0.05, q-value <0.01) in 1784 biological processes (BP), 105 molecular functions (MF), and 50 cellular components (CC) (Supplementary Table S8). Among these, BP mainly encompasses immune-inflammatory responses, oxidative stress, etc.; CC mainly involves membrane raft, and MF mainly includes tyrosine kinase activity, insulin receptor substrate binding, phosphatase binding, heme binding, etc. (Figures 7A–C). The KEGG enrichment analysis of potential therapeutic targets for GQD revealed pathways related to immunoinflammation, among others, indicating a broader spectrum of pathways beyond just immunoinflammatory regulation (Supplementary Table S9). We visualize the top 30 results, as well as results specifically related to immunization (Figures 7D,E).

3.8 Molecular docking

To validate our findings, we assessed the interactions between the identified active drugs and targets through molecular docking analysis. Our PPI network analysis revealed 7 core targets (AKT1, BCL2, EGFR, ESR1, PTGS2, STAT3, and TNF). Additionally, through drug-component-target network mapping, we identified 7 core chemical components of GQD: Berlambine, Palmatine, Moslosooflavone, Quercetin, Moupinamide, Panicolin, and Baicalein. Before docking, we transformed the core components and targets into the required format (Figure 8A and Table 2). The outcomes of the 49 docking combinations are represented through heatmaps and tables, highlighting the top 5 combinations exhibiting the strongest binding energy, which are then visualized in greater detail (Figures 8B–G and Table 3). In our molecular docking findings, it's evident that Berlambine and Palmatine exhibit the strongest binding affinity to the core target.

3.9 Single-cell sequencing analysis

The tSNE algorithm was utilized to cluster cells based on the GSE220939 and GSE231993 datasets, and subsequent labeling of each cluster was performed using SingleR. All cells from patients with T2DM were classified into seven groups: Epithelial Cells, Endothelial cells, Hepatocytes, Smooth Muscle Cells, Monocytes, B cells, and Natural Killer (NK) Cells (Figure 9A). Similarly, in UC, all cells were classified into eight classes: B cells, T cells, Epithelial cells, Monocytes, Fibroblasts, Endothelial cells, CMP, and Neurons (Figure 9B). The distribution of drug targets revealed that in T2DM, the primary cellular cluster targeted by GQD was Epithelial Cells, with subsequent impact on Hepatocytes (Figure 9C). In ulcerative colitis, GQD predominantly targets B cells, with subsequent involvement of T cells (Figure 9D). The seven core targets exhibit a broad distribution across various cell clusters (Figures 9E,F).

10.3389/fmed.2024.1406149

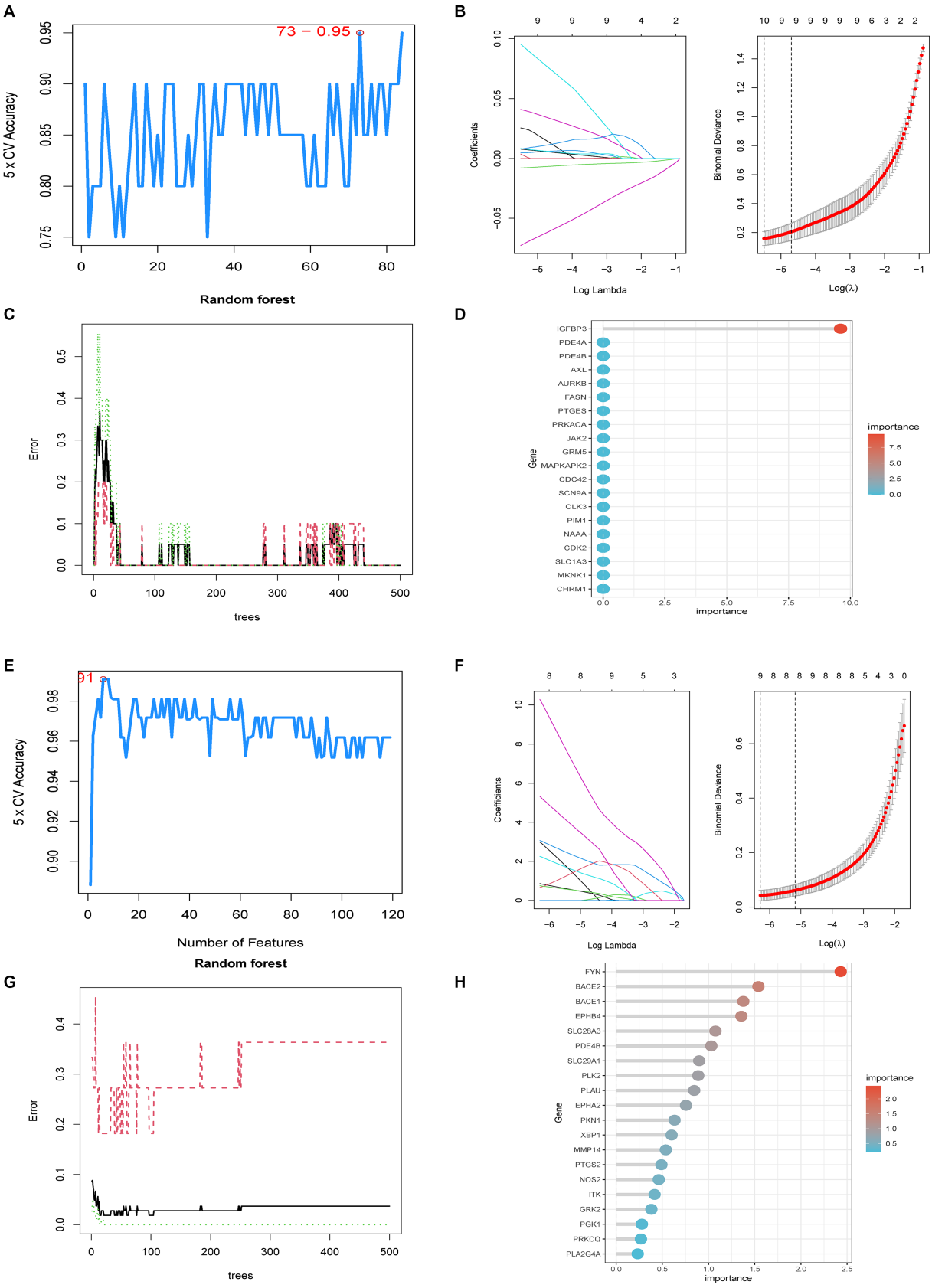

FIGURE 5 (Continued)

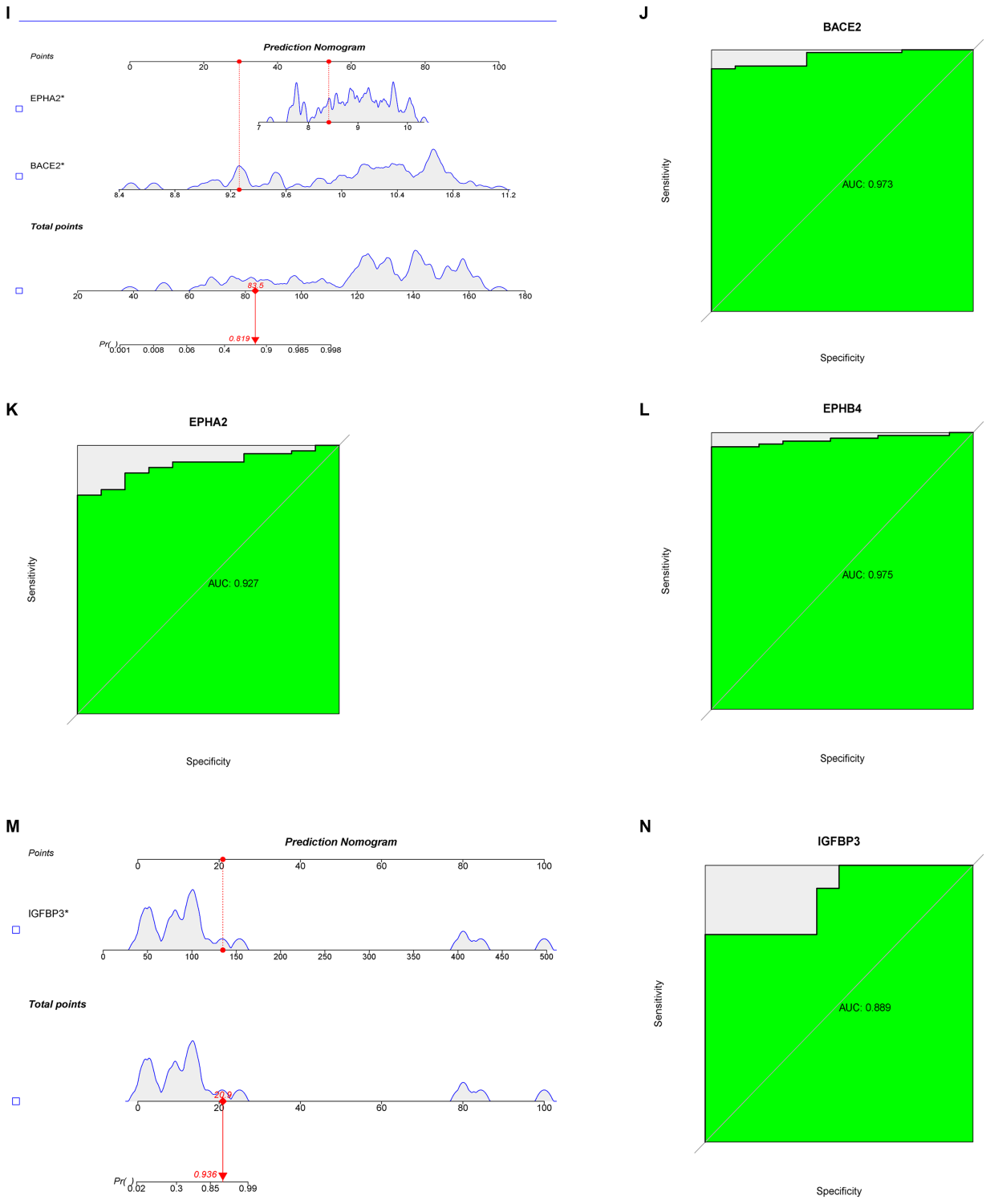

FIGURE 5
Predictive Biomarkers for GQD Treatment. **(A)** SVM-REF analysis of T2DM. **(B)** LASSO analysis of T2DM. **(C,D)** Random Forest analysis of T2DM. **(E)** SVM-REF analysis of UC. **(F)** LASSO analysis of UC. **(G,H)** Random Forest analysis of UC. **(I)** Nomograms of UC marker genes. **(J–L)** ROC curves for UC marker genes. **(M)** Nomograms of T2DM marker genes. **(N)** ROC curves for T2DM marker genes.

4 Discussion

An increasing body of research is corroborating the association between UC and T2DM. It has been demonstrated that diabetes is the most prevalent co-morbidity of UC (17). A population-based cohort study reveals a significantly heightened risk of T2DM among individuals with UC (18). The association between diabetes and UC holds significant implications across epidemiology, etiology, clinical

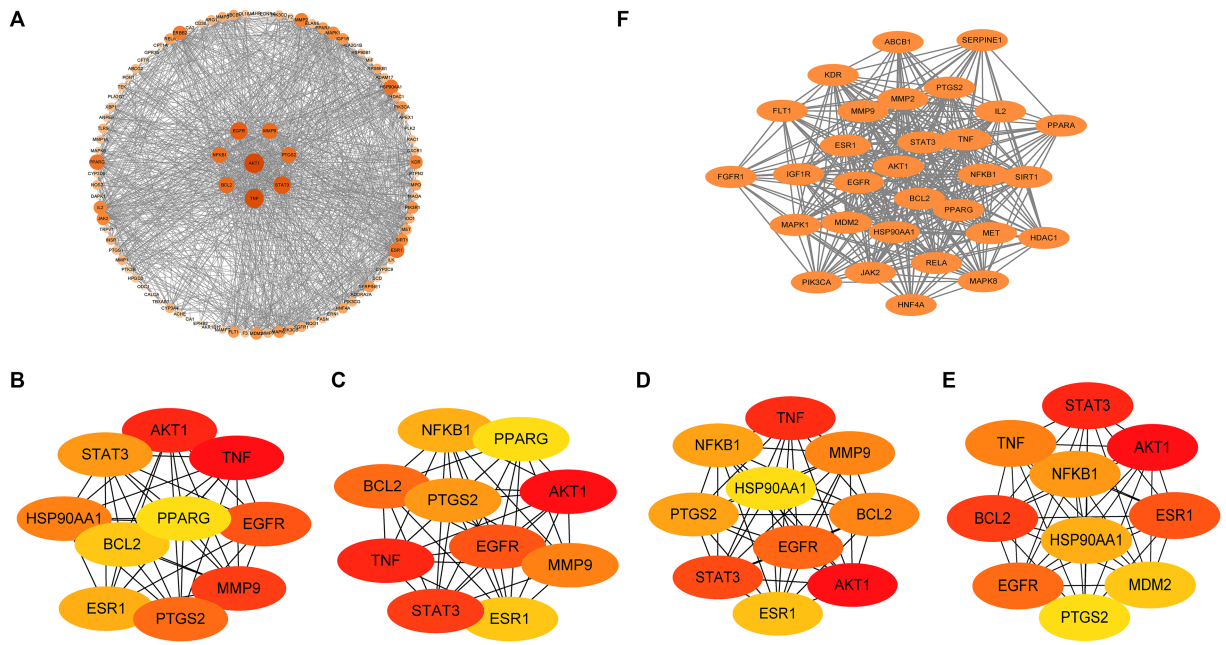

FIGURE 6
Protein-Protein interaction (PPI) network. **(A)** Analysis results of PPI network. **(B)** Betweenness centrality. **(C)** Closeness centrality. **(D)** Degree centrality. **(E)** Neighborhood Component Analysis. **(F)** MCODE plugin cluster analysis.

practice, and therapeutic strategies, signaling profound implications for research and patient care alike (19). Hence, investigating the mechanisms underlying the co-occurrence of UC and T2DM holds clinical significance, aiding in early disease detection and timely intervention. In this study, we conducted analyses on four microarray datasets related to UC and T2DM using diverse bioinformatics approaches. Based on our predictions, inflammatory and immune processes, along with Immunoinflammatory signaling pathways such as IL-17, TNF, chemokine, and Toll-like receptor, may represent potential mechanisms underlying the co-morbidity of T2DM with UC. Following subsequent network pharmacological analyses, pivotal targets of GQD for the concurrent treatment of T2DM and UC were identified, including AKT1, BCL2, EGFR, ESR1, PTGS2, STAT3, and TNF. Notably, GQD predominantly acts on immuno-inflammatory pathways, such as Toll-like receptors, IL-17, TNF, MAPK, and the PI3K-Akt signaling pathways, in the simultaneous treatment of T2DM and UC.

In recent years, the roles of intestinal flora, inflammation, and immune regulation in the pathogenesis of T2DM and UC have attracted increasing attention (20, 21). Single-cell sequencing coupled with immune infiltration analysis underscored the pivotal role of immune cells in driving the pathogenesis of both UC and T2DM. The intestinal mucosal immune system, comprising lymph nodes, lamina propria, and epithelial cells, serves as a vital barrier safeguarding intestinal integrity. The symbiotic relationship between the microbiome and the intestinal immune system is crucial for preserving mucosal homeostasis (22). However, deficiencies and dysbiosis in the intestinal flora can result in significant impairments to the intestinal mucosal immune system, precipitating the onset of T2DM alongside UC (23, 24). UC's development involves various factors within the gut microbiome, immune system dysfunctions, and compromised intestinal barriers, resulting in abnormal immune reactions to typical gut bacteria (25). UC is characterized by an imbalance between intestinal effector T cells and mucosal Treg, with effector T cells being overly active and Treg cells not expanding sufficiently (26). Balancing the population of Th17 and Treg cells in the intestines of mice markedly improves symptoms and reduces pathological damage in ulcerative colitis (27). In T2DM, disruptions in intestinal immunity and barrier function, alongside alterations in gut microbiota, foster heightened intestinal permeability. Consequently, intestinal bacterial components infiltrate circulation, fueling both local and systemic chronic inflammation, ultimately contributing to insulin resistance (28, 29). Dendritic cells, functioning as autocrine or paracrine modulators, synthesize and release classical neurotransmitters crucial for maintaining intestinal immune balance. Their abundance is markedly elevated in the gut of patients with UC and T2DM (30, 31). In ulcerative colitis, the usual equilibrium of intestinal B-cell reactions is disturbed, resulting in a notable decrease in regulatory B cells (32, 33). In parallel, B cells modulate Th17 proliferation and the production of pro-inflammatory factors in the intestines of T2DM patients (34). Research indicates that managing macrophage metabolism and polarization can alleviate symptoms in DSS-induced UC mice, hinting at the potential of targeting macrophage polarization to restore immune balance as a promising UC treatment strategy (35). In individuals with T2DM, there is a reduction in the quantity of anti-inflammatory T-cell subsets, such as regulatory T-cells (Treg), M2-like macrophages, and IgM-producing B-1 cells, alongside an elevation in the number and/or ratio of inflammatory effector T-cells (36). Individuals diagnosed with T2DM often exhibit irregularities in the frequency and functionality of B cells, potentially resulting in heightened inflammatory reactions and reduced insulin sensitivity. Moreover, the antibodies generated by B cells are pivotal in the progression of T2DM, notably contributing to neuroinflammation and cognitive deterioration (37, 38). In individuals with T2DM,

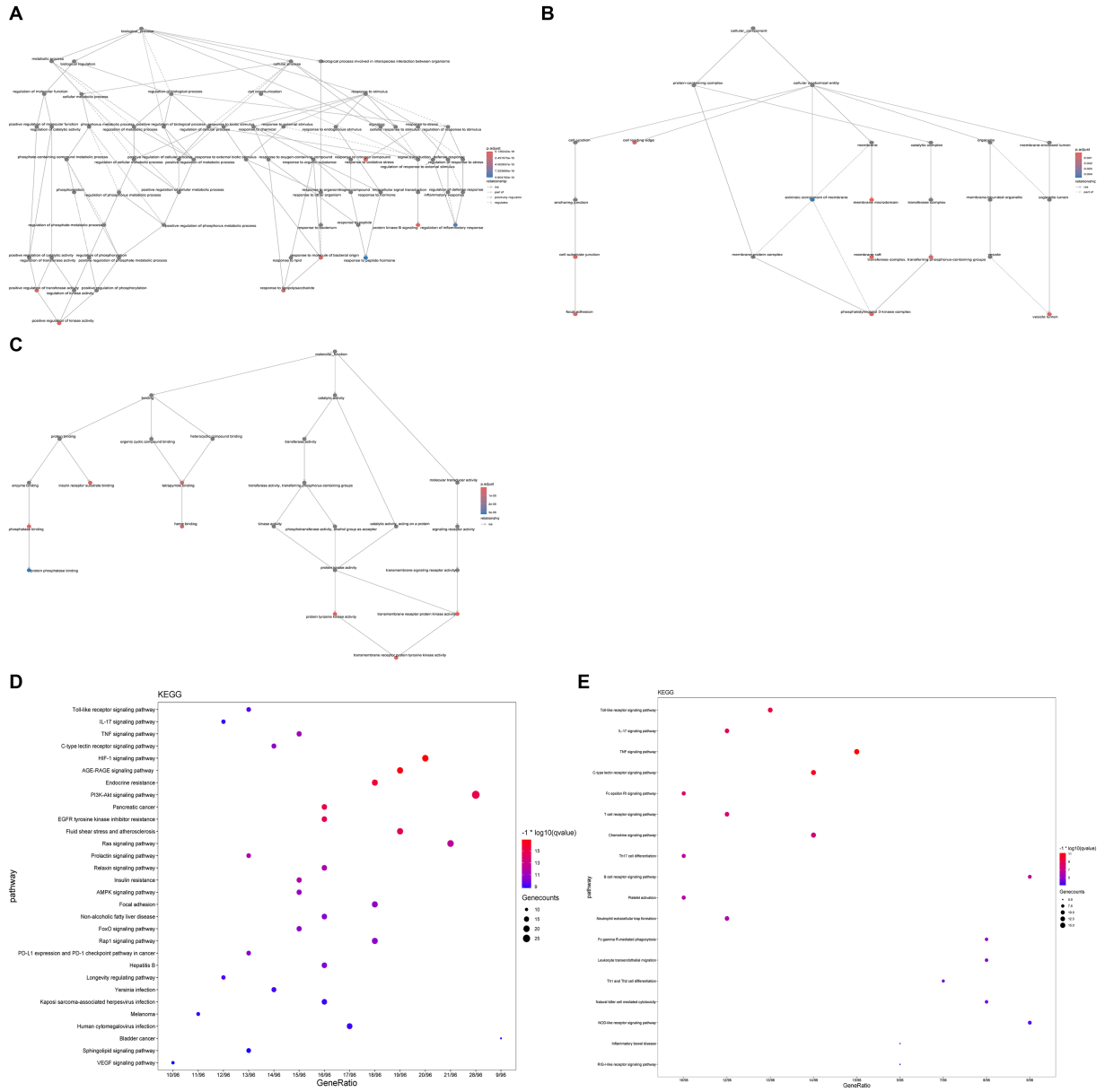

FIGURE 7
GO, KEGG enrichment analysis. **(A)** Biological process. **(B)** Cellular composition. **(C)** Molecular function. **(D)** Bubble plots of the first 30 pathways analyzed by KEGG enrichment. **(E)** KEGG enrichment analysis of immune-related pathways.

dendritic cells are implicated in vascular dysfunction. Research indicates an elevated accumulation of dendritic cells in the perivascular adipose tissue of diabetic mice, which consequently compromises their anticonstrictive and vasodilatory functions (39). Likewise, macrophages emerge as the primary immune cell driving inflammation within pancreatic islets in T2DM, posing a threat to the insulin-secreting function of β-cells through multiple mechanisms (40). In conclusion, the elevated permeability of intestinal mucosa caused by disturbances in intestinal flora and impairment of the intestinal mucosal immune system contributes to the onset of systemic chronic inflammatory responses, a shared mechanism underlying the development of UC and T2DM (41).

Ulcerative colitis manifests as recurring mucosal inflammation with periods of remission, necessitating treatment to induce and sustain remission (42). Concurrently, the incidence of T2DM is on the rise, contributing to escalating rates of disability and mortality, thereby compounding the burden on families (43). Thus, there is an imperative to discover additional routine serum biomarkers for the early diagnosis and treatment of T2DM and UC. Three distinct machine learning studies identified IGFBP3 as a biomarker for GQD in treating T2DM, while BACE2, EPHB4, and EPHA2 emerged as biomarkers for GQD in UC treatment. IGFBP3 interacts with cellular proteins involved in glucose metabolism regulation, consequently inducing insulin resistance and diminishing glucose uptake in adipose tissue (44). For every one-unit rise in genetically determined IGFBP3 levels, there's a 26 percent higher likelihood of developing T2DM (45). The degradation of pancreatic β-cells is a pivotal aspect of T2DM, and IGFBP3 signaling contributes to this decline in β-cell function and

10.3389/fmed.2024.1406149

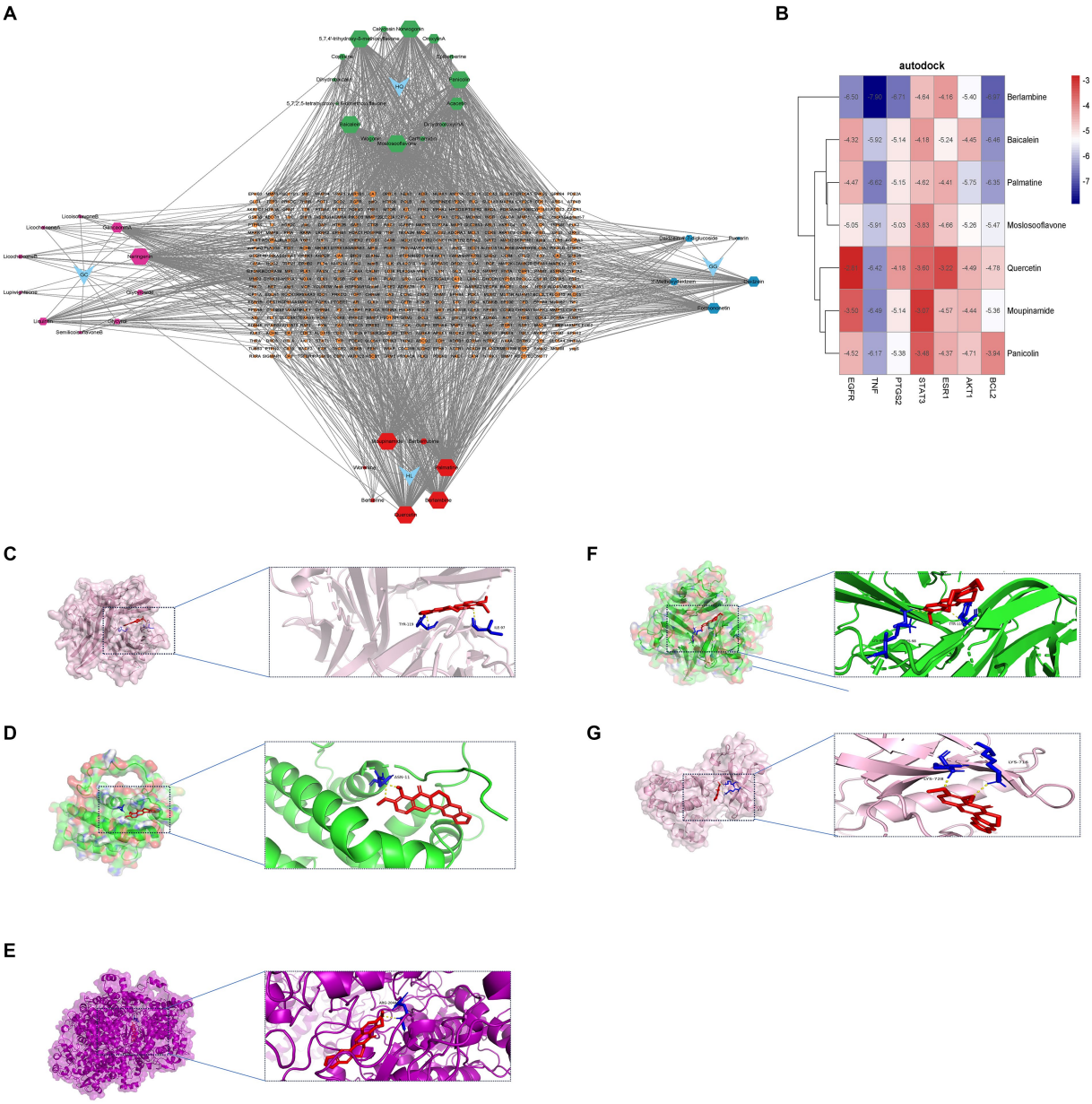

FIGURE 8
Molecular docking results. **(A)** Drug-constituent-target network diagram (the larger the value of degree in the diagram, the larger the node). **(B)** Heat map of molecular docking (kcal/mol). Berlambine – TNF **(C)**, Berlambine – BCL2 **(D)**, Berlambine – PTGS2 **(E)**, Palmatine – TNF **(F)**, Berlambine – EGFR **(G)**.

TABLE 2 Proteins and chemicals information.

Molecule name	PubChem ID	Target	PDB ID
Berlambine	11066	EGFR	2ITV
Palmatine	19009	TNF	7KP9
Moslosooflavone	188316	PTGS2	1PXX
Quercetin	5280343	STAT3	6NJS
Moupinamide	5280537	ESR1	4XI3
Panicolin	5320399	AKT1	5AAR
Baicalein	5281605	BCL2	1G5M

viability. Suppressing IGFBP3 activity can protect β-cells, potentially delaying or preventing the onset of diabetes, making it a promising therapeutic avenue for diabetes treatment (46). BACE2, a protease regulated by the JAK2/STAT5 signaling pathway, emerges as a pivotal contributor to UC development (47). The activity of IL-1R2, linked to ulcerative colitis, is influenced by the BACE2 gene. Therefore, BACE2 assumes a significant role in the pathogenesis of UC (48). The EphB/ephrin-B system has become a promising focus for tackling gut inflammatory diseases. Suppressing this system seems to provide a therapeutic benefit by regulating immune responses (49). Eph/ephrin proteins are implicated in numerous chronic inflammatory conditions. Targeting EPHB4 to disrupt EphB/ephrin B signaling

TABLE 3 Binding energy.

	EGFR	TNF	PTGS2	STAT3	ESR1	AKT1	BCL2
Berlambine	−6.5	−7.9	−6.71	−4.64	−4.16	−5.4	−6.97
Palmatine	−4.47	−6.62	−5.15	−4.62	−4.41	−5.75	−6.35
Moslosooflavone	−5.05	−5.91	−5.03	−3.83	−4.66	−5.26	−5.47
Quercetin	−2.81	−6.42	−4.18	−3.6	−3.22	−4.49	−4.78
Moupinamide	−3.5	−6.49	−5.14	−3.07	−4.57	−4.44	−5.36
Panicolin	−4.52	−6.17	−5.38	−3.48	−4.37	−4.71	−3.94
Baicalein	−4.32	−5.92	−5.14	−4.18	−5.24	−4.45	−6.46

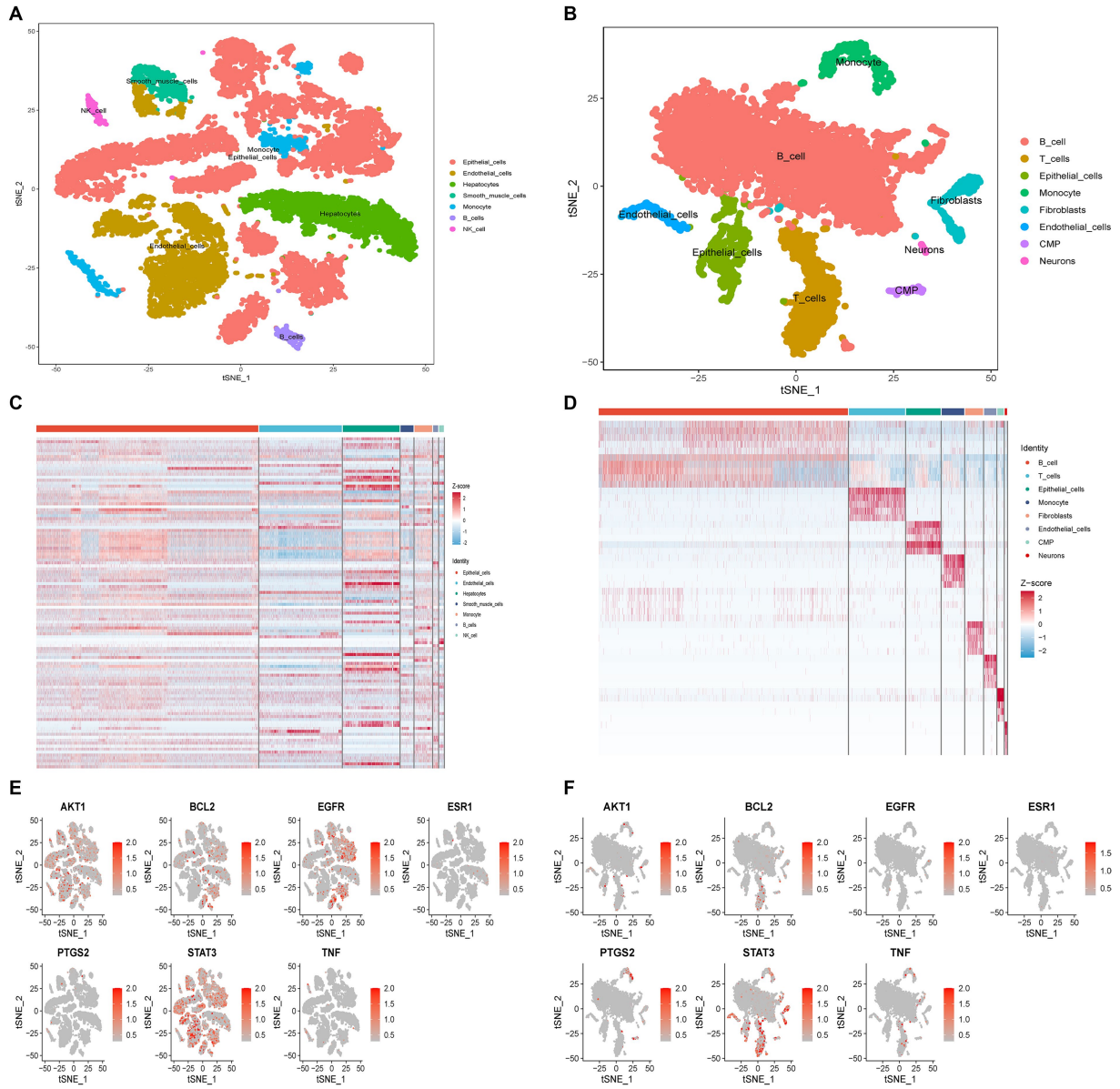

FIGURE 9
Single-cell sequencing analysis. **(A)** Cellular subtypes of T2DM. **(B)** Cellular subtypes of UC. **(C)** GQD expression in various cell clusters of T2DM. **(D)** GQD expression in various cell clusters of UC. **(E,F)** Distribution of the seven core targets in cell clusters of T2DM and UC.

holds potential as a pharmacological strategy for treating UC (50). In summary, the involvement of Eph/ephrin signaling in maintaining intestinal balance, managing inflammation, and regulating neuroimmune interactions offers exciting possibilities for future investigations and therapeutic advancements in gastrointestinal conditions (51).

In China, GQD is extensively employed for the treatment of both T2DM and UC. Through degree-value analysis of the herbal-chemical-target-protein network, we pinpointed seven active ingredients—Berlamine, Palmatine, Moslosooflavone, Quercetin, Moupinamide, Panicolin, and Baicalein—as potential compounds for treating the combined condition of T2DM and UC. Berlambine accomplishes the alleviation of inflammatory response and intestinal epithelial barrier dysfunction by diminishing the protein levels of TLR4 and MyD88, inhibiting the phosphorylation of I-κB α, and obstructing the translocation of NF-κB p65 from the cytoplasm to the nucleus (52). Concurrently, Berlambine notably increased the mRNA expression of the Nrf2 signaling pathway and elevated the activity of the pancreatic PI3K/Akt signaling pathway (53). Palmatine, a naturally occurring compound known for its anti-inflammatory and antioxidant properties, reverses the dysfunction in the insulin signaling pathway by increasing the expression of IRS-1, PI3K, AKT2, and GLUT4 genes while decreasing the expression of PKC (54). Furthermore, Palmatine alleviates ulcerative colitis symptoms by mitigating colon damage, preserving intestinal flora balance, and modulating tryptophan catabolism (55). Moslosooflavone markedly decreased the concentrations of inflammatory mediators like TNF-α, IL-1β, and IL-6 in mice (56). Quercetin's renowned anti-inflammatory properties position it as a promising natural remedy for various inflammatory conditions (57). Quercetin ameliorates UC by restoring intestinal barrier function via the activation of AHR-mediated enhancement of tight junctions (58). Additionally, quercetin provides therapeutic benefits in T2DM by inhibiting pancreatic iron accumulation and pancreatic β-cell death (59). Panicolin exhibited strong anti-inflammatory properties by significantly suppressing the production of IL-6 induced by LPS (60). Baicalein demonstrates anti-inflammatory properties by inhibiting T cell activation and suppressing the thioredoxin system to restrict NF-κB-dependent inflammatory responses (61). Moreover, baicalein exhibits multifaceted effects on glucose metabolism, enhancing glucose uptake and glycolysis while inhibiting hepatocyte gluconeogenesis through modulation of the InsR/IRS-1/PI3K/AKT pathway (62). Simultaneously, it enhances the integrity of the intestinal epithelial barrier via the AhR/IL-22 pathway in ILC3, thereby ameliorating ulcerative colitis (63). The therapeutic efficacy of the active constituents within the herbal formulation GQD for both T2DM and UC has been substantiated.

Ultimately, the affinity between seven key target proteins and active compounds was assessed through molecular docking techniques. Berlambine and Palmatine exhibited promising binding activity to the target, implying their potential relevance to the therapeutic role of GQD in treating T2DM and UC.

It is worth noting that our study also has some limitations. At the outset, our dataset originates from various public databases, each with its own set of inclusion criteria. These distinctions could potentially impact the precision of our findings. Secondly, the sample size in the GEO database is relatively small, potentially contributing to some degree of error. Additionally, variations in algorithms and parameter selections could yield divergent outcomes and interpretations. Hence, although employing various bioinformatics and machine learning approaches, validating the results through clinical trials and animal studies is imperative.

5 Conclusion

In summary, we delineated potential co-morbid mechanisms between T2DM and UC, primarily implicating pathways such as IL-17, TNF, chemokine, and Toll-like receptor signaling, alongside the involvement of immune-inflammatory pathways and various immune cells like T cells, B cells, and neutrophils. Three distinct machine learning studies identified IGFBP3 as a biomarker for GQD in treating T2DM, while BACE2, EPHB4, and EPHA2 emerged as biomarkers for GQD in UC treatment. Ultimately, our investigation identified Berlambine and Palmatine as key components of GQD, presenting promising therapeutic prospects for managing the concurrent occurrence of T2DM and UC. Additionally, our study clarifies the mechanisms underlying the therapeutic effects of GQD, employing strategies that involve multiple components, targets, and pathways. This highlights its capacity to regulate immune responses and inflammation, with a specific focus on targeting toll-like receptors, IL-17, TNF, MAPK, and PI3K-Akt signaling pathways. The therapeutic strategy involving multiple components, targets, and pathways plays a vital and effective role in enhancing treatment outcomes, mitigating drug resistance, customizing treatment plans, managing complications comprehensively, and minimizing therapeutic side effects. Consequently, this approach significantly benefits patients' clinical progress and enhances their quality of life.

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

FH: Data curation, Formal analysis, Software, Visualization, Writing – original draft. LX: Data curation, Formal analysis, Software, Visualization, Writing – original draft. ZL: Investigation, Validation, Writing – review & editing. LL: Investigation, Validation, Writing – review & editing. LW: Validation, Writing – review & editing. XW: Validation, Writing – review & editing. XZ: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – review & editing. YZ: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by Key Project of Anhui Provincial Education Department (2022AH050486) and 2021 High-level Talent Introduction Scientific Project of Anhui University of Chinese Medicine (2022rczd005) and Open Subjects of the Key Laboratory of the Ministry of Education of Xin'an Medicine (2022XAYX10).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2024.1406149/full#supplementary-material

References

1. Miao L, Liu C, Cheong MS, Zhong R, Tan Y, Rengasamy KRR, et al. Exploration of natural flavones' bioactivity and bioavailability in chronic inflammation induced-type-2 diabetes mellitus. *Crit Rev Food Sci Nutr*. (2023) 63:11640–67. doi: 10.1080/10408398.2022.2095349

2. Amar J, Chabo C, Waget A, Klopp P, Vachoux C, Bermúdez-Humarán LG, et al. Intestinal mucosal adherence and translocation of commensal bacteria at the early onset of type 2 diabetes: molecular mechanisms and probiotic treatment. *EMBO Mol Med*. (2011) 3:559–72. doi: 10.1002/emmm.201100159

3. Dubinsky M, Bleakman AP, Panaccione R, Hibi T, Schreiber S, Rubin D, et al. Bowel urgency in ulcerative colitis: current perspectives and future directions. *Am J Gastroenterol*. (2023) 118:1940–53. doi: 10.14309/ajg.0000000000002404

4. Franzosa EA, Sirota-Madi A, Avila-Pacheco J, Fornelos N, Haiser HJ, Reinker S, et al. Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nat Microbiol*. (2019) 4:293–305. doi: 10.1038/s41564-018-0306-4

5. Su J, Luo Y, Hu S, Tang L, Ouyang S. Advances in research on type 2 diabetes mellitus targets and therapeutic agents. *Int J Mol Sci*. (2023) 24:2–28. doi: 10.3390/ijms241713381

6. Paik J. Ozanimod: a review in ulcerative colitis. *Drugs*. (2022) 82:1303–13. doi: 10.1007/s40265-022-01762-8

7. Tong XL, Zhao LH, Lian FM, Zhou Q, Xia L, Zhang JC, et al. Clinical observations on the dose-effect relationship of gegen qin lian decoction on 54 out-patients with type 2 diabetes. *J Tradit Chin Med*. (2011) 31:56–9. doi: 10.1016/s0254-6272(11)60013-7

8. Xu X, Gao Z, Yang F, Yang Y, Chen L, Han L, et al. Antidiabetic effects of Gegen Qinlian decoction via the gut microbiota are attributable to its key ingredient Berberine. *Genomics Proteomics Bioinformatics*. (2020) 18:721–36. doi: 10.1016/j.gpb.2019.09.007

9. Fan Y, Yi W, Huang H, Mei Z, Feng Z. Efficacy of herbal medicine (Gegen Qinlian decoction) on ulcerative colitis: a systematic review of randomized controlled trials. *Medicine (Baltimore)*. (2019) 98:e18512. doi: 10.1097/md.0000000000018512

10. Wang Y, Zhang J, Zhang B, Lu M, Ma J, Liu Z, et al. Modified Gegen Qinlian decoction ameliorated ulcerative colitis by attenuating inflammation and oxidative stress and enhancing intestinal barrier function in vivo and in vitro. *J Ethnopharmacol*. (2023) 313:116538. doi: 10.1016/j.jep.2023.116538

11. Zhao Y, Luan H, Jiang H, Xu Y, Wu X, Zhang Y, et al. Gegen Qinlian decoction relieved DSS-induced ulcerative colitis in mice by modulating Th17/Treg cell homeostasis via suppressing IL-6/JAK2/STAT3 signaling. *Phytomedicine*. (2021) 84:153519. doi: 10.1016/j.phymed.2021.153519

12. Chen J, Li M, Chen R, Xu Z, Yang X, Gu H, et al. Gegen Qinlian standard decoction alleviated irinotecan-induced diarrhea via PI3K/AKT/NF-κB axis by network pharmacology prediction and experimental validation combination. *Chin Med*. (2023) 18:46. doi: 10.1186/s13020-023-00747-3

13. Wang X, Shen Y, Wang S, Li S, Zhang W, Liu X, et al. PharmMapper 2017 update: a web server for potential drug target identification with a comprehensive target pharmacophore database. *Nucleic Acids Res*. (2017) 45:W356–w360. doi: 10.1093/nar/gkx374

14. Xu M, Zhou H, Hu P, Pan Y, Wang S, Liu L, et al. Identification and validation of immune and oxidative stress-related diagnostic markers for diabetic nephropathy by WGCNA and machine learning. *Front Immunol*. (2023) 14:2–25. doi: 10.3389/fimmu.2023.1084531

15. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards suite: from gene data mining to disease genome sequence analyses. *Curr Protoc Bioinformatics*. (2016) 54:1.30.1–1.30.33. doi: 10.1002/cpbi.5

16. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res*. (2018) 46:D1074–d1082. doi: 10.1093/nar/gkx1037

17. Lan YZ, Bai YL, Zhu XD. Integrated traditional Chinese and Western medicine for ulcerative colitis with diabetes: a protocol for systematic review and meta-analysis. *Medicine (Baltimore)*. (2021) 100:e24444. doi: 10.1097/md.0000000000024444

18. Jess T, Jensen BW, Andersson M, Villumsen M, Allin KH. Inflammatory bowel diseases increase risk of type 2 diabetes in a Nationwide Cohort Study. *Clin Gastroenterol Hepatol*. (2020) 18:881–888.e1. doi: 10.1016/j.cgh.2019.07.052

19. Maconi G, Furfaro F, Sciurti R, Bezzio C, Ardizzone S, de Franchis R. Glucose intolerance and diabetes mellitus in ulcerative colitis: pathogenetic and therapeutic implications. *World J Gastroenterol*. (2014) 20:3507–15. doi: 10.3748/wjg.v20.i13.3507

20. Geremia A, Biancheri P, Allan P, Corazza GR, Di Sabatino A. Innate and adaptive immunity in inflammatory bowel disease. *Autoimmun Rev*. (2014) 13:3–10. doi: 10.1016/j.autrev.2013.06.004

21. Postler TS, Peng V, Bhatt DM, Ghosh S. Metformin selectively dampens the acute inflammatory response through an AMPK-dependent mechanism. *Sci Rep*. (2021) 11:18721. doi: 10.1038/s41598-021-97441-x

22. Shi N, Li N, Duan X, Niu H. Interaction between the gut microbiome and mucosal immune system. *Mil Med Res*. (2017) 4:14. doi: 10.1186/s40779-017-0122-9

23. Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci USA*. (2007) 104:13780–5. doi: 10.1073/pnas.0706625104

24. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature*. (2012) 490:55–60. doi: 10.1038/nature11450

25. Xavier RJ, Podolsky DK. Unravelling the pathogenesis of inflammatory bowel disease. *Nature*. (2007) 448:427–34. doi: 10.1038/nature06005

26. Voskens C, Stoica D, Rosenberg M, Vitali F, Zundler S, Ganslmayer M, et al. Autologous regulatory T-cell transfer in refractory ulcerative colitis with concomitant primary sclerosing cholangitis. *Gut*. (2023) 72:49–53. doi: 10.1136/gutjnl-2022-327075

27. Xu M, Duan XY, Chen QY, Fan H, Hong ZC, Deng SJ, et al. Effect of compound sophorae decoction on dextran sodium sulfate (DSS)-induced colitis in mice by regulating Th17/Treg cell balance. *Biomed Pharmacother*. (2019) 109:2396–408. doi: 10.1016/j.biopha.2018.11.087

28. Tilg H, Zmora N, Adolph TE, Elinav E. The intestinal microbiota fuelling metabolic inflammation. *Nat Rev Immunol*. (2020) 20:40–54. doi: 10.1038/s41577-019-0198-4

29. Winer DA, Luck H, Tsai S, Winer S. The intestinal immune system in obesity and insulin resistance. *Cell Metab*. (2016) 23:413–26. doi: 10.1016/j.cmet.2016.01.003

30. Pergolizzi S, Rizzo G, Favaloro A, Alesci A, Pallio S, Melita G, et al. Expression of VAChT and 5-HT in ulcerative colitis dendritic cells. *Acta Histochem*. (2021) 123:151715. doi: 10.1016/j.acthis.2021.151715

31. Qiao Z, Wang X, Zhang H, Han J, Feng H, Wu Z. Single-cell Transcriptomics reveals that metabolites produced by Paenibacillus bovis sp. nov. BD3526 ameliorate type 2 diabetes in GK rats by downregulating the inflammatory response. *Front Microbiol*. (2020) 11:568805. doi: 10.3389/fmicb.2020.568805

32. Uzzan M, Martin JC, Mesin L, Livanos AE, Castro-Dopico T, Huang R, et al. Ulcerative colitis is characterized by a plasmablast-skewed humoral response associated with disease activity. *Nat Med*. (2022) 28:766–79. doi: 10.1038/s41591-022-01680-y

33. Wang X, Zhu Y, Zhang M, Wang H, Jiang Y, Gao P. Ulcerative colitis is characterized by a decrease in regulatory B cells. *J Crohns Colitis*. (2016) 10:1212–23. doi: 10.1093/ecco-jcc/jjw074

34. Zhou T, Hu Z, Yang S, Sun L, Yu Z, Wang G. Role of adaptive and innate immunity in type 2 diabetes mellitus. *J Diabetes Res*. (2018) 2018:1–9. doi: 10.1155/2018/7457269

35. Wu MM, Wang QM, Huang BY, Mai CT, Wang CL, Wang TT, et al. Dioscin ameliorates murine ulcerative colitis by regulating macrophage polarization. *Pharmacol Res*. (2021) 172:105796. doi: 10.1016/j.phrs.2021.105796

36. SantaCruz-Calvo S, Bharath L, Pugh G, SantaCruz-Calvo L, Lenin RR, Lutshumba J, et al. Adaptive immune cells shape obesity-associated type 2 diabetes mellitus and less

prominent comorbidities. *Nat Rev Endocrinol*. (2022) 18:23–42. doi: 10.1038/s41574-021-00575-1

37. Zhai X, Qian G, Wang Y, Chen X, Lu J, Zhang Y, et al. Elevated B cell activation is associated with type 2 diabetes development in obese subjects. *Cell Physiol Biochem*. (2016) 38:1257–66. doi: 10.1159/000443073

38. Zhan J, Huang L, Ma H, Chen H, Yang Y, Tan S, et al. Reduced inflammatory responses of follicular helper T cell promote the development of regulatory B cells after roux-en-Y gastric bypass. *Clin Exp Pharmacol Physiol*. (2017) 44:556–65. doi: 10.1111/1440-1681.12740

39. Qiu T, Li M, Tanner MA, Yang Y, Sowers JR, Korthuis RJ, et al. Depletion of dendritic cells in perivascular adipose tissue improves arterial relaxation responses in type 2 diabetic mice. *Metabolism*. (2018) 85:76–89. doi: 10.1016/j.metabol.2018.03.002

40. Ying W, Fu W, Lee YS, Olefsky JM. The role of macrophages in obesity-associated islet inflammation and β-cell abnormalities. *Nat Rev Endocrinol*. (2020) 16:81–90. doi: 10.1038/s41574-019-0286-3

41. Zhang B, Liu K, Yang H, Jin Z, Ding Q, Zhao L. Gut microbiota: the potential key target of TCM's therapeutic effect of treating different diseases using the same method-UC and T2DM as examples. *Front Cell Infect Microbiol*. (2022) 12:1–19. doi: 10.3389/fcimb.2022.855075

42. Ungaro R, Mehandru S, Allen PB, Peyrin-Biroulet L, Colombel JF. Ulcerative colitis. *Lancet*. (2017) 389:1756–70. doi: 10.1016/s0140-6736(16)32126-2

43. Dugani SB, Mielke MM, Vella A. Burden and management of type 2 diabetes in rural United States. *Diabetes Metab Res Rev*. (2021) 37:e3410. doi: 10.1002/dmrr.3410

44. Rajpathak SN, He M, Sun Q, Kaplan RC, Muzumdar R, Rohan TE, et al. Insulin-like growth factor axis and risk of type 2 diabetes in women. *Diabetes*. (2012) 61:2248–54. doi: 10.2337/db11-1488

45. Pigeyre M, Sjaarda J, Mao S, Chong M, Hess S, Yusuf S, et al. Identification of novel causal blood biomarkers linking metabolically favorable adiposity with type 2 diabetes risk. *Diabetes Care*. (2019) 42:1800–8. doi: 10.2337/dc18-2444

46. D'Addio F, Maestroni A, Assi E, Ben Nasr M, Amabile G, Usuelli V, et al. The IGFBP3/TMEM219 pathway regulates beta cell homeostasis. *Nat Commun*. (2022) 13:684. doi: 10.1038/s41467-022-28360-2

47. Cardinale CJ, March ME, Lin X, Liu Y, Spruce LA, Bradfield JP, et al. Regulation of Janus kinase 2 by an inflammatory bowel disease causal non-coding single nucleotide polymorphism. *J Crohns Colitis*. (2020) 14:646–53. doi: 10.1093/ecco-jcc/jjz213

48. Madore AM, Vaillancourt VT, Bouzigon E, Sarnowski C, Monier F, Dizier MH, et al. Genes involved in Interleukin-1 receptor type II activities are associated with asthmatic phenotypes. *Allergy Asthma Immunol Res*. (2016) 8:466–70. doi: 10.4168/aair.2016.8.5.466

49. Grandi A, Zini I, Palese S, Giorgio C, Tognolini M, Marchesani F, et al. Targeting the Eph/Ephrin system as anti-inflammatory strategy in IBD. *Front Pharmacol*. (2019) 10:1–16. doi: 10.3389/fphar.2019.00691

50. Qiao Z, Liao M, Xiao M, Luo S, Wang K, Niu M, et al. Ephrin B3 exacerbates colitis and colitis-associated colorectal cancer. *Biochem Pharmacol*. (2024) 220:116004. doi: 10.1016/j.bcp.2023.116004

51. Qiu P, Li D, Xiao C, Xu F, Chen X, Chang Y, et al. The Eph/ephrin system symphony of gut inflammation. *Pharmacol Res*. (2023) 197:106976. doi: 10.1016/j.phrs.2023.106976

52. Li C, Ai G, Wang Y, Lu Q, Luo C, Tan L, et al. Oxyberberine, a novel gut microbiota-mediated metabolite of berberine, possesses superior anti-colitis effect: impact on intestinal epithelial barrier, gut microbiota profile and TLR4-MyD88-NF-κB pathway. *Pharmacol Res*. (2020) 152:104603. doi: 10.1016/j.phrs.2019.104603

53. Dou Y, Huang R, Li Q, Liu Y, Li Y, Chen H, et al. Oxyberberine, an absorbed metabolite of berberine, possess superior hypoglycemic effect via regulating the PI3K/Akt and Nrf2 signaling pathways. *Biomed Pharmacother*. (2021) 137:111312. doi: 10.1016/j.biopha.2021.111312

54. Nwabueze OP, Sharma M, Balachandran A, Gaurav A, Abdul Rani AN, Małgorzata J, et al. Comparative studies of Palmatine with metformin and glimepiride on the modulation of insulin dependent signaling pathway in vitro, in vivo & ex vivo. *Pharmaceuticals (Basel)*. (2022) 15:2–16. doi: 10.3390/ph15111317

55. Zhang XJ, Yuan ZW, Qu C, Yu XT, Huang T, Chen PV, et al. Palmatine ameliorated murine colitis by suppressing tryptophan metabolism and regulating gut microbiota. *Pharmacol Res*. (2018) 137:34–46. doi: 10.1016/j.phrs.2018.09.010

56. Zhang J, Zhao T, Zhang P, Shi Z, Da Q, Ma H, et al. Moslosooflavone protects against brain injury induced by hypobaric hypoxic via suppressing oxidative stress, neuroinflammation, energy metabolism disorder, and apoptosis. *J Pharm Pharmacol*. (2024) 76:44–56. doi: 10.1093/jpp/rgad109

57. Rogerio AP, Dora CL, Andrade EL, Chaves JS, Silva LF, Lemos-Senna E, et al. Anti-inflammatory effect of quercetin-loaded microemulsion in the airways allergic inflammatory model in mice. *Pharmacol Res*. (2010) 61:288–97. doi: 10.1016/j.phrs.2009.10.005

58. Wang X, Xie X, Li Y, Xie X, Huang S, Pan S, et al. Quercetin ameliorates ulcerative colitis by activating aryl hydrocarbon receptor to improve intestinal barrier integrity. *Phytother Res*. (2024) 38:253–64. doi: 10.1002/ptr.8027

59. Li D, Jiang C, Mei G, Zhao Y, Chen L, Liu J, et al. Quercetin alleviates ferroptosis of pancreatic β cells in type 2 diabetes. *Nutrients*. (2020) 12:2–15. doi: 10.3390/nu12102954

60. Chandrasekaran CV, Thiyagarajan P, Deepak HB, Agarwal A. In vitro modulation of LPS/calcimycin induced inflammatory and allergic mediators by pure compounds of *Andrographis paniculata* (king of bitters) extract. *Int Immunopharmacol*. (2011) 11:79–84. doi: 10.1016/j.intimp.2010.10.009

61. Patwardhan RS, Sharma D, Thoh M, Checker R, Sandur SK. Baicalein exhibits anti-inflammatory effects via inhibition of NF-κB transactivation. *Biochem Pharmacol*. (2016) 108:75–89. doi: 10.1016/j.bcp.2016.03.013

62. Yang Z, Huang W, Zhang J, Xie M, Wang X. Baicalein improves glucose metabolism in insulin resistant HepG2 cells. *Eur J Pharmacol*. (2019) 854:187–93. doi: 10.1016/j.ejphar.2019.04.005

63. Li YY, Wang XJ, Su YL, Wang Q, Huang SW, Pan ZF, et al. Baicalein ameliorates ulcerative colitis by improving intestinal epithelial barrier via AhR/IL-22 pathway in ILC3s. *Acta Pharmacol Sin*. (2022) 43:1495–507. doi: 10.1038/s41401-021-00781-7

Glossary

GQD	Gegen Qinlian Decoction
T2DM	Type 2 diabetes
UC	Ulcerative colitis
DEGs	differentially expressed genes
GEO	Gene Expression Omnibus
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
PPI	Protein–protein interaction
DC	degree centrality
BC	betweenness centrality
CC	closeness centrality
TCM	traditional Chinese medicine
DL	drug-likeness
OB	Oral bioavailability
WGCNA	Weighted Gene Co-Expression Network Analysis
FBG	fasting blood glucose
UCEIS	ulcerative colitis endoscopic index of severity
TOM	Topological Overlap Matrix
LASSO	Least Absolute Shrinkage and Selection Operator
SVM-RFE	Support Vector Machine-Recursive Feature Elimination
RF	Random Forest

frontiers | Frontiers in Medicine

TYPE Review
PUBLISHED 27 June 2024
DOI 10.3389/fmed.2024.1360966

OPEN ACCESS

EDITED BY
Hari S. Sharma,
Erasmus Medical Center, Netherlands

REVIEWED BY
Himansu Kumar,
University of Texas Health Science Center at Houston, United States
Ashraf Hamed,
Minia University, Egypt

*CORRESPONDENCE
Cailian Ruan
✉ yadxjp20210325@163.com
Yayun Wang
✉ wangyayunfmmu@qq.com
Ying Yang
✉ yying1930@163.com

RECEIVED 24 December 2023
ACCEPTED 29 May 2024
PUBLISHED 27 June 2024

CITATION
Ni Z, Ma Z, Qiao X, Guo Y, Ruan C, Wang Y and Yang Y (2024) Prediction and analysis of components and functions of *Ixeris chinensis* based on network pharmacology and molecular docking. *Front. Med.* 11:1360966. doi: 10.3389/fmed.2024.1360966

Prediction and analysis of components and functions of *Ixeris chinensis* based on network pharmacology and molecular docking

Ziwei Ni[1], Zhe Ma[2], Xiaoting Qiao[1], Yaqian Guo[1], Cailian Ruan[1,3]*, Yayun Wang[4]* and Ying Yang[5]*

[1]Medical School of Yan'an University, Yan'an, Shaanxi, China, [2]Ultrasound Room of Shaanxi Provincial Hospital of Traditional Chinese Medicine, Xi'an, Shaanxi, China, [3]Xi'an Jiaotong University School of Medicine, Xi'an, Shaanxi, China, [4]National Experimental Center of Air Force Medical University, Xi'an, Shaanxi, China, [5]Xi'an Children's Hospital Research Institute, Xi'an, Shaanxi, China

Background: It is reported that the *Ixeris chinensis* has high medicinal value, but there are few reports about its potential molecular mechanism. We used a network pharmacology approach to predict the active ingredients, targets of action and possible interventions in diseases of *Ixeris chinensis*.

Methods: We employed various databases and software to predict the active ingredients, target genes, protein interactions, signaling pathways, network diagrams, and molecular docking of *Ixeris chinensis*. Simultaneously, we searched multiple Chinese and English databases and conducted meta-analyses of five randomized controlled trials.

Results: The analysis results revealed 12 effective components, including apigenin β-sitosterol, baicalin, baicalein, and luteolin; and selected 40 key targets, including AKT1, TNF, EGFR, ESR1, SRC, among others. GO analysis generated 225 biological processes, 39 cellular components, and 65 molecular functions; KEGG analysis revealed 103 signaling pathways. Molecular docking results indicated that the main active components of *Ixeris chinensis* can bind well with key targets. Five randomized controlled trials were included. Meta-analysis showed that Ixeris extract can effectively reduce animal blood lipid levels.

Conclusion: This study revealed the main active ingredients and key targets of *Ixeris chinensis*, analyzed the signaling pathways of potential targets, conducted disease prediction, and performed molecular docking prediction, providing a basis for research on the pathways of Ixeris treatment for related diseases and subsequent new drug development.

KEYWORDS

***Ixeris chinensis*, network pharmacology, molecular docking, meta analysis, drug development**

1 Introduction

Ixeris chinensis (Thunb.) Nakai is the whole herb of *Sonchus* of the family Asteraceae (1). According to the textual research on the herbal literature, the authentic source of Chinese herbal medicine bitter vegetables should be the main stream of *Sonchus oleraceus L*, which belongs to *Sonchus*, *Ixeris polycephala* Cass of Ixeris Cass, and the dried whole grass of Chinese Ixeris (2).

I. chinensis is a traditional medicinal plant, which can be used to treat tumors, inflammation and some infections. *I. chinensis* contains a variety of flavonoids, which have various pharmacological or health effects such as antioxidant, anti-cancer, antihypertensive, improving cardiovascular diseases, lowering blood lipids, lowering cholesterol, and protecting our liver (3). *I. chinensis* is one of the traditional Chinese herbs in China, with a wide range of distribution, easy survival and a wide variety of species (4).

In recent years, with the deepening of the research on *I. chinensis*, the clinical application scope of the herb has been continuously expanded (5). It is often used in combination with other Chinese medicines for treating various inflammatory diseases such as gastritis, enteritis, lung fever and cough, sore throat, etc. (6). The pharmacological effects of *I. chinensis* have been studied more extensively, but the mechanism of action of the disease is still unknown and needs to be studied thoroughly.

Network pharmacology research uses public databases and literature to screen out the target information of known drug components (7). Then uses network visualization tools to build a multi-dimensional biological network model, through multi-level analysis of specific signal nodes in the network at the cellular, molecular and overall biological levels, to identify key nodes, and from the perspective of the overall biological network balance to discover the target drug's interference with the "pathogenic network," and then predicts the pharmacological active ingredients, targets and possible involvement in the regulation of cellular signal transduction pathways, potential pharmacological mechanisms and formulation and combination patterns of known drugs to combat diseases (8). To predict the active ingredients, targets, cellular signal transduction pathways, potential pharmacological mechanisms and prescriptions that may be involved in the regulation of known drugs against diseases (9). Finally, validation at the animal or cellular level, will reveal the modern pharmacological mechanisms of drugs against diseases and explore new indications for drug interventions (10).

This artical adopted the research method of network pharmacology. By screening the active components and key targets of *I. chinensis*, we systematically analyzed its potential action targets to construct protein interaction networks and the enriched biological processes and pathways, so as to elaborate the pharmacological action mechanism of *I. chinensis*, which laid a foundation for further in-depth study on the mechanism of action of *I. chinensis* in the treatment of specific diseases and was of great significance in the field of new drug development.

2 Materials and methods

2.1 Database

Traditional Chinese Medicine Systems Pharmacology (TCMSP[1]), HERB database[2], Pubchem database[3], Swiss Target Prediction database[4], Uniprot database[5], STRING database[6], DAVID database[7], pubmed[8], Web of Science[9], China National Knowledge Infrastructure (CNKI[10]), Wangfang database[11].

2.2 Composition software

Cytoscape3.9.1, Rstudio, Microlife Letter, PyMOL, AutoDockTools, RevMan 5.4.1.

2.3 Screening of monomeric active compounds in Chinese medicine

By searching the Traditional Chinese Medicine Systems Pharmacology Database and Analysis Platform. The results were obtained by searching for "*Sonchus oleraceus* L," "*Ixeris polycephala* Cass," "*I. chinensis* (Thunb.) Nakai," "*Lactuca tatarica*." The active ingredients were searched in the HERB database by searching for "*I. chinensis*."

2.4 Target screening of active compounds

We inputted the active components of *I. chinensis* into the swiss target prediction to predict the target genes, and used the Uniprot database (see foot note 5) to convert the full name of the target gene to the abbreviation, deleted the target without corresponding gene name, and used Cytoscape 3.9.1 to construct the network diagram of *I. chinensis* – component-target.

2.5 Critical targets screening

The potential targets were entered into the STRING database for protein interaction analysis, free targets were excluded, and the Protein–protein interaction (PPI) network was constructed using Cytoscape 3.9.1 software to screen for key targets, which could be considered as important proteins for bittercress to exert its medicinal effects.

2.6 Enrichment analysis

We used the DAVID database to perform Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis on key targets, including Biological Process (BP), Cell Component (CC) and Molecular Function (MF). With human gene as the background, set p-value<0.01, we screened the top 10 BP, CC, MF entries and the top 20 KEGG pathway entries, and used the microbiota website to make enrichment histogram and bubble chart.

1 https://old.tcmsp-e.com/tcmsp.php
2 http://herb.ac.cn/Search/
3 https://pubchem.ncbi.nlm.nih.gov/
4 http://swisstargetprediction.ch/
5 https://www.uniprot.org/
6 https://cn.string-db.org/
7 https://david.ncifcrf.gov/
8 https://pubmed.ncbi.nlm.nih.gov/
9 http://isiknowledge.com/
10 https://www.cnki.net/
11 http://www.wanfangdata.com.cn/

2.7 Prediction of target diseases

We used the Rstudio software to analyze DO data, and used the microbiota website for visual analysis. According to the *p*-value, we screened the top 20 target diseases with the highest credibility, so as to achieve the effect of disease prediction.

2.8 Molecular docking technology

According to the PPI analysis results, we selected the top ten key targets AKT1, TNF, EGFR, ESR1, SRC, PTGS2, MAPK1, MMP9, IL2 and AR for docking. Based on the comprehensive analysis of the results of "component-target network" and "component- target-disease network," the key components in *I. chinensis* luteolin, apigenin, chinensiolide c, chinensiolide b, chinensiolide a, *β*-sitosterol, luteolin-7-*O*-*β*-D-glucoside played a leading role in the network. We adopted semi flexible docking to find the best binding position and strength of substrate molecules and receptor molecules (11). We searched the 3D structure of key proteins in the PDB protein structure database, and selected the proteins with high resolution and long structure in the human background. PyMOL software was used to remove water and solvent molecules in the protein, and AutoDockTools 1.5.7 software was used to conduct hydrogenation treatment, set as ligand. After downloading the 3D structure of small molecules using Pubchem, swiss target prediction, hydrogenation was performed in AutoDockTools 1.5.7 software to automatically assign charges, set up as ligands and set up torsional bonds. We docked large and small molecules, set docking boxes, docked parameters and arithmetic methods. The screened macromolecules and small molecules from *I. chinensis* were docked separately, and the binding energy less than 0 indicated that the ligand and the receptor could bind spontaneously, the binding energy ≤ -5.0 kcal·mol^{-1} proved that the component was well docked with the target site, and the binding energy ≤ -7.0 kcal·mol^{-1} indicated that the component and the binding conformation of the target was strongly active (12). The docking results were thermographed using Rstudio, and those with high binding energy were visualized using PyMOL.

2.9 Meta analysis

2.9.1 Research type

Randomized controlled trial (RCT) of *I. chinensis* water extract in the treatment of diseases.

2.9.2 Research objects

Mice fed with high-fat diet or injected with ccl4 and successfully modeled.

2.9.3 Intervention measures

Test group treated rats with *I. chinensis* water extract. The control group used placebo or no special treatment.

2.9.4 Outcome indicators

Total Cholesterol (TC), Triglycerides (TG), High-Density Lipoprotein Cholesterol (HDL-C), Low-Density Lipoprotein Cholesterol (LDL-C), Malondialdehyde (MDA).

2.9.5 Literature retrieval strategy

Subject term retrieval (see footnote 8), Web of Science (see footnote 9), CNKI (see footnote 10), Wanfang database (Wanfang, see footnote 11), to be included in the randomized controlled trial with *I. chinensis* as the treatment method. Take Pubmed as an example. See the table for its specific retrieval strategy (Table 1).

2.9.6 For literature quality evaluation

Cochrane risk bias assessment tool and SYRCLE bias risk assessment tool were used to evaluate the quality of the included studies. The evaluation contents included: ① Random sequence generation (selection bias) ② Allocation consideration (selection bias), ③ Blinding of participants and persons (performance bias), ④ Blinding of outcome assessment (detection bias), ⑤ Incomplete outcome data (attribute bias), ⑥ Selective reporting (reporting bias), ⑦ Other biases.

2.9.7 Statistical analysis

RevMan 5.4.1 software was used for meta analysis. The standardized mean difference (SMD) was used as the effect index for the measurement data. The point estimates and 95% CI were given for each effect. If $p \geq 0.1$ and $I^2 < 50\%$, it was considered that the heterogeneity between the studies was not significant, and the fixed effect model was selected; If $p < 0.1$, $I^2 \geq 50\%$, it was considered that there was significant heterogeneity among the studies, and the random effect model was selected (13).

3 Results

3.1 Screening for active compounds of *I. chinensis*

We searched the HERB database for "*I. chinensi*" and found that there were 12 active ingredients, namely Apigenin, *β*-sitosterol, calcium carbonate, chinensiolide a, chinensiolide b, chinensiolide c, ixerisosidea, ixerochinolide, ixerochinoside, Lactucin, Luteolin,

TABLE 1 Pubmed retrieval strategy.

#1 TS = (("asteraceae"[MeSH Terms] OR "asteraceae"[All Fields] OR "ixeris"[All Fields]) AND chinensis[All Fields]) AND (("disease"[MeSH Terms] OR "disease"[All Fields] OR "diseases"[All Fields]) OR ("disease"[MeSH Terms] OR "disease"[All Fields]))
#2 TS = (("asteraceae"[MeSH Terms] OR "asteraceae"[All Fields] OR "ixeris"[All Fields]) AND chinensis[All Fields]) AND (("disease"[MeSH Terms] OR "disease"[All Fields] OR "diseases"[All Fields]) OR ("disease"[MeSH Terms] OR "disease"[All Fields]))
#3 TS = (("asteraceae"[MeSH Terms] OR "asteraceae"[All Fields] OR "ixeris"[All Fields]) AND chinensis[All Fields]) AND (("inflammation"[MeSH Terms] OR "inflammation"[All Fields]) OR ("inflammation"[MeSH Terms] OR "inflammation"[All Fields] OR ("innate"[All Fields] AND "inflammatory"[All Fields] AND "response"[All Fields]) OR "innate inflammatory response"[All Fields]) OR (Inflammatory[All Fields] AND Response[All Fields]) OR Innate[All Fields] OR ("inflammation"[MeSH Terms] OR "inflammation"[All Fields] OR ("innate"[All Fields] AND "inflammatory"[All Fields] AND "responses"[All Fields]) OR "innate inflammatory responses"[All Fields]))

luteolin-7-*O*-*β*-D-glucoside. The above active ingredients were used as research subjects.

3.2 Analysis of *I. chinensis*–active compound–target network

Since there was no match between the targets of HERB database and the active ingredients of *I. chinensis*, so we entered into Pubchem for searching, and the smileIDs of 10 active ingredients were obtained and then entered into Swiss Target Prediction for target gene prediction, among which the remaining two ingredients, "Lactucin" was used for target gene prediction in TCMSP website and "calcium carbonate" was used for target gene prediction in National Library of Medicine website. By screening the target genes with confidence greater than 0.01 or the top 15 target classes, converting the full names of target genes to abbreviations with the help of Uniprot database (see footnote 5) and deleting the targets without corresponding gene names, a total of 238 targets were derived, and using Cytoscape 3.9.1 to construct the *I. chinensis*–active compound–target network (Figure 1).

3.3 Key targets prediction

After entering 238 potential targets into STRING database for protein interaction analysis and eliminating free targets, we used

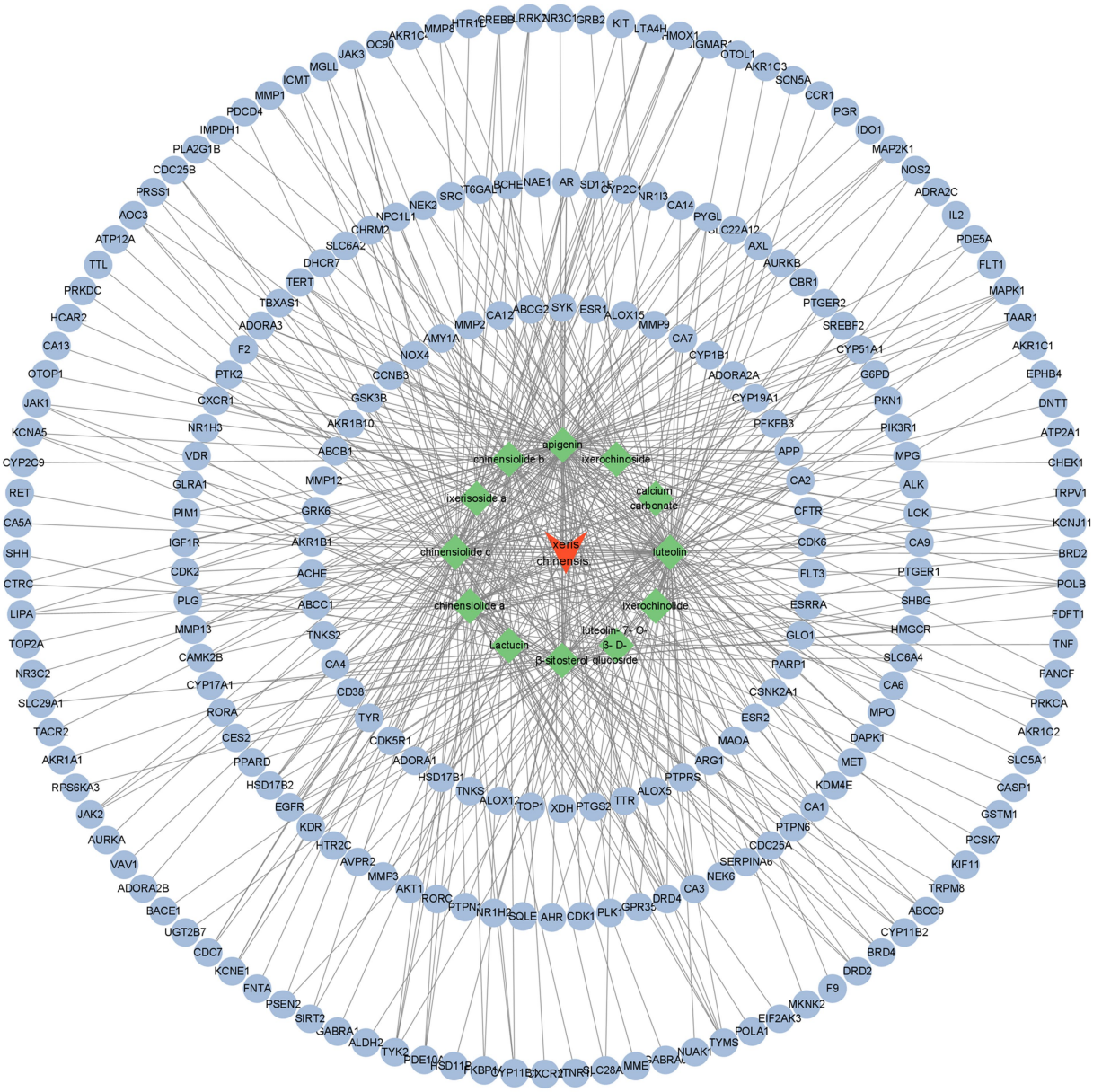

FIGURE 1
Ixeris chinensis active ingredient-target network diagram. The diamond shape represented the active ingredients of this herb, the oval represented the target site.

Cytoscape 3.9.1 software to construct PPI network (Figure 2) and calculated parameter thresholds, screened Degree (node centrality, reflecting the number of other targets that participate in different disease pathological processes together with a certain target) greater than 30, BC (mediator centrality value, the target with high proximity centrality was directly connected to more targets, indicating the importance of the target in different disease pathomechanical processes) was greater than 369, CC (closeness centrality, reflecting the proximity between a node and other nodes in the network) was greater than 0.0017, a total of 40 key targets were screened, which could be regarded as important proteins for *I. chinensis* to exert their medicinal effects (Table 2).

3.4 GO enrichment analysis of targets

GO-BP analysis showed that 225 biological processes were related to this, and the most significant top 5 were positive regulation of transcription from RNA polymerase II promoter, signal transduction, response to xenobiotic stimulus, positive regulation of gene expression, positive regulation of transcription, DNA-templated (Table 3). GO-CC analysis revealed 39 cellular compositions associated with this, the five most significant being cytoplasm, nucleoplasm, plasma membrane, nucleus, cytosol (Table 4). GO-MF analysis showed that 65 molecular functions were related to this, and the most significant 5 were protein binding, ATP binding, identical protein binding, enzyme binding, protein homodimerization

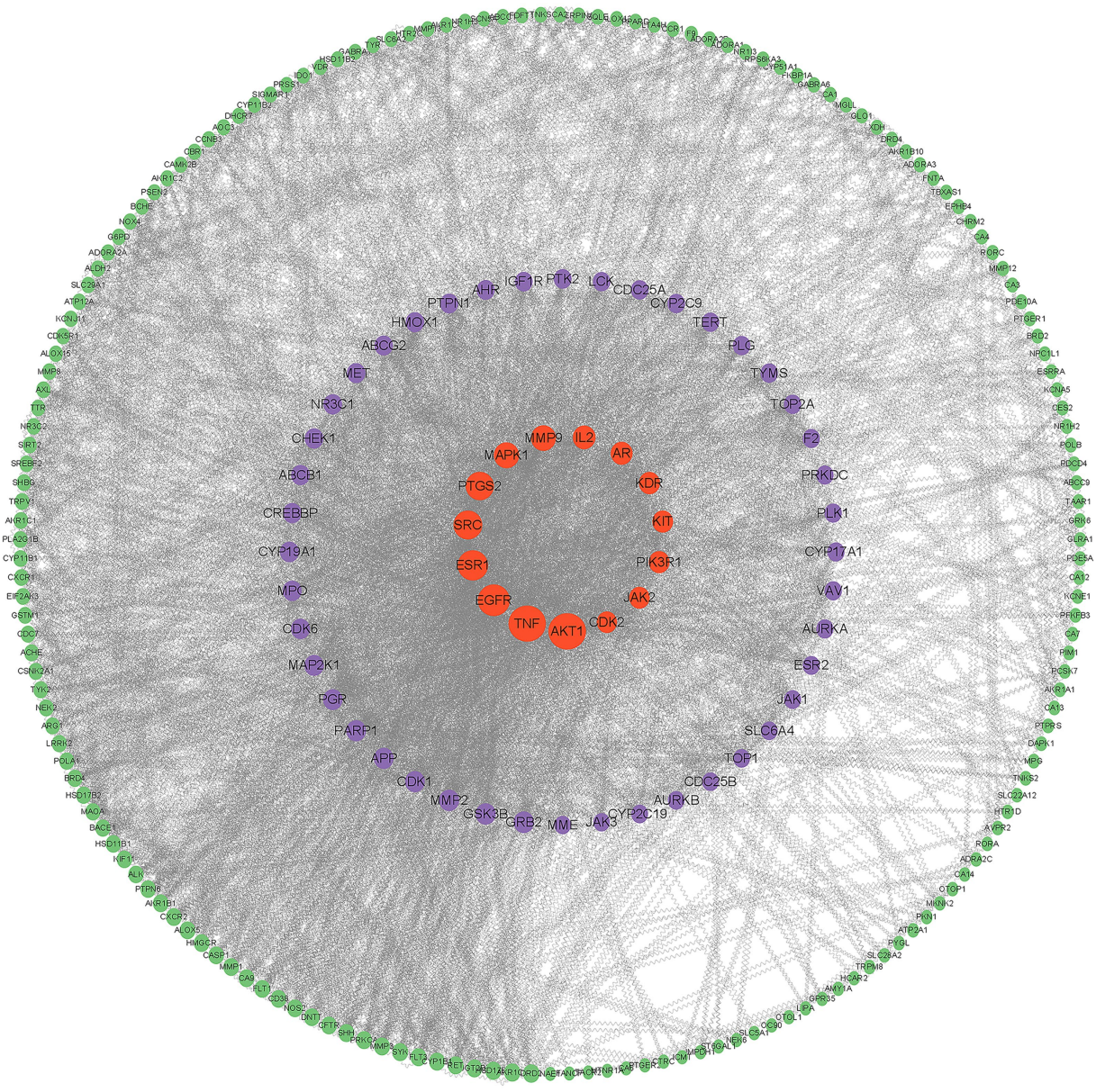

FIGURE 2
PPI network diagram. The larger the node shape in the figure and the closer to the center, the higher the criticality.

(Table 5). The results suggested that *I. chinensis* could exert their therapeutic effects by participating in the regulation of various biological processes (Figure 3).

3.5 Analysis of KEGG enrichment pathway

KEGG is the main database for systematic analysis of gene function, genomic and proteomic information, where different proteins exercise their biological behaviors in coordination with each other, and it helps researchers to study protein and expression information as a whole network. Included in the database are illustrated cellular biochemical processes such as metabolism, membrane transport, signaling, cell cycle, and also information on homologous conserved subpathways (14).

KEGG analysis showed that the five most significantly enriched pathways were Pathways in cancer, Human papillomavirus infection, Prostate cancer, Breast cancer, Chemical carcinogenesis-receptor activation (Table 6 and Figure 4). Among the pathways associated with disease classes, the main ones are the Pathways in cancer, Human papillomavirus infection, Prostate cancer, Breast cancer, Chemical carcinogenesis – receptor activation, Human cytomegalovirus infection, Cushing syndrome, Hepatitis C, Hepatitis B, Kaposi sarcoma-associated herpesvirus infection, Viral carcinogenesis, Proteoglycans in cancer, Lipid and atherosclerosis, Human T-cell leukemia virus 1 infection, MicroRNAs in cancer, Alzheimer disease, etc. Among the pathways associated with signaling functions were PI3K-Akt signaling pathway, Estrogen signaling pathway, Prolactin signaling pathway, etc. It was suggested that the potential targets were involved in multiple signaling pathways acting in concert.

3.6 Analysis of KEGG signaling pathways related to cardiovascular and cerebrovascular

Cardiovascular and cerebrovascular diseases based on atherosclerosis are the main killers threatening human health. Years of research showed that the common pathological basis of cardiovascular and cerebrovascular diseases comes from atherosclerosis, so the key to prevent cardiovascular and cerebrovascular diseases is to prevent atherosclerosis (15).

There are many factors affecting cardiovascular and cerebrovascular diseases, such as vascular endothelial growth factor A (VEGF-A), a dimer glycoprotein encoded by VEGF-A gene, which plays a crucial role in the process of inducing vascular growth (16). Many studies have also confirmed that Akt and downstream molecular activity changes are related to some cardiovascular diseases. Minamino et al. (17) detected a vascular endothelial cell senescence phenotype in atheromatous plaque tissue. The findings suggested that statins improved endothelial function and reduced myocardial remodeling in the early stages of myocardial infarction through activation of Akt. Yang Hui et al. (18) prepared a hyperlipidemia mouse model, treated the water extract of *I. chinensis* and injected it intraperitoneally. The study showed that *I. chinensis* has an important contribution to improving the antioxidant capacity of hyperlipidemia mice, and could reduce the level of lipid peroxidation, but the specific deep mechanism needs further study.

TABLE 2 Key targets prediction.

Gene	DC	BC	CC
AKT1	194	7225.132999	0.00255102
TNF	192	7903.35269	0.002557545
EGFR	154	3971.227254	0.002409639
ESR1	142	4009.212674	0.00234192
SRC	130	2895.883562	0.002325581
PTGS2	126	3198.407973	0.002293578
MAPK1	104	1673.561859	0.002150538
MMP9	102	1417.735444	0.002136752
IL2	86	595.0008977	0.002061856
AR	80	1267.538745	0.002118644
KIT	72	525.9431779	0.00203666
CDK2	68	709.8494537	0.002012072
GSK3B	66	423.3720516	0.00203666
APP	64	1312.657413	0.002057613
PARP1	64	725.8641357	0.001976285
CDK1	64	463.6579469	0.001941748
PGR	62	591.4637203	0.002074689
CYP19A1	58	1020.976118	0.002
CDK6	58	376.8362946	0.001915709
ABCB1	56	1720.684138	0.002114165
NR3C1	56	939.3539641	0.002053388
CREBBP	56	620.1867078	0.00203252
CHEK1	56	898.8009262	0.001915709
ABCG2	54	1410.390052	0.002020202
AHR	52	909.7985089	0.002004008
CYP2C9	48	638.5874397	0.001964637
TYMS	48	907.4959802	0.001886792
TOP2A	48	524.5084321	0.001808318
F2	46	790.2361428	0.001953125
CYP17A1	46	439.2279507	0.001769912
TOP1	42	381.8070735	0.001901141
SLC6A4	42	1335.630618	0.001883239
CYP2C19	42	622.4386451	0.001841621
DRD2	40	2328.251479	0.001945525
MME	40	682.4157568	0.001883239
AKR1B1	34	1114.208445	0.001937984
CFTR	34	1748.372878	0.001893939
CA9	34	762.4022156	0.001886792
HMGCR	34	633.4694013	0.001865672
HSD11B1	32	376.5306058	0.001782531

Signaling pathways are important for the study of diseases, based on the key targets in *I. chinensis* and cardiovascular diseases, KEGG enrichment analysis showed that lipids and atherosclerosis was more significant signaling pathways. Therefore, taking this pathway as an example, it is of great significance to explore the role of the key target

TABLE 3 GO-BP analysis.

Signal pathway	*p*-value	Count
Positive regulation of transcription from RNA polymerase II promoter	3.57028×10^{-7}	14
Signal transduction	2.82142×10^{-5}	12
Response to xenobiotic stimulus	3.3837×10^{-11}	11
Positive regulation of gene expression	6.72839×10^{-7}	10
Positive regulation of transcription, DNA-templated	1.08731×10^{-5}	10
Response to drug	1.04544×10^{-7}	9
Negative regulation of transcription from RNA polymerase II promoter	5.99392×10^{-4}	9
Protein phosphorylation	5.6852×10^{-05}	8
Negative regulation of apoptotic process	7.72192×10^{-05}	8
Peptidyl-serine phosphorylation	1.57615×10^{-06}	7

TABLE 4 GO-CC analysis.

Signal pathway	*p*-value	Count
Cytoplasm	2.85031×10^{-5}	24
Nucleoplasm	3.156×10^{-5}	20
Plasma membrane	9.68707×10^{-4}	20
Nucleus	7.52085×10^{-3}	20
Cytosol	7.28669×10^{-3}	19
Integral component of membrane	3.2415493×10^{-2}	17
Macromolecular complex	2.7946×10^{-10}	14
Membrane	6.305045×10^{-3}	12
Endoplasmic reticulum membrane	1.24183×10^{-4}	10
Mitochondrion	4.887025×10^{-3}	9

TABLE 5 GO-MF analysis.

Signal pathway	*p*-value	Count
Protein binding	9.727706×10^{-3}	34
ATP binding	1.22316×10^{-6}	15
Identical protein binding	2.32923×10^{-5}	14
Enzyme binding	2.54902×10^{-10}	12
Protein homodimerization activity	1.36166×10^{-06}	11
DNA binding	1.311848×10^{-3}	10
Zinc ion binding	3.51011×10^{-4}	9
Chromatin binding	4.26989×10^{-5}	8
Protein kinase activity	1.17564×10^{-4}	7
Protein serine/threonine kinase activity	1.41325×10^{-4}	7

in the signal pathway of *I. chinensis*, and to explore the specific mechanism of disease and treat disease.

3.6.1 Lipid and atherosclerosis signal pathway

The elevated level of low density lipoprotein (LDL) cholesterol is a major risk factor for atherosclerosis. LDL can accumulate in the blood vessel wall and be modified by oxidation. Oxidized LDL (oxLDL) leads to endothelial dysfunction, the expression of adhesion molecules and the recruitment of monocytes in the subendothelial space (19). The extracellular oxLDL binds to the membrane receptor Lox-1. LOX-1 mediates the recognition and internalization of oxLDL, activates the substrate ROCK2, which inhibits PI3K, and then generates PIP3, thereby activating AKT, which promotes Bclxl phosphorylation, thus inducing cell apoptosis (20, 21).

Advanced glycosylation end products (AGEs) are harmful compounds formed by the combination of protein or fat in the blood with sugar. They bind with the RAGE receptor on the cell membrane, and then activate and generate ROS (22). ROS indirectly activates Src, which in turn activates Rac1. Rac1 further activates MKK4/7, promotes JNK phosphorylation, and then promotes AP-1 phosphorylation. The phosphorylated AP-1 will enter the nucleus, regulate DNA transcription, and affect the expression of MCP-1, ICAM-1, and IL-8, thus affecting the attachment and activation of downstream monocytes. ERK activated in cytoplasm can indirectly activate NF-κB, NF-κB affects the expression of MMP1,3,9 and TNF-α by promoting DNA transcription, thus affecting damage, cell apoptosis and inflammation (23).

Minimally modified low density lipoprotein (mmLDL) is LDL in which only the lipid fraction is oxidized and the lysine residue in the Apo B100 structure is not destroyed (24). MmLDL binds with TLR2/4/6 receptor on cell membrane to activate ERK, which in turn indirectly activates rap1, and rap1 activates Rac, which in turn activates NADPH oxidase, promotes the expression of ROS *in vivo*, and indirectly promotes NF-κB. In addition, mmLDL promotes the separation of intracytoplasmic phosphorylated IκBα from NF-κB through a series of pathways, and the separated NF-κB promotes DNA transcription, thus affecting the expression of MMP1,3,9 and TNF-α, which have effects on monocyte attachment and activation, injury and apoptosis, and inflammatory response.

Tumor necrosis factor (TNF-α) is a cytokine involved in systemic inflammation, and also one of the many cytokines that cause acute response. Extracellular TNF-α binds to the TNF-R1 receptor on the cytomembrane, indirectly activating CASP8, and then activating CASP3, thereby affecting cell apoptosis and plaque instability (25).

Very low density lipoprotein(VLDL)is a kind of lipoprotein which is composed of apolipoprotein synthesized in the liver by using the intermediate metabolites of chyle granules, bile acid, fatty acid, sugar and protein. VLDL binds to VLDLR receptor on the cell membrane, activates substrate PI3K, connects with PI3K-AKT signal, promotes the expression of PIP3, and then activates AKT (26). AKT phosphorylates GSK3β, and the phosphorylated GSK3β phosphorylates NFAT, which promotes DNA transcription and has an effect on cell proliferation (27).

The signal pathways are complex, and the proteins related to the key targets in the *I. chinensis* were all marked red (Figure 5), such as Src, TNFα, ERK, AKT, CYP, etc. It could be seen from the figure that most of the proteins in this pathway were distributed in the middle reaches, and they had important effects on inflammatory reaction, foam cell formation, cytokines, inflammatory signals, plaque

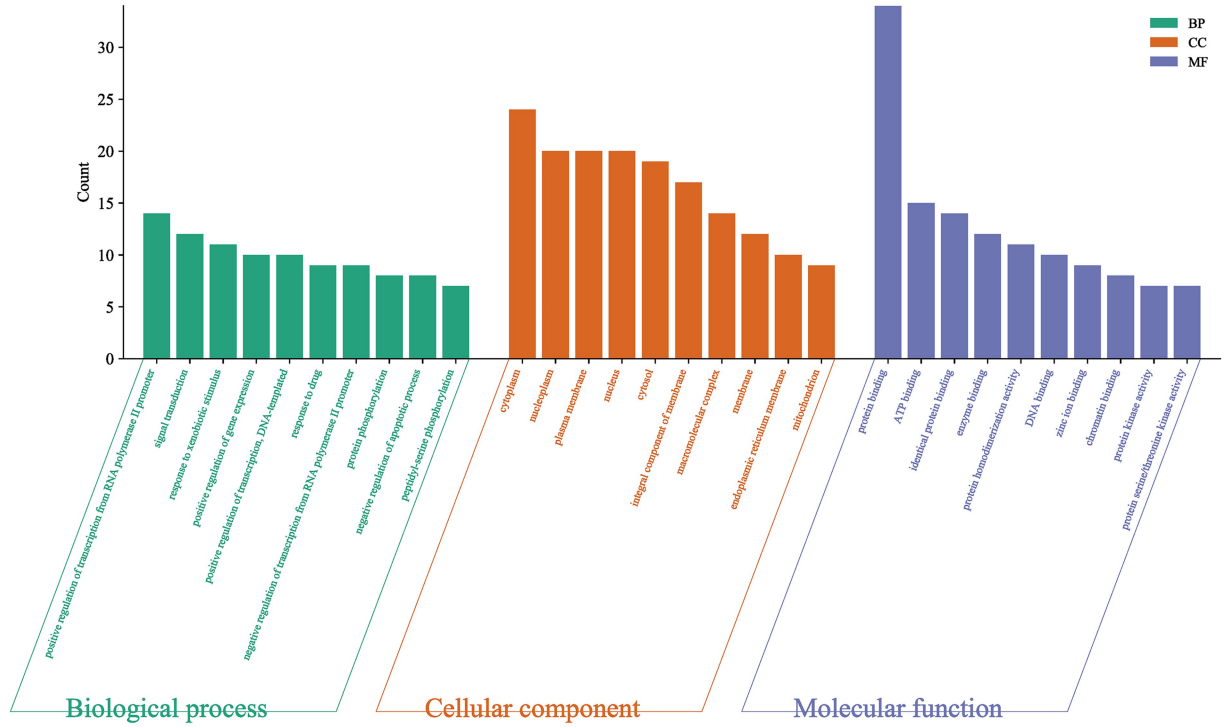

FIGURE 3
GO enrichment analysis results of key target proteins. Including biological process, cellular components, molecular functions.

TABLE 6 Key target protein KEGG enrichment results.

Term	*p*-value	GENE
Pathways in cancer	3.85841×10^{-7}	GSK3B, CREBBP, F2, PTGS2, ESR1, MMP9, EGFR, IL2, AR, CDK6, KIT, CDK2, AKT1, MAPK1
Human papillomavirus infection	1.12948×10^{-4}	GSK3B, CREBBP, CDK6, CDK2, MAPK1, AKT1, PTGS2, TNF, EGFR
Prostate cancer	2.50934×10^{-7}	GSK3B, AR, CREBBP, CDK2, MAPK1, AKT1, MMP9, EGFR
Breast cancer	4.20554×10^{-6}	GSK3B, CDK6, KIT, MAPK1, AKT1, PGR, ESR1, EGFR
Chemical carcinogenesis - receptor activation	4.5876×10^{-5}	AR, SRC, MAPK1, AKT1, PGR, AHR, ESR1, EGFR
Human cytomegalovirus infection	6.69942×10^{-5}	GSK3B, CDK6, SRC, MAPK1, AKT1, PTGS2, TNF, EGFR
PI3K-Akt signaling pathway	1.070434×10^{-3}	GSK3B, CDK6, KIT, CDK2, MAPK1, AKT1, EGFR, IL2
Estrogen signaling pathway	3.70105×10^{-5}	SRC, MAPK1, AKT1, PGR, ESR1, MMP9, EGFR
Cushing syndrome	7.10189×10^{-5}	GSK3B, CDK6, CDK2, MAPK1, AHR, EGFR, CYP17A1
Hepatitis C	7.62765×10^{-5}	GSK3B, CDK6, CDK2, MAPK1, AKT1, TNF, EGFR
Hepatitis B	9.07914×10^{-5}	CREBBP, SRC, CDK2, MAPK1, AKT1, TNF, MMP9
Kaposi sarcoma-associated herpesvirus infection	2.44053×10^{-4}	GSK3B, CREBBP, CDK6, SRC, MAPK1, AKT1, PTGS2
Viral carcinogenesis	3.20198×10^{-4}	CREBBP, CDK6, SRC, CHEK1, CDK2, CDK1, MAPK1
Proteoglycans in cancer	3.28737×10^{-4}	SRC, MAPK1, AKT1, ESR1, TNF, MMP9, EGFR
Lipid and atherosclerosis	4.24369×10^{-4}	GSK3B, CYP2C9, SRC, MAPK1, AKT1, TNF, MMP9
Human T-cell leukemia virus 1 infection	5.03384×10^{-4}	CREBBP, CHEK1, CDK2, MAPK1, AKT1, TNF, IL2
MicroRNAs in cancer	2.82329×10^{-3}	CREBBP, ABCB1, CDK6, MAPK1, PTGS2, MMP9, EGFR
Alzheimer disease	8.007899×10^{-3}	GSK3B, APP, MME, MAPK1, AKT1, PTGS2, TNF
Prolactin signaling pathway	1.61592×10^{-5}	GSK3B, SRC, MAPK1, AKT1, ESR1, CYP17A1
Endocrine resistance	8.24887×10^{-5}	SRC, MAPK1, AKT1, ESR1, MMP9, EGFR

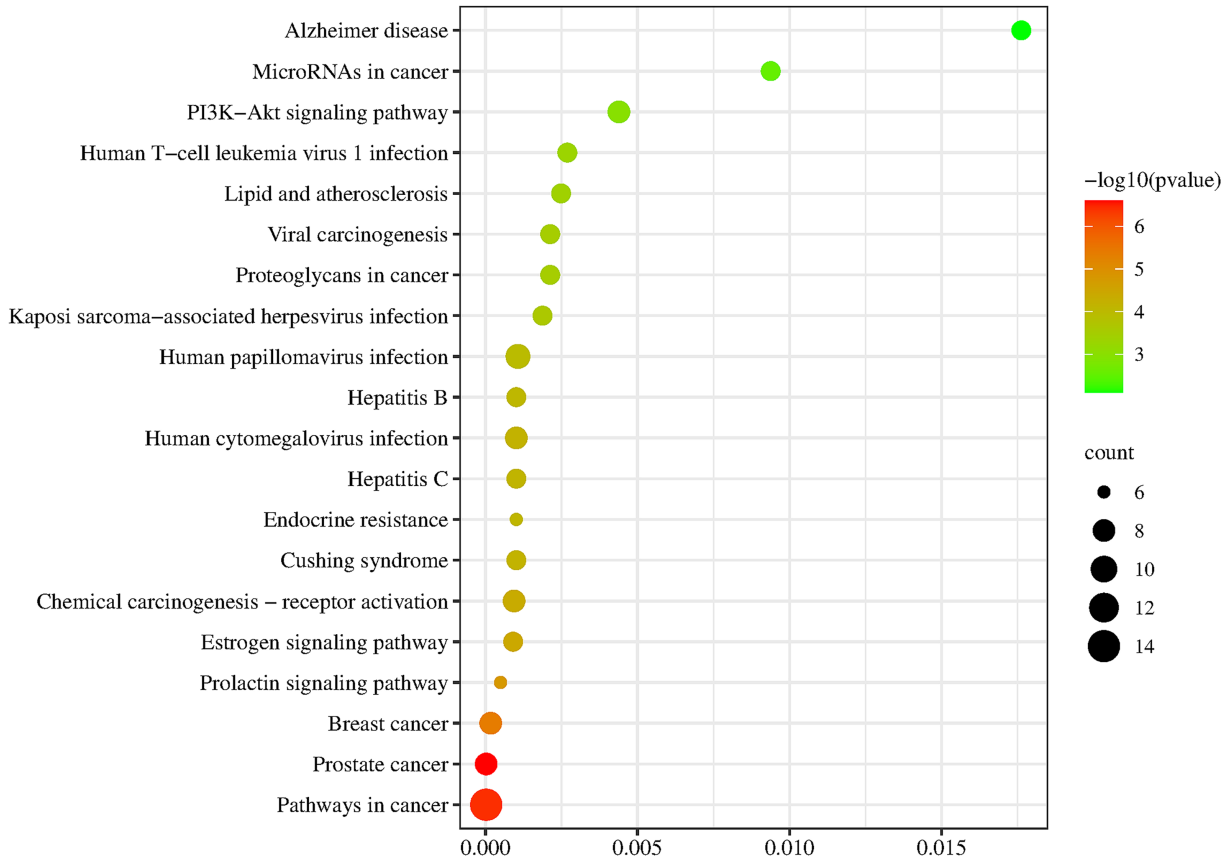

FIGURE 4
Enrichment results of key target protein KEGG. The larger the point was, the more relevant target points were; The redder the color, the higher the credibility.

instability, etc. The study of this pathway is helpful for anti atherosclerosis.

3.7 Analysis of related diseases enriched by targets

DO data analysis was performed using Rstudio software, setting p-value <0.05 and Q value = 1. Disease enrichment analysis was performed on the key targets of *I. chinensis* action, removing the same diseases, and then extracting the 20 diseases with the highest number of targets and visualizing them using the Microbiology website (Figure 6 and Table 7). The main diseases included autosomal dominant disease, hereditary breast ovarian cancer, cell type benign neoplasm, female reproductive organ cancer, adenoma, biliary tract cancer, brain disease, diarrhea, renal cell carcinoma, atherosclerosis, arteriosclerotic cardiovascular disease, leiomyoma, osteoporosis, Alzheimer's disease, tauopathy, bone cancer, connective tissue cancer, lymphoblastic leukemia, obesity, infertility, etc. The key targets proved to be highly relevant to these diseases. In order to explore the association between *I. chinensis* and diseases, Cytoscape 3.9.1 software was used to construct and visualize the "component-target-disease" network map using collected components, targets and disease data (Figure 7). The network consisted of 12 components, 238 targets and 20 diseases, which visually showed that multiple targets contained in one component could participate in the regulation of multiple diseases at the same time.

3.8 Molecular docking technique to predict the binding ability of the active components and key targets of *I. chinensis*

Molecular docking technology is of great significance for further research in plant chemistry and biology (28). The current study docked each of the 7 active ingredients in *I. chinensis* with 10 key targets, yielding 70 docking results (Figure 8). The results revealed that luteolin, apigenin, chinensiolide c, chinensiolide b, chinensiolide a, β-sitosterol, and luteolin-7-*O*-β-D-glucoside had better docking ability with EGFR, ESR1, PTGS2, MAPK1, MMP9, and AR. The lower the binding energy, the better the molecule's ability to bind to the protein (Table 8). Among them, MMP9 showed strong docking ability with each active ingredient, and β-sitosterol also showed strong binding ability with each protein. The docking results of the top 8 components with strong binding activity (binding energy $\leq$ −7 kcal·mol-1) to the targets were visualized and presented using PyMOL software (Figure 9). β-sitosterol forms one hydrogen bond with MMP9 via amino acid residue GLU-111, one hydrogen bond

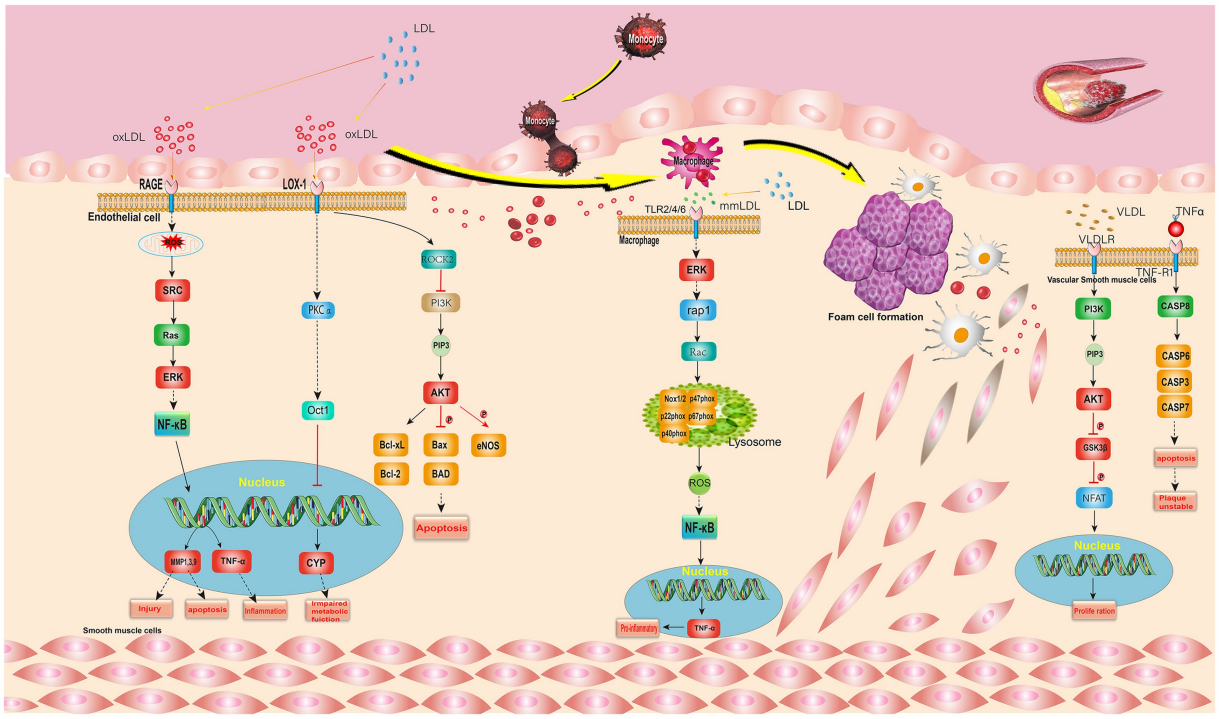

FIGURE 5
Lipid and atherosclerosis signal pathway. The gene marked red is the key target in *I. chinensis*.

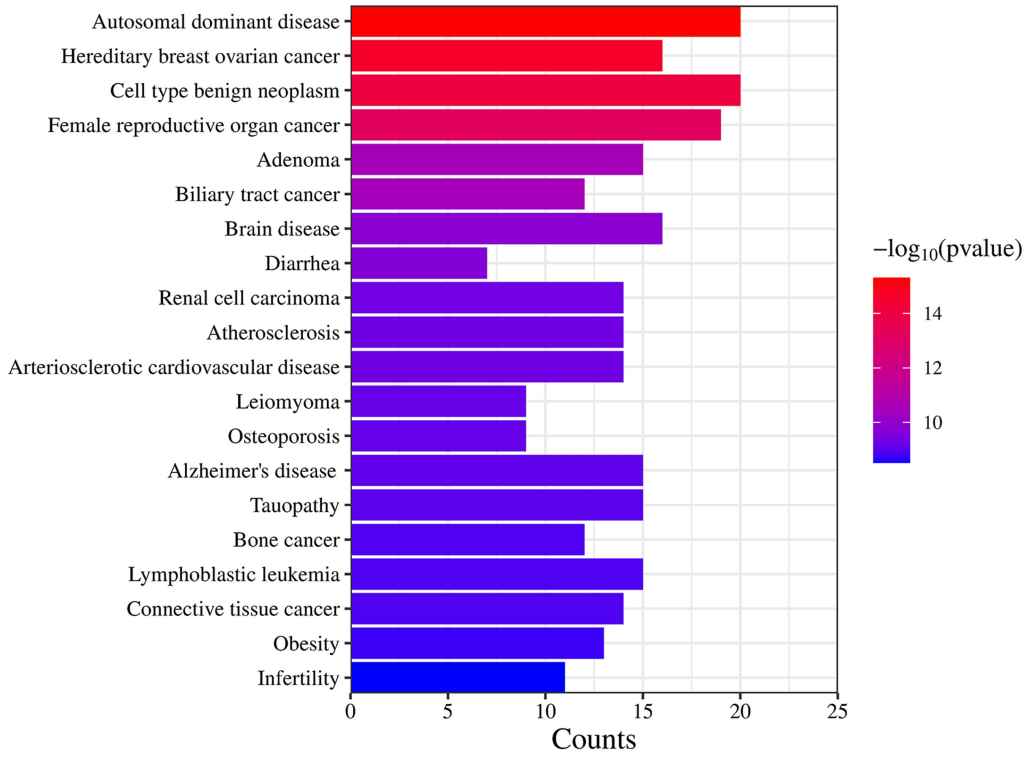

FIGURE 6
DO enrichment results for key target proteins. The longer the bar graph was, the more relevant targets were; The redder the color, the higher the credibility.

10.3389/fmed.2024.1360966

TABLE 7 DO enrichment results of key target protein.

Description	*p*-value	Gene
Autosomal dominant disease	4.97947×10^{-16}	AKT1, TNF, EGFR, ESR1, SRC, MAPK1, AR, KIT, CDK2, GSK3B, PARP1, PGR, CYP19A1, ABCB1, CREBBP, CHEK1, AHR, TOP2A, CYP17A1, TOP1
Hereditary breast ovarian cancer	2.47998×10^{-15}	AKT1, EGFR, ESR1, MAPK1, AR, CDK2, PARP1, PGR, CYP19A1, ABCB1, CREBBP, CHEK1, AHR, TOP2A, CYP17A1, TOP1
Cell type benign neoplasm	8.98661×10^{-15}	AKT1, TNF, EGFR, ESR1, SRC, PTGS2, MAPK1, MMP9, AR, PARP1, PGR, CYP19A1, ABCB1, AHR, TOP2A, CYP17A1, DRD2, MME, CA9, HSD11B1
Female reproductive organ cancer	6.49841×10^{-14}	AKT1, TNF, EGFR, ESR1, SRC, PTGS2, MAPK1, MMP9, AR, KIT, CDK1, PGR, CYP19A1, ABCB1, ABCG2, TYMS, TOP2A, MME, CA9
Adenoma	2.96803×10^{-11}	AKT1, EGFR, ESR1, PTGS2, MAPK1, MMP9, AR, PGR, ABCB1, AHR, TOP2A, DRD2, MME, CA9, HSD11B1
Biliary tract cancer	3.26756×10^{-11}	AKT1, TNF, EGFR, ESR1, PTGS2, MAPK1, MMP9, ABCB1, ABCG2, TOP2A, CYP2C19, MME
Brain disease	1.5651×10^{-10}	TNF, EGFR, ESR1, PTGS2, MMP9, IL2, AR, APP, PGR, ABCB1, ABCG2, CYP2C9, SLC6A4, CYP2C19, DRD2, HSD11B1
Diarrhea	2.17837×10^{-10}	TNF, EGFR, PTGS2, ABCB1, ABCG2, SLC6A4, CFTR
Renal cell carcinoma	4.37165×10^{-10}	AKT1, TNF, EGFR, ESR1, PTGS2, MAPK1, MMP9, IL2, AR, KIT, ABCB1, TYMS, MME, CA9
Atherosclerosis	4.72123×10^{-10}	AKT1, TNF, ESR1, PTGS2, MMP9, AR, CYP19A1, AHR, CYP2C9, F2, SLC6A4, CYP2C19, AKR1B1, HMGCR
Arteriosclerotic cardiovascular disease	4.90545×10^{-10}	AKT1, TNF, ESR1, PTGS2, MMP9, AR, CYP19A1, AHR, CYP2C9, F2, SLC6A4, CYP2C19, AKR1B1, HMGCR
Leiomyoma	6.40375×10^{-10}	TNF, EGFR, ESR1, AR, PARP1, PGR, CYP19A1, CYP17A1, DRD2
Osteoporosis	7.04139×10^{-10}	TNF, ESR1, MAPK1, AR, KIT, CYP19A1, ABCB1, CYP17A1, HSD11B1
Alzheimer's disease	8.15987×10^{-10}	AKT1, TNF, ESR1, PTGS2, MAPK1, GSK3B, APP, PARP1, CDK1, CYP19A1, ABCB1, F2, SLC6A4, MME, CFTR
Tauopathy	9.28254×10^{-10}	AKT1, TNF, ESR1, PTGS2, MAPK1, GSK3B, APP, PARP1, CDK1, CYP19A1, ABCB1, F2, SLC6A4, MME, CFTR
Bone cancer	1.22472×10^{-9}	AKT1, TNF, EGFR, ESR1, PTGS2, MAPK1, MMP9, AR, KIT, ABCB1, CHEK1, TOP1
Connective tissue cancer	1.22959×10^{-9}	AKT1, TNF, EGFR, ESR1, SRC, PTGS2, MAPK1, MMP9, AR, KIT, ABCB1, CHEK1, TOP2A, TOP1
Lymphoblastic leukemia	1.23466×10^{-9}	AKT1, TNF, PTGS2, MAPK1, MMP9, IL2, GSK3B, PARP1, CDK6, ABCB1, ABCG2, AHR, TYMS, TOP2A, MME
Obesity	1.80466×10^{-9}	TNF, ESR1, MAPK1, MMP9, AR, APP, PGR, CYP19A1, ABCB1, NR3C1, SLC6A4, DRD2, HSD11B1
Infertility	3.01154×10^{-9}	TNF, ESR1, MMP9, AR, PGR, CYP19A1, ABCB1, AHR, F2, CYP17A1, CFTR

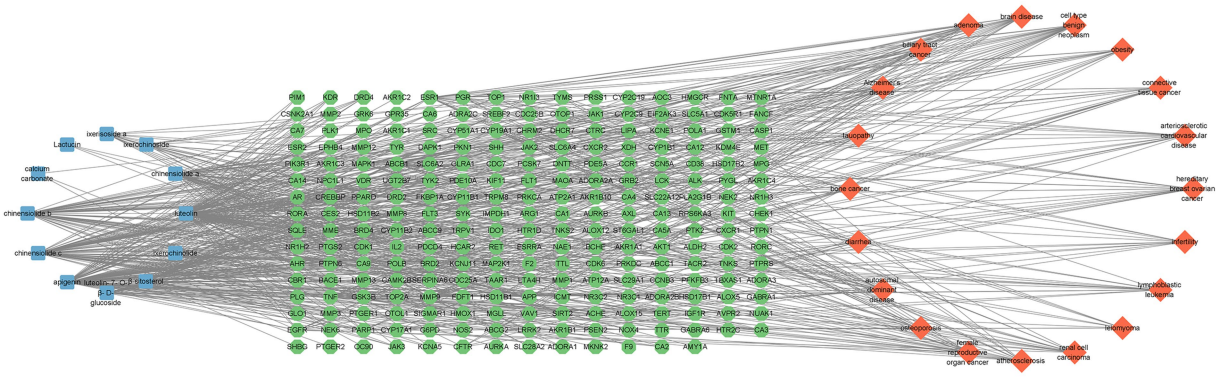

FIGURE 7
Ixeris chinensis component-disease-target map. In the figure, the left circle represented active chemical components, the middle square matrix represented target genes, and the right circle represented diseases.

with PTGS2 via amino acid residue ARG-44, two hydrogen bonds with AR via amino acid residues GLN-802 and GLU-687, and one hydrogen bond with EGFR via amino acid residue GLU-758 forming one hydrogen bond; luteolin forming four hydrogen bonds with MMP9 via amino acid residues ALA-189, ALA-417, GLU-402, and LEU-188; apigenin forming three hydrogen bonds with MMP9 via amino acid residues ALA-189, ALA-417, and LEU-188, and three hydrogen bonds with MMP9 via amino acid residues ARG-752, ASN-705, GLN-711, and LEU-873 to form four hydrogen bonds with AR, and chinensiolide b to form three hydrogen bonds with PTGS2 via amino acid residues HIS-388, TRP-387, and TYR-385. The high binding energy between the main active components and key targets

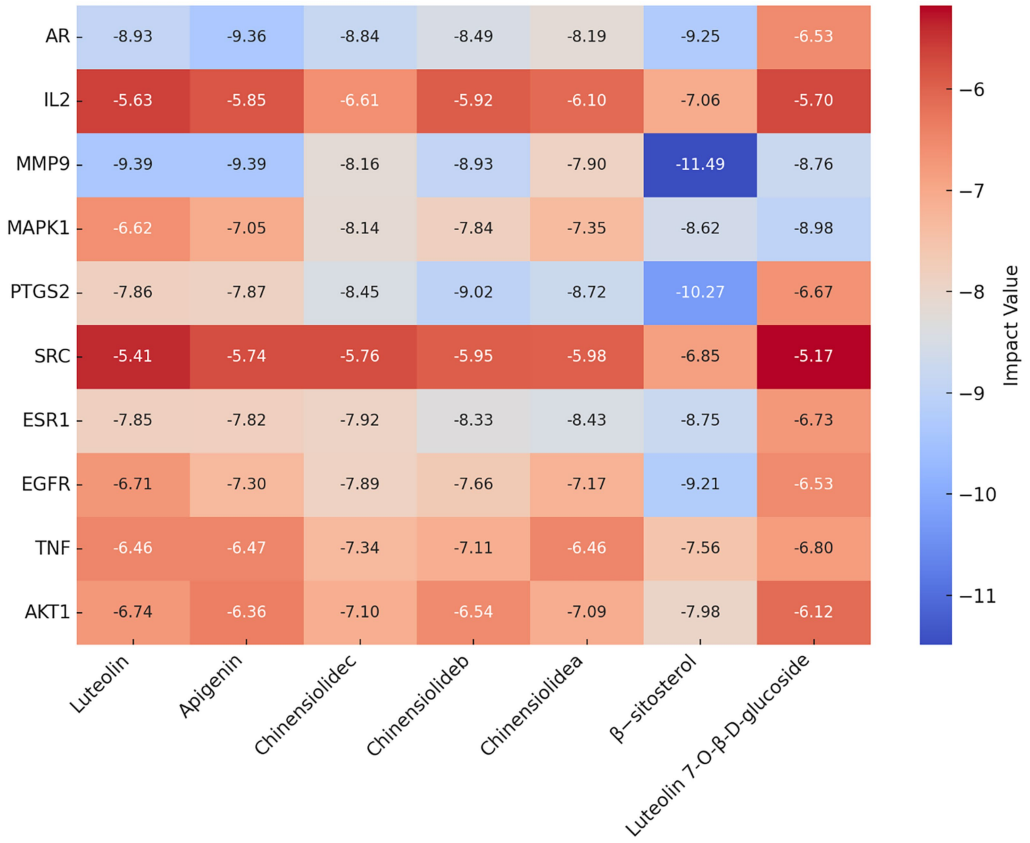

FIGURE 8
Docking and binding energy information of active ingredients and key target molecules in *Ixeris chinensis*. The redder the color, the better the binding ability.

in *I. chinensis* indicated that *I. chinensis* might play its therapeutic role by regulating the above related targets.

3.9 Meta analysis results

3.9.1 Literature search and screening results

The document screening process and basic characteristics of included documents are shown in the figure below (Figure 10 and Table 8). A total of five randomized controlled trials (RCTs) were conducted, including studies employing Sprague–Dawley (SD) rats, ICR mice, C57BL/6J mice, Wistar rats, and KM mice. For modeling, one study was fed with high-fat and high-sugar diets, two studies used carbon tetrachloride (CCl4), and two studies were fed with high-fat diets. The interventions were mainly treated with *I. chinensis* water extract. The diseases included in the studies mainly consisted of two cases of lipid-based diseases and three cases of hepatitis-based diseases. The included outcome indicators contained TC, TG, HDL-C, LDL-C, and MDA.

3.9.2 Document quality evaluation

The five research baselines included were comparable. In the Random sequence generation, four studies described the random allocation method in detail and were rated as low risk bias, while one study only mentioned the word "random" without specifying the specific method and was rated as unknown risk bias. In the Allocation consideration, only one described the allocation concealment scheme, which was rated as low risk bias, and the rest as unknown risk bias. Among Blinding, only one described the allocation concealment scheme, which was rated as low risk bias, and the rest as unknown risk bias. In the Incomplete outcome data, all studies reported complete results and were rated as low risk bias; In selective reporting, none of the five studies reported selectively, which was rated as low risk bias. In Other bias, five studies were unable to judge whether there were other sources of bias, and all were rated as unknown risk bias (Figure 10).

3.9.3 Statistical analysis results

Included in the study, three articles reported four changes in blood lipids (29–31) (Figure 11A). The above three articles were divided into four subgroups according to blood lipid indicators. Heterogeneity between studies ($I^2 > 50\%$) was analyzed by random effect model. Meta analysis results showed that *I. chinensis* could significantly reduce the blood TC level, [SMD = −1.85, 95% CI (−2.87, −0.38), $I^2 = 53\%$, $p = 0.0004$], indicating significant heterogeneity and statistically significant difference. *I. chinensis* could also significantly reduce the blood TG level, [SMD = −1.95, 95% CI (−3.02, −0.87), I2 = 51%, $p = 0.0004$], indicating significant heterogeneity and statistically significant difference. *I. chinensis* showed no statistically significant difference in improving HDL-C [SMD = 1.34, 95% CI (0.26, 2.42), $I^2 = 65\%$,

TABLE 8 The binding ability of the active components and key targets of *Ixeris chinensis*.

Gene	Luteolin	Apigenin	Chinensiolide c	Chinensiolide b	Chinensiolide a	β-sitosterol	Luteolin 7-*O*-*β*-D-glucoside
AKT1	−6.74	−6.36	−7.1	−6.54	−7.09	−7.98	−6.12
TNF	−6.46	−6.47	−7.34	−7.11	−6.46	−7.24	−6.8
EGFR	−6.71	−7.3	−7.89	−7.66	−7.17	−9.21	−6.53
ESR1	−7.85	−7.82	−7.92	−8.33	−8.43	−8.75	−6.73
SRC	−5.41	−5.74	−5.76	−5.95	−5.98	−6.85	−5.17
PTGS2	−7.86	−7.87	−8.45	−9.02	−8.72	−10.27	−6.67
MAPK1	−6.62	−7.05	−8.14	−7.84	−7.35	−8.62	−8.98
MMP9	−9.39	−9.39	−8.16	−8.93	−7.9	−11.49	−8.76
IL2	−5.63	−5.85	−6.61	−5.92	−6.1	−7.06	−5.7
AR	8.93	−9.36	−8.84	−8.49	−8.19	−9.25	−6.53

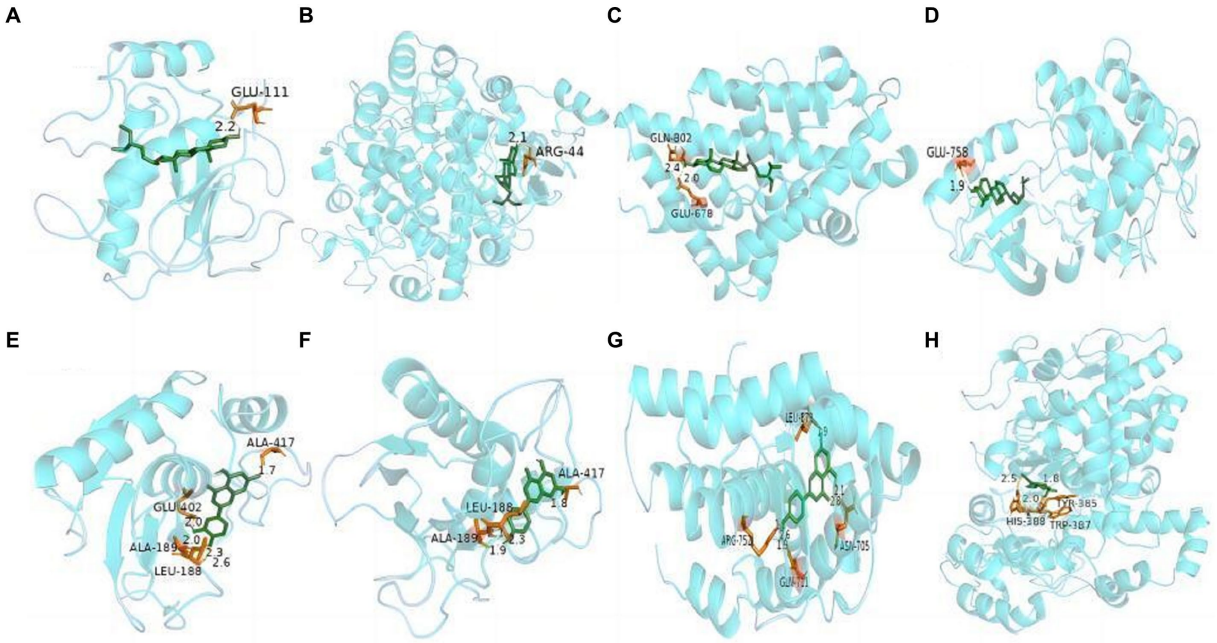

FIGURE 9
Docking analysis of the active components and key targets of *I. chinensis*. In the figure, the blue macromolecule is the target protein, and the green smaller molecule is the active component. **(A)** *β*-sitosterol and MMP9, **(B)** *β*-sitosterol and PTGS2, **(C)** *β*-sitosterol and AR, **(D)** *β*-sitosterol and EGFR, **(E)** luteolin and MMP9, **(F)** apigenin and MMP9, **(G)** apigenin and AR, **(H)** chinensiolide b and PTGS2.

$p = 0.02$]. *I. chinensis* significantly reduced blood LDL-C levels, [SMD = −2.77, 95% CI (−3.87, −1.66), $I^2 = 43\%$, $p < 0.00001$], suggesting that the difference was statistically significant. The sensitivity analysis of the included literatures was carried out one by one with the method of exclusion. None of the literatures had a great impact on the results of this Meta-analysis, which meant that this study has a good stability.

Among the included studies, two articles reported changes in MDA with heterogeneity between studies ($I^2 > 50\%$), analyzed using a random-effects model (32, 33) (Figure 11B). The difference in *I. chinensis* in improving HDL-C was not statistically significant [SMD = −9.76, 95% CI (−26.39, 6.86), $I^2 = 96\%$, $p = 0.25$].

3.9.4 Publication bias analysis

We used the four indicators of blood lipids as an example to create a funnel plot. Funnel diagram showed that the included articles were asymmetric with the symmetry axis as the boundary, indicating that there was publication bias, which might be related to the small sample size of the included articles (Figure 12).

4 Discussion

As a traditional Chinese herbal medicine, *I. chinensis* has complex types and various functions. It has been used as a medicinal material

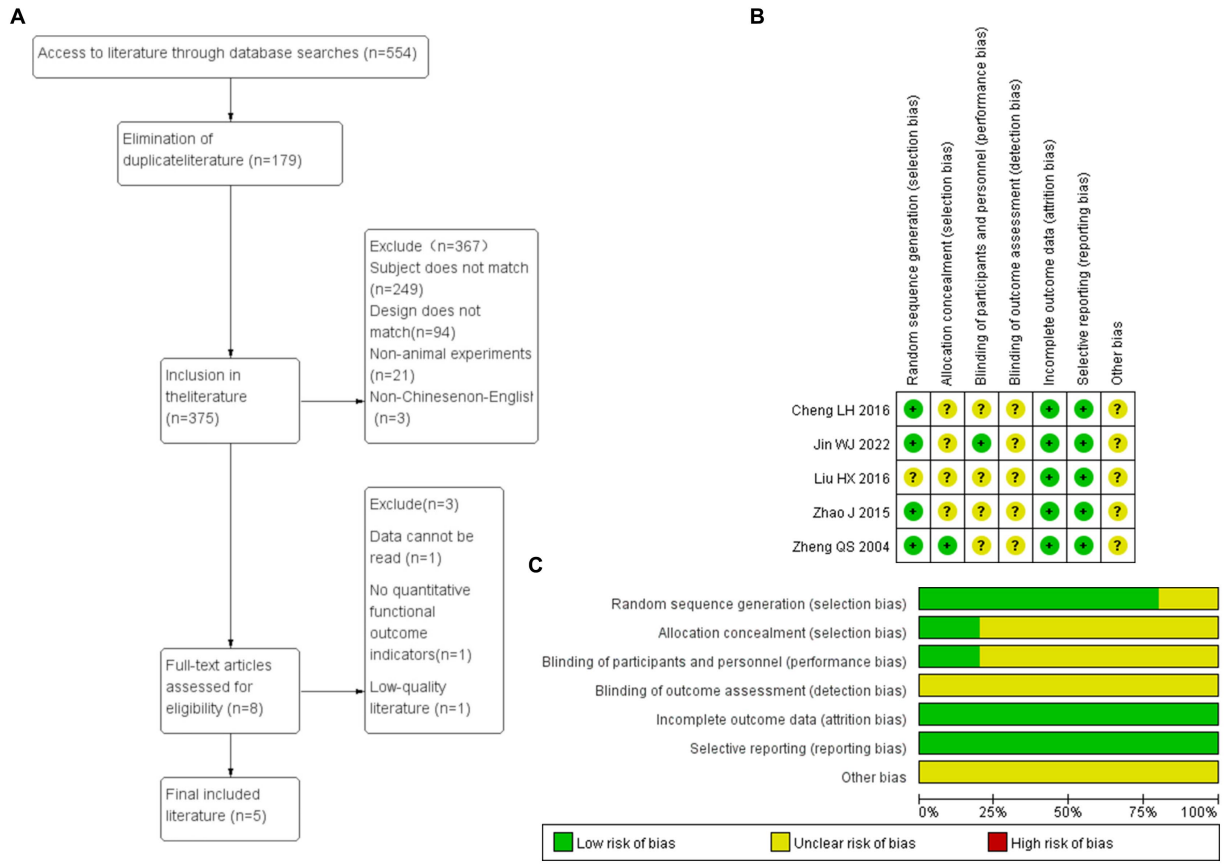

FIGURE 10
(A) Flow chart of document screening. **(B,C)** Risk assessment of literature bias. Green is low risk bias, yellow is unknown risk bias, and red is high risk bias. The length of the bar indicates the proportion of this bias in the total number of articles.

for lowering blood glucose, cholesterol, diuresis, bloodpressure, anticoagulation, liver protection, antiinflammatory, antibacterial, anti-tumor and so on (6). The emergence of traditional.

Chinese medicine network pharmacology provides an opportunity to systematically explore the molecular complexity of traditional Chinese medicine prescriptions and the molecular relationship between traditional Chinese medicine ingredients and complex diseases (34). In recent years, the pharmacological effects of *I. chinensis* have been widely studied, but the specific mechanism of disease action is still unclear and needs to be further studied. Therefore, we should start from the analysis of active components to explore the key role targets of *I. chinensis*, and carry out GO and KEGG analysis, conduct molecular docking which is of great significance to predict the biological process, cell composition, molecular function and signal pathway of *I. chinensis*, and has a profound influence on the research of specific disease mechanism and the development of new drugs.

As a traditional Chinese herbal medicine, there are many kinds of *I. chinensis*. Now we pay more attention to the pharmacodynamic material basis of *I. chinensis*, that is, the effective components in the curative effect, which requires us to screen and analyze the effective components in it. By searching "*I. chinensis* "in HERB database, the current study found 12 active ingredients of it, including apigenin, *β*-sitosterol, luteolin, luteolin-7-*O*-*β*-D-glucoside and so on. These active ingredients were used as the basis of the whole analysis of *I. chinensis*.

The active compounds of *I. chinensis* have many targets. The active compounds of *I. chinensis* were input into swiss target prediction for target gene prediction. The targets without corresponding gene names were deleted by Uniprot database, and the related targets were screened out. To explore the relationship between the active compounds of *I. chinensis* and the target sites, Cytoscape3.9.1 was used to construct the *I. chinensis* composition-target network diagram. There were 238 targets of *I. chinensis*. Two hundred and thirty-eight potential targets were input into STRING database for protein interaction analysis, and free targets were eliminated. Cytoscape 3.9.1 software was used to construct PPI network, and 40 key targets were screened out. They were mainly AKT1, TNF, EGFR, ESR1, SRC, PTGS2, MAPK1, which were regarded as important proteins of *I. chinensis*. It was suggested that it might play a therapeutic role in diseases through the interaction of active compounds with key targets. Most of the key targets were closely related to monocyte attachment and activation, injury and apoptosis, inflammation, matrix digestion, cell migration, differentiation, proliferation and other biological functions. They had similar biological effects and synergistic effects of multiple targets. Akt1 is an important gene regulating cell survival and proliferation, which can regulate tumor proliferation, metastasis and invasion. Transplanting mesenchymal stem cells overexpressing Akt1 into rat myocardium can reduce infarct size and delay cardiac remodeling (35). TNF is a substance that can damage tumor cells and make them necrotic. TNF-α plays a role in inducing cell apoptosis, regulating the immune response of the body, regulating the vascular system of tumor tissue, and inducing programmed cell

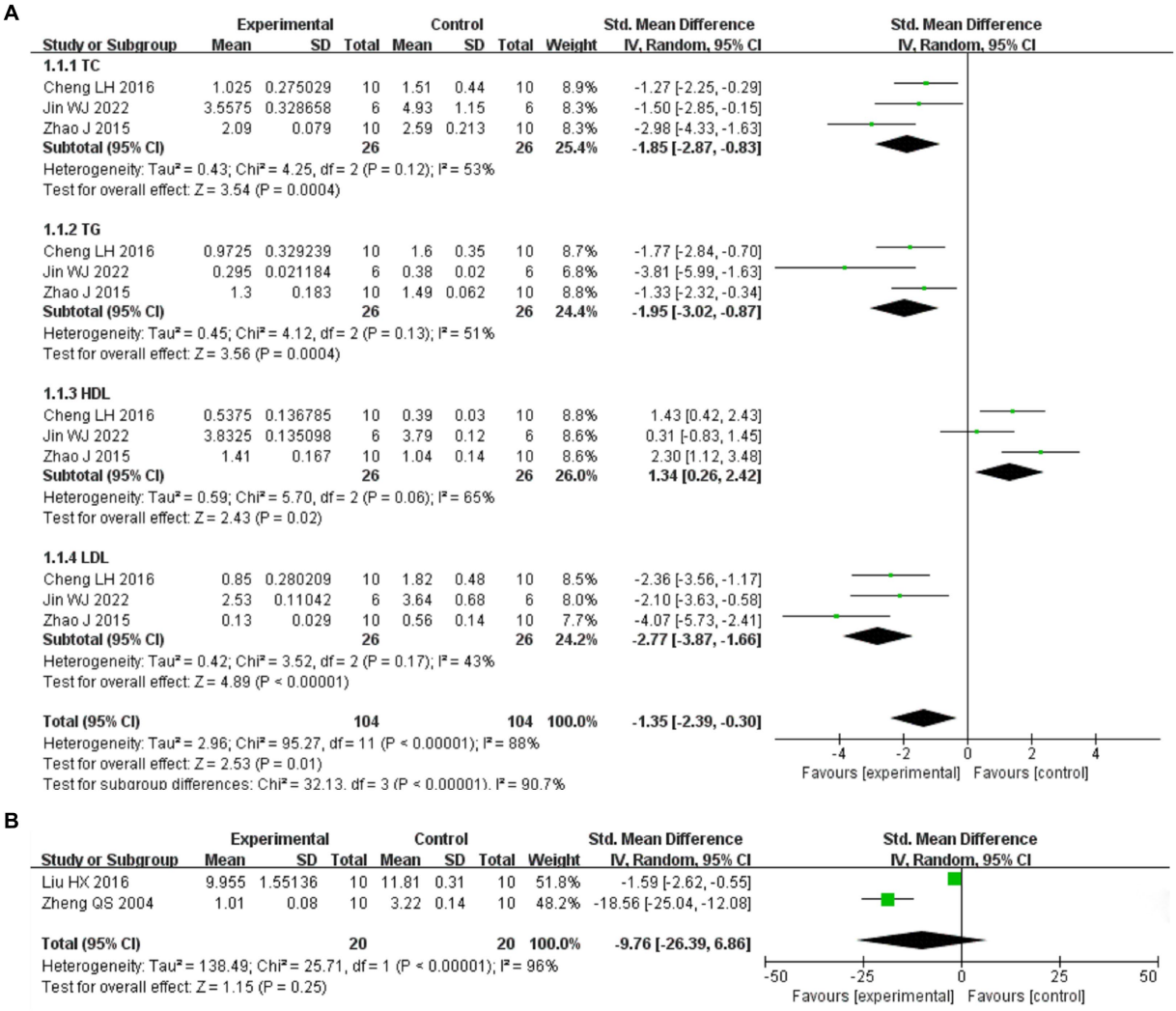

FIGURE 11
The study was grouped according to outcome indicators. The random effect model was used for analysis. Heterogeneity was calculated using I2. The summary estimates of the standard mean difference and 95% confidence zone were calculated using Review Manager 5.4.1. **(A)** Take the four items of blood lipid as the outcome index. **(B)** Take MDA as the outcome indicator.

necrosis. It plays different regulatory roles in many malignant tumors, such as gastric cancer, liver cancer, breast cancer, etc. (36). EGFR belongs to the epidermal growth factor receptor family, which is activated after binding with ligands. It transmits information through Ras/Raf/MEK/ERK/MAPK pathway, PI3K/AKT (PKB) pathway and JAK/STAT pathway, thus affecting tumor cell proliferation, angiogenesis, invasion and metastasis (37). Estrogen Receptor 1 (ESR1) expression is lost or diminished in human Hepatocellular carcinoma (HCC) cells and liver tumors, suggesting a potential protective role of estrogen signaling in HCC and that low ESR1 gene expression plays.

an important role in the development of hepatocellular carcinoma (38). The sparse representation based classifier gene is the earliest proto oncogene discovered by human beings. Its protein product, Src, is the most widely distributed protein kinase *in vivo*. It participates in a series of physiological activities such as regulating cell proliferation and differentiation by phosphorylating various signal molecules in cells (39). PTGS2 gene is an inducible immediate response gene, and its expression is rapidly upregulated by certain cytokines, growth factors, inflammatory mediators, pro-oncogenic factors and other stimulating factors during the occurrence of pathological responses such as inflammation or tumor, and its main product catalyzed by PG is an important inflammatory mediator in liver injury, which is an important link in inflammation (40). Mitogen-activated protein Kinase (MAPK) signal pathway is an important inflammation related signal pathway. Some studies have shown that MAPK1 has neuroprotective effect after stroke, and some studies have shown that MAPK1 has harmful effects on stroke due to its activation promoting inflammation and oxidative stress and inhibiting and reducing ischemic injury (41).

Comprehensive analysis of the results of "component-target" network and "component-target-disease network," The key components in *I. chinensis* were apigenin, chinensiolide c, chinensiolide a, chinensiolide b, luteolin, luteolin-7-*O*-β-D-glucoside and β-sitosterol, which were dominant in the network. According to relevant literature reports, apigenin, luteolin, luteolin-7-*O*-β-D-glucoside belong to flavonoids (6, 42–44), chinensiolide c, chinensiolide a and chinensiolide b belong to sesquiterpenes (45)

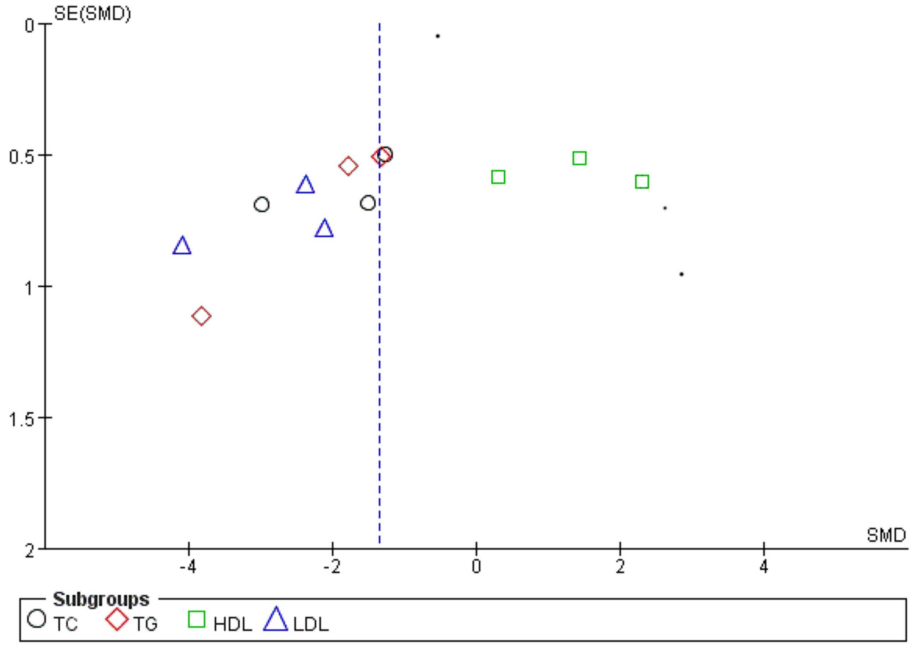

FIGURE 12
Funnel chart analysis of publication bias with four blood lipids as outcome indicators.

, β-sitosterol belongs to triterpenes and steroids (6). Flavonoids have the advantages of scavenging free radicals, protecting liver, protecting heart and brain, regulating blood lipids and so on (46). Sesquiterpenes have a wide range of biological activities, such as anti-tumor, antibacterial, anti-inflammatory, anti-neurotoxic, antiviral, immunosuppressive, hepatoprotective and heart-strengthening activities (47). Triterpenes and steroids have anti-inflammatory, antibacterial and anti-tumor effects (48). Studies have shown that apigenin can inhibit the proliferation of colorectal cancer CL187 cells and promote apoptosis by inhibiting PI3K/Akt signaling pathway and regulating the expression of MAPK signaling pathway related proteins (49). Apigenin can inhibit the EMT of liver tissue cells by inhibiting PDK1/AKT signal pathway through liver bypass, and play an anti-fibrosis role (50). Luteolin can eliminate free radicals, enhance the activity of antioxidant enzymes, regulate proinflammatory mediators, and inhibit IκB kinase β Phosphorylation of subunits, down regulating TNF-α and the mRNA level of IL-6 (51). Luteolin-7-*O*-β-D-glucoside has protective effect on myocardial cells cultured under ischemia and hypoxia. Its mechanism may be related to scavenging oxygen free radicals, stabilizing cell membrane and inhibiting apoptosis (52). The water extract of *I. chinensis* A can effectively inhibit the growth of lung adenocarcinoma A549 cells, liver cancer Blx10-7402 cells and LoVo cells *in vitro*, and has strong anti-tumor activity (53). The sesquiterpene lactones, chinensiolide A-C, show strong anti-inflammatory and anti-tumor activities (54, 55). The pharmacological studies of the above active ingredients were in accordance with the results of *I. chinensis* network analysis, which was speculated to be an important component of *I. chinensis* to exert drug effects, and to a certain extent, it provided the value of medical research for the treatment of diseases.

Molecular docking technology can analyze the optimal binding sites between active components and targets, providing valuable insights into the mechanisms of drug action in treating diseases. Lin Shenghua and others (56) conducted molecular docking on the top 6 targets and the top 3 active components of Xuefu Zhuyu oral liquid for anti-thrombotic activity. The binding energies were between −5 and −9.5KJ/mol, indicating that the active components of Xuefu Zhuyu oral liquid effectively bind with disease-related targets. Through molecular docking technology, the docking results between the main active components of 7 kinds of *I. chinensis* found in this study and key targets were pretty good, with binding energies concentrated between −7 and -9KJ/mol, indicating that the main active components of *I. chinensis* and key targets have good affinity, providing a basis for future drug design.

GO and KEGG analysis revealed that the biological processes of *I. chinensis* were mainly enriched in positive regulation of transcription from RNA polymerase II promoter, signal transduction, response to xenobiotic stimulus, positive regulation of gene expression, positive regulation of transcription, DNA-templated, response to drug, negative regulation of transcription from RNA polymerase II promoter, protein phosphorylation, negative regulation of apoptotic process, peptidyl-serine phosphorylation, etc. The potential action pathways of *I. chinensis* included 103 pathways of Pathways in cancer, Chemical carcinogenesis-receptor activation, PI3K-Akt signaling pathway, Hepatitis B, Proteoglycans in cancer, Lipid and atherosclerosis, etc. AKT1, TNF, EGFR and other proteins that regulated tumor cell proliferation, migration, apoptosis, inflammatory expression, immune response, MAPK1, MMP9, ESR1 and other proteins participated in nerve cell regulation, brain tissue repair, and regulation of vascular function were also at the center of the PPI network, indicating that *I. chinensis* may play a role in promoting tumor cell death, regulating blood lipids, and protecting the heart and brain by regulating these targets. In addition, the active compounds of *I. chinensis* could regulate inflammatory pathways, and the pathways with a high degree of significance include TNF

signaling pathway, PI3K/AKT signaling pathway, and Estrogen signaling pathway, and the inflammatory response was related to the occurrence and development of diseases. Tumor necrosis factor-α (TNF-α) is a cytokine involved in systemic inflammation and is also among the many cytokines that cause acute phase responses. Extracellular TNF- α bound to the TNF-R1 receptor on the cytosolic membrane and indirectly activates CASP8, which in turn activated CASP3, and thus had an impact on apoptosis and plaque instability (57, 58). In addition, lignans and lignans 7-*O*-*β*-D-glucoside scavenged free radicals and enhanced antioxidant enzyme activity. It was evident that the active compounds of *I. chinensis* cought have myocardial protective and antiatherosclerotic effects by modulating the above-mentioned inflammatory pathways (59).

The DO analysis found that, *I. chinensis* could be applied to tumors such as hereditary breast, ovarian cancer, cell type benign neoplasm, female reproductive organ cancer, adenoma, biliary tract cancer, renal cell carcinoma, bone cancer, connective tissue cancer. It could also be applied to diseases such as autosomal dominant disease, brain disease, diarrhea, atherosclerosis, arteriosclerotic cardiovascular disease, leiomyoma, osteoporosis, Alzheimer's, tauopathy, lymphoblastic leukemia, obesity, infertility, etc. It could be seen that the *I. chinensis* could be mainly used to treat cancer related diseases, visceral system diseases, reproductive system diseases, nervous system diseases, musculoskeletal system diseases, lymphatic system diseases, etc.

In conclusion, network pharmacology research can shorten the research process of molecular targets (60). Researchers can use network pharmacology studies to help discover effective and specific compounds for treating diseases (61). Based on meta analysis, network pharmacology and molecular docking methods, this study preliminarily explored the active components, key targets, docking binding energy, signal pathways and disease prediction of *I. chinensis*. It further verified the feasibility of *I. chinensis* as a drug to treat many diseases, and provided a basis for exploring the specific mechanism in treating diseases and developing new drugs. At the same time, current study found that in the five articles included, the extract of *I. chinensis* had a good effect on the improvement of blood lipid indicators. However, due to the small sample size, general literature quality, and fewer types of diseases studied, the reliability of outcomes had been affected to some extent, so more *in vitro* and *in vivo* experiments are needed to verify in the future.

Author contributions

ZN: Resources, Writing – original draft. ZM: Writing – original draft, Conceptualization. XQ: Data curation, Writing – original draft. YG: Formal analysis, Software, Writing – review & editing. CR: Methodology, Project administration, Writing – original draft, Writing – review & editing. YW: Resources, Writing – original draft. YY: Investigation, Writing – original draft.

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 National Natural Science Foundation of China (82201627), the Military Medicine Upgrade Program of Air Force Military Medical University (2020SWAQ04), Shaanxi Provincial Innovation Capacity Support Programme (2023-CX-PT-33), Shaanxi Provincial Natural Science Basic Research Programme (2024JC-ZDXM-60, 2022JQ-820), Shaanxi Provincial Natural Science Basic Research Programme Key Projects (2024JC-ZDXM-51), and Xijing Hospital Clinical New Technology (2023XJSY27).

Conflict of interest

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

Publisher's note

References

1. Tai WJ, Chen SK, Gan XH. Flavonoid extraction from *Ixeris chinensis* (Thunb) Nakai and its antioxidant activity with DPPH. *J Guizhou Norm Univ.* (2016) 32:59–63. doi: 10.13391/j.cnki.issn.1674-7798.2016.12.014

2. Cui GZ, Li F, Liu Y. Literature research of the sow thistle. *Chin J Experiment Formulas*. (2012) 18:360–2. doi: 10.13422/j.cnki.syfjx.2012.23.021

3. Hou YG, Wang F, Li HY, Tian C, Yang L. Optimization on the extraction conditions of flavonoids from *Lactuca Tatarica* using response surface methodology. *J Liaoning Univ*. (2019) 39:21–6. doi: 10.3969/j.issn.1672-6952.2019.02.004

4. Song YX, Chen YQ, Min YT. Purification of flavonoids from variegated *Picris hieracioides* using macroporous resin and their antioxidant activity study. *Food Sci Technol.* (2023) 48:154–60. doi: 10.13684/j.cnki.spkj.2023.09.016

5. Xu Z, Jin YT, Zhao ZY, Wang Y, Gong BY, Zhou HF, et al. Study on the compatibility rules and action mechanism of Chinese herbal compound in the treatment of ulcerative colitis based on national patent database and network pharmacology. *Tianjin J Traditional Chin Med*. (2023) 40:219–27. doi: 10.11656/j.issn.1672-1519.2023.02.16

6. Wang ZM. *Anti-inflammatory effects and mechanism of water extract of Sonchus oleraceus*. Shijiazhuang: Hebei Medical University (2015).

7. Li MY, Chao B, Zhou T, Zhang JP. Effects of *Sonchus oleraceus* on inflammatory factors and HMGB1_TLR4_NF-κB signaling pathway in mice with pneumonia. *J Chin Med Materials*. (2023) 46:911–7. doi: 10.16431/j.cnki.1671-7236.2023.12.044

8. Cao Y, Wang P, Liu S, Yang H. Exploration of the mechanism of action of modified Zhixiao Pingchuan decoction in the treatment of bronchial asthma based on network pharmacology. *China Med Pharmacy*. (2024) 14:129–32. doi: 10.20116/j.issn2095-0616.2024.05.29

9. Zhou L, Liu MH. Exploration of the mechanism of action of honeysuckle in the treatment of 2019-nCoV infection with myocarditis based on network pharmacology. *Hunan J Trad Chin Med*. (2024) 40:160–7. doi: 10.16808/j.cnki.issn1003-7705.2024.03.034

10. Xie J, Gao S, Li L, Xu YL, Gao SM, Yu CQ. Research progress and application strategy on network pharmacology in Chinese materia medica. *Chin Tradit Herb Drug*. (2019) 50:2257–65. doi: 10.7501/j.issn.0253-2670.2019.10.001

11. Fu K. *Prediction and screening of active components in Chinese medicine for SLE treatment using semi-flexible molecular docking*. Nanchang, Jiangxi: Jiangxi University of Traditional Chinese Medicine (2019).

12. Ding J, Zheng P, Sun YY, Wang XR, Feng YC. Analysis of Wumei pills in treating chronic digestive system diseases with concept of "treating different diseases with same method" based on network pharmacology and molecular docking. *China J Chinese Materia Medica*. (2022) 47:4164–76. doi: 10.19540/j.cnki.cjcmm.20220412.401

13. Han KY, Dang H, Zhang H. Efficacy of repetitive transcranial magnetic stimulation on cognitive dysfunction after traumatic brain injury: a meta-analysis. *Chin J Rehabil Theory Pract*. (2021) 27:1282–90. doi: 10.3969/j.issn.1006-9771.2021.11.007

14. Li QH, Ju LW, Wang N, Liu ZZ. Analysis of the effect of ginsenoside Rh1 on gene expression in breast cancer SKBR3 cells by KEGG biological pathway enrichment analysis. *Chinese J Med*. (2019) 25:36–41. doi: 10.13862/j.cnki.cn43-1446/r.2019.20.010

15. Guo P, Guan LK. The significance of JAK1 signal pathway in cardiovascular and cerebrovascular diseases. *Jilin Med J*. (2017) 38:175–8. doi: 10.3969/j.issn.1004-0412.2017.01.071

16. Zheng J, Li XS, Peng F, Luo J, Tang XH. Chitosan modulates neointimal hyperplasia in arteriovenous fistula models in rabbits by downregulating serum MCP-1 and VEGF-A expression. *Chin J Gerontol*. (2023) 43:336–41. doi: 10.3969/j.issn.1005-9202.2023.03.033

17. Minamino T, Yoshida T, Tateno K, Miyauchi H, Zou Y, Toko H, et al. Ras induces vascular smooth muscle cell senescence and inflammation in human atherosclerosis. *Circulation*. (2003) 108:2264–9. doi: 10.1161/01.CIR.0000093274.82929.22

18. Yang H, Zeng CP. Effects of water extract from *Scorzonera sinensis* Lipsch，*Pteridium aquilinum* and *Sonchus oleraceus* L. on plasma-lipids metabolism in mice fed high fats diet. *Food Res Develop*. (2015) 36:11–3. doi: 10.3969/j.issn.1005-6521.2015.06.003

19. Lou CX, Wang Y, Dai LL, Long SX, Huang XQ. Effects of three traditional Chinese medicine ingredients on oxidized low-density lipoprotein-induced injury in human aortic endothelial cells. *China Pharmac*. (2024) 33:39–42. doi: 10.3969/j.issn.1006-4931.2024.04.010

20. Ni Q, Sun P. Research Progress on the correlation between LOX-1 and Tumors. *Advances Clin Med*. (2023) 13:8319–24. doi: 10.12677/acm.2023.1351164

21. Wang MQ, Liu KK, Liu XY, Zhang XL, Li HX. Research progress on inhibitors of the PI3K/Akt signaling pathway in the treatment of gynecological malignant tumors. *World J Cancer Res*. (2024) 14:35–40. doi: 10.12677/WJCR.2024.141006

22. Liu QB, Li HW. Advanced glycation end products and atherosclerosis. *Chin J Cardiovasc Diseases*. (2018) 23:87–91. doi: 10.3969/j.issn.1007-5410.2018.01.020

23. Cao NB, Yu WX. *Role of Connexin 43 in oxidative damage of periodontal tissues through regulation of the JNK/NF-κB signaling pathway*. Changchun: School of Stomatology, Jilin University (2023).

24. Chen G. *The mechanism of mmLDL impaired endothelium-dependent relaxation in mice mesenteric arteries*. Hengyang: University of South China (2014).

25. Chu MY, Lu YT, Liu T, Liu J. Expression and clinical application prospects of tumor necrosis factor alpha in autoimmune diseases. *J Hebei Med Univ*. (2020) 41:853–63. doi: 10.3969/j.issn.1007-3205.2020.03.027

26. Zhu BJ, Zhou XD. Study of the PI3K/AKT pathway in lung cancer metastasis and drug resistance. *Chin J Lung Cancer*. (2011) 14:689–94. doi: 10.3779/j.issn.1009-3419.2011.08.10

27. Krishnankutty A, Kimura T, Saito T, Aoyagi K, Asada A, Takahashi S, et al. In vivo regulation of glycogen synthase kinase 3β activity in neurons and brains. *Sci Rep*. (2017) 7:8602. doi: 10.1038/s41598-017-09239-5

28. Mohammed MH, Hamed AN, Sayed AM, Afifi AH, Rateb ME, Thissera B, et al. Antiproliferative potential of *Dypsis decaryi* seeds supported by metabolic profiling and molecular docking (J/OL). *J Herbal Med*. (2024) 44:100846. doi: 10.1016/j.hermed.2024.100846

29. Cheng LH. *The study of the preparation of Ixeris Chinensis (thumb.) Nakai total flavonoids extract and the effect and mechanism to T2DM*. Haerbin: Heilongjiang University Of Chinese Medicine (2016).

30. Jin W, Cho S, Laxi N, Bao T, Dai L, Yu H, et al. Hepatoprotective effects of *Ixeris chinensis* on nonalcoholic fatty liver disease induced by high-fat diet in mice: an integrated gut microbiota and metabolomic analysis. *Molecules*. (2022) 27:3148. doi: 10.3390/molecules27103148

31. Zhao J. *Formulation of plant milk-tea products and comparing of its lipid-lowering activity*. Changchun: Jilin Agricultural University (2015).

32. Zheng QS, Sun XL, Xu B, Song M, Wang CH. Protective effects of luteolin-7-glucoside against liver injury caused by carbon tetrachloride in rats. *Pharmazie*. vol. 5 (2004).

33. Liu HX, Pei XP, Pei MR, Li HF, Xue YM. Research on anti-inflammatory and hepatoprotective activities between ixeris chinensis and s.brachyotus. Taiyuan: Journal of Shanxi University of Traditional Chinese Medicine (2016) 17:19–20+56.

34. Meng FC, Tang LD. Challenges and prospect in research of Chinese materia medica network pharmacology. *Chin Tradit Herb Drug*. (2020) 51:2232–7. doi: 10.7501/j.issn.0253-2670.2020.08.034

35. Huang XF, Yu GZ, Tong JJ. Analysis of the pharmacological mechanism of action of Chen pi based on network pharmacology. *Chin Trad Patent Med*. (2019) 41:3038–45. doi: 10.3969/i.iSsn.1001-1528.2019.12.043

36. Gao SY, Li D. Advances in tumor necrosis factor and cancer-related research. *Chin Pharm Bull*. (2020) 36:1209–13. doi: 10.3969/j.issn.1001-1978.2020.09.006

37. Huang JX, Wang H. Targeted therapy and mechanism of drug resistance in non-small cell lung Cancer with epidermal growth factor receptor gene mutation. *Chin J Lung Cancer*. (2022) 25:183–92. doi: 10.3779/j.issn.1009-3419.2022.101.05

38. Han XW. *Expression and significance of ESR1 and ACOX2 in hepatocellular carcinoma*. Hefei: Anhui Medical University (2021).

39. Bin W. *Expression of Src gene in glioma and its effect on the malignant biological behavior of glioma cell*. Shanghai: Naval Medical University (2016).

40. Liu CL. *Association between polymorphisms of PTGS2 gene and Clinicopathologic characteristics and prognosis to hepatocellular carcinoma*. Nanning: Guangxi Medical University (2013).

41. Wu YL. *Association of Mitogen-activated protein kinase 1 (MAPK1) and mitogen-activated protein kinase 4(MAP2K4)gene polymorphism with risk of ischemic stroke and IS-related traditional Chinese medicine syndrome*. Nanning: Guangxi University of Chinese Medicine (2017).

42. Liu C, Yan MM, Shao S, Wu CY, Fu ML, Tian S, et al. Chemical constituents and pharmacological effects of *Ixeris chinensis*. *Chin J Exp Tradit Med Formulae*. (2015) 21:231–4. doi: 10.13422/j.cnki.syfjx.2015200231

43. Chen L, Lin X, Xiao J, Tian Y, Zheng B, Teng H. *Sonchus oleraceus* Linn protects against LPS-induced sepsis and inhibits inflammatory responses in RAW264.7 cells. *J Ethnopharmacol*. (2019) 236:63–9. doi: 10.1016/j.jep.2019.02.039

44. Chen L, Lin X, Fan X, Qian Y, Lv Q, Teng H. *Sonchus oleraceus* Linn extract enhanced glucose homeostasis through the AMPK/Akt/GSK-3β signaling pathway in diabetic liver and HepG2 cell culture. *Food Chem Toxicol*. (2020) 136:111072. doi: 10.1016/j.fct.2019.111072

45. Wang JL, Lv D, Liang XY, Zhao M, Zhang SJ. Study on Chemical Constituents of the Ixeris chinensis. *J Chin Med Mater*. Qiqihar University 34: 1706–1708. (2011).

46. Wang HY, Ben LF, Jiao LH, Ma Y, Su XH. Research status of Total flavonoids from *Sonchus Oleraceus* L. *Shandong Chem Industry*. (2018) 47:64–5. doi: 10.19319/j.cnki.issn.1008-021x.2018.21.024

47. Guo LM, Lv JL, Zhang LB. Research progress on anti-inflammatory mechanism of natural sesquiterpenoids. *China J Chin Materia Medic*. (2018) 43:3989–99. doi: 10.19540/j.cnki.cjcmm.20180726.013

48. Lan GC, Wu ZL, Wang SM, Wu ZL, Wang SM, Li HL, et al. A new triterpenoid and a new steroid from roots of *Vladimiria souliei*. *Chin Tradit Herb Drug*. (2019) 50:793–7. doi: 10.7501/j.issn.0253-2670.2019.04.001

49. Lin S, Qin HZ, Deng LY, Li ZY, Xie FF, Zhang M, et al. Apigenin induces apoptosis of human colorectal Cancer CL187 cells via PI3K/Akt and MAPK Signaling pathways. *Chin J Exp Tradit Med Formulae*. (2022) 28:97–104. doi: 10.13422/j.cnki.syfjx.20221122

50. Chen XD, Zhong WL, Yan PY, Kuang JN, Qiao KL, Sun T, et al. Apigenin inhibits mouse liver fibrosis via PDK1/AKT signaling pathway. *Chinese Pharmacol Bull*. (2022) 38:1010–6. doi: 10.12360/CPB202106091

51. Han NX, Sun YL, Sheng S, Zhang W. Mechanisms of lignocaine regulation of oxidative stress and inflammation. *Chin J Anim Nutr*. (2022) 34:2856–61. doi: 10.3969/j.issn.1006-267x.2022.05.015

52. Zhou L, Xie YY, Li J, Wang XJ, Mou YL. Protective effects of Luteolin-7-O-β-D-glucopyranoside on hypoxic-ischemic injury of myocardial cells in neonatal rats. *Trad Chin Drug Res Clin Pharmacol*. (2008) 4:259–61. doi: 10.19378/j.issn.1003-9783.2008.04.006

53. Zhou L, Zhao Y, Wang Y, Li LB, Zhang SJ, Li T. In vitro anti-tumor activity of extract Chinensiolide a from the *Ixeris Chinensis*. *Chin Med Innov*. (2015) 12:109–11. doi: 10.3969/j.issn.1674-4985.2015.34.034

54. Zhang S, Zhao M, Bai L, Hasegawa T, Wang J, Wang L, et al. Bioactive Guaianolides from Siyekucai (Ixerischinensis). *J Nat Prod*. (2006) 69:1425–8. doi: 10.1021/np068015j

55. Zhang S, Wang J, Xue H, Deng Q, Xing F, Ando M. Three new Guaianolides from Siyekucai (*Ixeris chinenseis*). *J Nat Prod*. (2002) 65:1927–9. doi: 10.1021/np0202632

56. Lin SH, Xue C, Ma HL, Fan W, Shen CM, Chen JY, et al. Revealing the active components and mechanisms of action of Xuefu Zhuyu oral liquid against thrombosis based on network pharmacology and molecular docking technology. *Shandong Sci*. (2023):1–8. doi: 10.3976/j.issn.1002-4026.20230130

57. Qing P, Zhao JF, Zhao CH, Hu J, Lin YL, He KJ. Effect of acupuncture on cancer-related fatigue and its impact on serum levels of CRP, IL-6, TNF-α, and sTNF-R1. *Chin Acupuncture Moxibust*. (2020) 40:505–7. doi: 10.13703/j.0255-2930.20190423-k0002

58. Wang L, Du F, Wang X. TNF-alpha induces two distinct caspase-8 activation pathways (J/OL). *Cell*. (2008) 133:693–703. doi: 10.1016/j.cell.2008.03.036

59. Jang WY, Kim M-Y, Cho JY. Antioxidant, anti-inflammatory, anti-menopausal, and anti-Cancer effects of Lignans and their metabolites. *Int J Mol Sci*. (2022) 23:15482. doi: 10.3390/ijms232415482

60. Mohammed MHH, Hamed ANE, Elhabal SF, Mokhtar FA, Abdelmohsen UR, Fouad MA, et al. Metabolomic profiling and anti-proliferative activities of Dypsis leptocheilos ethanol extract and its green synthesized silver nanoparticles supported by network pharmacology analysis. *S Afr J Bot*. (2023) 161:648–65. doi: 10.1016/j.sajb.2023.08.026

61. Mohammed MHH, Hamed ANE, Elhabal SF, Abdelmohsen UR, Fouad MA, Kamel MS. Chemical composition and anti-proliferative activities of Hyophorbe lagenicaulis aerial parts and their biogenic nanoparticles supported by network pharmacology study. *S Afr J Bot*. (2023) 156:398-410. doi: 10.1016/j.sajb.2023.03.018

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 27 September 2024
DOI 10.3389/fmed.2024.1462649

OPEN ACCESS

EDITED BY
J. Francis Borgio,
Imam Abdulrahman Bin Faisal University, Saudi Arabia

REVIEWED BY
Wen-Jing Wang,
Beijing Genomics Institute (BGI), China
Andrey Glotov,
D. O. Ott Research Institute of Obstetrics, Gynecology and Reproductology, Russia
Nicoletta Di Simone,
Humanitas University, Italy

*CORRESPONDENCE
Ying Cui
✉ cuiying388@126.com
Yang Zou
✉ zouyang81@163.com

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

RECEIVED 10 July 2024
ACCEPTED 17 September 2024
PUBLISHED 27 September 2024

CITATION
Ye Q, Liu F-Y, Xia X-J, Chen X-Y, Zou L, Wu H-M, Li D-D, Xia C-N, Huang T, Cui Y and Zou Y (2024) Whole exome sequencing identifies a novel mutation in *Annexin A4* that is associated with recurrent spontaneous abortion.
Front. Med. 11:1462649.
doi: 10.3389/fmed.2024.1462649

Whole exome sequencing identifies a novel mutation in *Annexin A4* that is associated with recurrent spontaneous abortion

Qian Ye[1,2,3†], Fa-Ying Liu[2,3,4†], Xiao-Jian Xia[1,2,3], Xiao-Yong Chen[1,2,3], Li Zou[5], Hui-Min Wu[6], Dan-Dan Li[6], Chen-Nian Xia[6], Ting Huang[6], Ying Cui[1,2,3]* and Yang Zou[2,3,4]*

[1]Department of Traditional Chinese Medicine, Jiangxi Maternal and Child Health Hospital, Nanchang, China, [2]Key Laboratory of Women's Reproductive Health of Jiangxi Province, Jiangxi Maternal and Child Health Hospital, Nanchang, China, [3]Key Research Unit of Female Reproduction with Integrated Chinese and Western Medicine of Jiangxi Province, Jiangxi Maternal and Child Health Hospital, Nanchang, China, [4]Central Laboratory, Jiangxi Maternal and Child Health Hospital, Nanchang, China, [5]Quality Control Office, Ganzhou People's Hospital, Ganzhou, China, [6]Graduate School of Clinical Medicine, Jiangxi University of Traditional Chinese Medicine, Nanchang, China

Background: Recurrent spontaneous abortion (RSA) is a multifactorial disease, the exact causes of which are still unknown. Environmental, maternal, and genetic factors have been shown to contribute to this condition. The aim of this study was to investigate the presence of mutations in the *ANXA4* gene in patients with RSA.

Methods: Genomic DNA was extracted from 325 patients with RSA and 941 control women with a normal reproductive history for whole-exome sequencing (WES). The detected variants were annotated and filtered, and the pathogenicity of the variants was predicted through the SIFT online tool, functional enrichment analyses, Sanger sequencing validation, prediction of changes in protein structure, and evolutionary conservation analysis. Furthermore, plasmid construction, Western blotting, RT–qPCR, and cell migration, invasion and adhesion assays were used to detect the effects of *ANXA4* mutations on protein function.

Results: An *ANXA4* mutation (p.G8D) in 1 of the 325 samples from patients with RSA (RSA-219) was identified through WES. This mutation was not detected in 941 controls or included in public databases. Evolutionary conservation analysis revealed that the amino acid residue affected by the mutation (p.G8D) was highly conserved among 13 vertebrate species, and the SIFT program and structural modeling analysis predicted that this mutation was harmful. Furthermore, functional assays revealed that this mutation could inhibit cell migration, invasion and adhesion.

Conclusion: Our study suggests that an unreported novel *ANXA4* mutation (p.G8D) plays an important role in the pathogenesis of RSA and may contribute to the genetic diagnosis of RSA.

KEYWORDS

***ANXA4* mutation, recurrent spontaneous abortion, cell invasion, migration, adhesion**

1 Introduction

Recurrent spontaneous abortion (RSA) is a common complication in women of childbearing age during pregnancy, with an incidence rate ranging from 1 to 5% (1).

The number of miscarriages required for a diagnosis of RSA is controversial because the risk of another miscarriage after two consecutive miscarriages is similar to that after three consecutive miscarriages, and recent studies have suggested that predicting pregnancy loss after two consecutive miscarriages is more reasonable (2, 3). The American Reproductive Association defines recurrent spontaneous abortion as two or more consecutive miscarriages (4). Obviously, the pathogenesis of RSA is complex. To date, the potential risk factors for RSA reported include genetic, immunologic, anatomic, and infectious anomalies (5–10). However, approximately 50% of RSAs are unexplained or poorly understood (11).

For almost all human diseases, individual susceptibility is influenced by genetic variations to some degree (12). In recent years, whole-genome sequencing (WGS) and whole-exome sequencing (WES) have proven to be powerful new approaches for identifying disease-associated variants across the full minor allele frequency (MAF) spectrum in animals and humans (13–15). Genome-wide screening has shown that siblings of patients with RSA have a greater risk of miscarriage (16), Variants associated with RSA have been detected in genes involved in the immune response (*IFNG, IL10, KIR2DS2, KIR2DS3, KIR2DS4, MBL,* and *TNF*), coagulation (*F2, F5, PAI-1,* and *PROZ*), metabolism (*GSTT1* and *MTHFR*), hormonal regulation (*ESR1* and *ADRB2*), angiogenesis (*NOS3* and *VEGFA*), and endometrial and placental function (*ENOS* and *ACE*) (17–19). Moreover, several studies have shown that rare variants play crucial roles in multiple human diseases, including RSA (20–22). However, few studies have confirmed that rare variants may play certain roles in the pathogenesis of RSA at the whole-exome/genome scale.

Annexin A4 (*ANXA4*) is a cytosolic calcium-binding protein with four repeat domains, each containing one calcium-binding site (CBS), and belongs to a ubiquitous family of Ca^{2+}-dependent membrane-binding proteins thought to be involved in membrane trafficking and membrane organization within cells (23). Previous studies have shown that *ANXA4* modulates membrane permeability and membrane trafficking, participates in cellular growth and apoptosis, and enhances tumor invasion (24). *ANXA4* has been shown to be upregulated in various clinical epithelial tumors, including colorectal, pancreatic, ovarian, breast, and prostate cancers (24). Furthermore, a previous study revealed that the *ANXA4* protein is localized to the glandular and luminal epithelium of the human endometrium and is present at high levels throughout the menstrual cycle, except during the early secretory (ES) phase (25).

In this study, we used WES to explore the potential involvement of rare variants in RSA risk in 325 patients with RSA. Among these rare variants, a novel missense mutation in the *ANXA4* gene (c.23G > A, p.G8D) was detected in 1 patient with RSA (RSA-219). This variant was not found in either 941 controls or public databases. Functional assays revealed that this novel mutation could influence cell migration, invasion and adhesion, indicating that the *ANXA4* mutation p.G8D contributes to the pathogenesis of RSA in women of Chinese ethnicity.

2 Materials and methods

2.1 Clinical samples

A total of 325 Han Chinese RSA patients and 941 control individuals at the Maternal and Child Health Hospital of Jiangxi Province between August 2018 and October 2019 were enrolled in this study. The RSA patients had experienced more than 2 unexplained and consecutive spontaneous abortions at less than 10 weeks of gestation. Patients who had genital abnormalities; chronic hypertension; diabetes; liver, kidney, cardiovascular, and thyroid diseases; autoimmune diseases; or infectious diseases were excluded. The control group was composed of women who underwent induced abortions of a normal pregnancy at a gestational age matched to that of the RSA group. The individuals in the control group were not treated with any drugs and had no pregnancy risk. The basic information of our study population is summarized in Table 1.

All patients signed an informed consent form before being enrolled in the study, and this study was approved by the Ethics Committee of Jiangxi Provincial Maternal and Child Health Hospital (EC-KY-202149).

2.2 Exome sequencing and data analysis

Genomic DNA was extracted from the peripheral blood of each sample with the AxyPrep Blood Genomic DNA Miniprep Kit (Axygen Scientific, Inc., 33,210 Central Avenue, Union City, CA 94587, United States) according to the instructions provided by the manufacturer. DNA quality and concentration were determined with a NanoDrop-1000 spectrophotometer (Thermo Fisher, United States) and gel electrophoresis, respectively. Exome capture was performed with a BGI Exome V4 (59 Mb) Kit (BGI, China) according to the manufacturer's protocols. DNA sequencing was performed on BGISEQ-500 sequencers (BGI, China) in high-output mode with 100 bp paired-end reads. Exome data were mapped to the human reference genome GRCh37 using the Burrows–Wheeler Aligner (BWA-MEM, version 0.6.2) (26). The credibility and quality of single-nucleotide polymorphisms (SNPs) and small insertions/deletions (InDels) were detected with the Genome Analysis Tool Kit (GATK, version 3.7) (27). ANNOVAR was used for the annotation and classification of SNPs and InDels, respectively (28). Variants with a read depth > 20 were extracted for further analysis. Finally, the variants identified through the above pipeline were filtered to eliminate benign variants with minor allele frequency (MAF) ≥1% in the dbSNP database,[1] 1,000 Genomes database,[2] Exome Aggregation Consortium (ExAC, http://exac.broadinstitute.org/) and BGI's in-house databases containing exome data from >100,000 subjects (29, 30). We ranked the genes and their potential for damage using the online prediction program SIFT to analyze whether these variants were potentially pathogenic. Variants that received lower SIFT scores (typically ≤0.05), indicating a greater likelihood of affecting protein function, were predicted to be damaging.

1 http://www.ncbi.nlm.nih.gov/projects/snp/

2 www.1000genomes.org/

TABLE 1 The basic clinical features of RSA patients and controls.

Characteristic		RSA patients	Controls
No. of samples		325	941
Age at diagnosis (years)		30.4 ± 5.27	29.4 ± 5.07
Gravidity		4.3 ± 1.38	1.9 ± 1.72
	<2	0	347
	2–4	232	556
	5–7	86	38
	>7	7	0
Parity		0.5 ± 0.64	NA
	0	189	
	1	102	
	≥2	34	
Number of spontaneous abortions		2.4 ± 0.74	NA
	2	237	
	3	57	
	4	20	
	≥5	11	

NA, not available; RSA, recurrent spontaneous abortion. All data are presented as mean ± SEM.

2.3 Validation by Sanger sequencing

Sanger sequencing was used to verify the mutation in the gene of interest in RSA patients. The polymerase chain reaction (PCR) primer was designed according to the mutation site in the *ANXA4* gene with Primer 3 (F: 5'-GGCCTCGAAGAACTTCTGCT-3′, R: 5'-TGGGCATCTTCCATGGCATT-3′). PCR amplification was performed using the rTaq enzyme (Takara Biotechnology, Dalian, China) with 50 ng of genomic DNA as a template. PCR was performed with an ABI 2720 thermocycler (Applied Biosystems, Waltham, MA), and the reaction conditions were as follows: the first step was 94°C for 5 min, followed by 30 cycles of denaturation at 94°C for 30 s, annealing at 52°C for 30 s, and extension at 72°C for 30 s and a final extension at 72°C for 10 min. The amplified PCR products were subjected to 1% agarose gel electrophoresis and then purified from the gel with a DNA purification kit (Tiangen, Beijing, China). The purified PCR products were sequenced with an ABI 3730 Automatic Capillary DNA Sequencer (Applied Biosystems, Waltham, MA). The obtained DNA sequences were analyzed using DNASTAR Lasergene version 7.2 software (DNASTAR, Inc., Madison, WI, United States) with the reference sequence from the NCBI database.

2.4 Evolutionary conservation analysis of the rare *ANXA4* variant (p.G8D)

To predict the potential pathogenicity of the mutation, evolutionary conservation analysis was performed by aligning the amino acid sequences of *ANXA4* proteins from 13 species obtained from the GenBank database,[3] including *Human* (NP_001144.1), *Chimpanzee* (XP_016804171.1), *Pongo abelii* (XP_002811982.2), *Green monkey* (XP_007968539.1), *Nomascus leucogenys* (XP_030683909.1), *Mouse* (NP_001318049.1), *Rat* (NP_077069.3), *Cattle* (NP_001001440.2), *Pig* (NP_001161111.1), *Dog* (NP_001003039.2), *Horse* (XP_023474726.1), *Rabbit* (XP_017196184.1) and *Goat* (XP_017910949.1). Evolutionary conservation analysis was performed using MEGA4 software.

2.5 Protein structural modeling

The modeling of the protein template between the reference and the modified (p.G8D) mutation of the *ANXA4* gene was conducted using the SWISS-MODEL repository database.[4] Then, we employed the Chimera 1.14rc package to concurrently compare the protein models.

2.6 Plasmid construction

Wild-type and mutant *ANXA4* (NM_001153.5) were obtained from GeneCreate Biological Engineering Co., Ltd. (Wuhan, China). Human *ANXA4* cDNA was inserted into the pcDNA3.1 vector to generate the wild-type plasmid. The mutant plasmid was generated with the wild-type plasmid as the template using a KOD-Plus-Mutagenesis Kit (Toyobo, Osaka, Japan). All the plasmids were confirmed by Sanger sequencing.

2.7 Cell culture and transfection

The THESCs cell line (human endometrial stromal cell line) was obtained from ATCC (Manassas, VA, http://www.lgcstandards-atcc.org). The cells were cultured in DMEM/F12 medium (HyClone, GE,

3 https://www.ncbi.nlm.nih.gov/gene/

4 http://www.expasy.org/

United States) supplemented with 10% fetal bovine serum (FBS) (SA211.02, Minhai Bioengineering Co., Ltd., Lanzhou, China) and 1% penicillin–streptomycin (PS) (P1400, Beijing Solarbio Science & Technology Co., Ltd., Beijing, China) in an incubator at 37°C with a humidified atmosphere of 5% CO2.

The wild-type or mutant *ANXA4* plasmids were transfected into THESCs with FuGENE HD transfection reagent (Promega, E2311) according to the manufacturer's instructions. Briefly, 1×10^5 THESCs were seeded in 6-well plates and cultured for 24 h. The medium in the 6-well plate was subsequently discarded, and 800 μL of DMEM (without FBS or antibiotics) was added to each well. A total of 100 μL of DMEM (without FBS or antibiotics), 2 μg of plasmid and 5 μL of FuGENE HD transfection reagent were mixed together, and the mixture was incubated for 10 min at room temperature. Finally, the transfection mixture was added to the cells gently and incubated in an incubator with 5% CO_2 at 37°C for 2 h. Then, 1 mL of DMEM with 10% FBS was added to each well, and the cells were cultured in an incubator with 5% CO_2 at 37°C.

2.8 Western blotting

Protein expression was detected by Western blotting analysis. Forty-eight hours after transfection, the cells were treated with cell lysis buffer (Applygen Technologies Co., Ltd., Beijing). The protein concentration was determined by using a bicinchoninic acid (BCA) protein analysis kit (Thermo Fisher Biochemical, Beijing). The protein samples (20–30 μg) were separated by 10% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS–PAGE), transferred to polyvinylidene difluoride (PVDF) membranes (Merck KGaA, United States), and blocked in 5% skim milk for 1 h at room temperature. The membranes were subsequently incubated with the following primary antibodies at 4°C overnight: anti-*ANXA4* (1:1000, TA803051; Origene, China) and anti-β-actin (1,2000, sc-47778; Santa Cruz). After being washed with 1× TBST three times, the membranes were incubated with the corresponding secondary antibodies for 2 h at room temperature. Protein bands were visualized using luminol reagent (sc-2048; Santa Cruz Biotechnology Company, United States) or enhanced chemiluminescence reagent (Santa Cruz Biotechnology Company, United States) with a ChemiDoc XRS instrument (Bio-Rad Laboratories, CA, United States). Relative protein expression levels were measured using ImageJ software (NIH).

2.9 RNA isolation and real-time quantitative PCR

RT–qPCR was used to assess the mRNA expression of *ANXA4*. The quality and concentration of total RNA were confirmed with a Nanodrop 2000 instrument (Thermo Scientific, United States), followed by extraction from treated cells using TRIzol reagent (Invitrogen, United States). Five hundred nanograms of RNA was reverse transcribed to synthesize cDNA (20 μL) using a PrimeScript 1st Strand cDNA Synthesis Kit (TaKaRa, Japan) according to the manufacturer's protocol. Then, qPCR was performed with SYBR Green PCR Master Mix (TaKaRa, Japan) on an ABI 7500 thermocycler (Applied Biosystems, Waltham, MA). *ANXA4* was used as the target gene, and *GAPDH* was used as an internal reference. The primer sequences for both genes were designed and validated for specificity. The primer sequences used for RT–qPCR were as follows: *ANXA4*-F (5'-TGCCCTGCTGAGCTGGACTT-3′) and *ANXA4*-R (5'-AAAGCT GCTCAGGACCATGT-3′); *GAPDH*-F (5'-GGAGCGAGATCCCTCC AAAAT-3′) and *GAPDH*-R (5′- GGCTGTTGTCATACTTCTCAT GG-3′). The following qPCR conditions were used: initial denaturation at 95°C for 2 min, followed by 30 cycles of denaturation at 95°C for 15 s, annealing at 60°C for 10 s, and extension at 72°C for 15 s and a final extension at 72°C for 10 min. The $2^{-\Delta\Delta Ct}$ method was used to calculate and normalize the relative expression of *ANXA4* to that of *GAPDH*.

2.10 Cell migration and invasion assays

The migration and invasion capabilities of THESCs transfected with wild-type or mutant *ANXA4* plasmids were assessed with Transwell assays (Corning, Toledo, OH, United States).

For the migration analysis, after 48 h of transfection, the cells (3×10^5/ml) were collected, resuspended in DMEM/F12 without FBS, and then inoculated into the upper Transwell chamber, while the lower chamber was filled with 500 μL of DMEM/F12 supplemented with 20% FBS. After incubation for 24 h, the upper chambers were washed, fixed with methanol for 20 min, stained with 1% crystal violet dye solution for 30 min at room temperature, and washed with phosphate-buffered saline (PBS) several times. The chamber membranes were then photographed via light microscopy (IX71, OLYMPUS, Japan), and the number of migratory cells on the lower surface was counted in five random fields to assess the migration capability.

For invasion analysis, a total of 4×10^5 cells were seeded into the upper Transwell chamber, after which the upper chamber was precoated with Matrigel matrix (3.5 mg/mL) (BD Biosciences, United States). After 48 h, the same protocol as in the migration analysis was performed.

2.11 Cell adhesion assay

Briefly, after 48 h of transfection, a total of 4×10^3 cells/well were seeded in a 96-well plate, which was pretreated with 50 μL/well of Matrigel matrix (BD Biosciences, United States) and incubated at 37°C for 30 min to allow polymerization. Then, the 96-well plates were incubated in an incubator with 5% CO_2 at 37°C for 20 min, 40 min or 60 min. The plates were washed with PBS three times to remove nonadherent cells, and the adherent cells were fixed with methanol for 15 min and stained with 1% crystal violet dye solution for 30 min at room temperature. The numbers of adherent cells were counted using light microscopy (IX71, OLYMPUS, Japan).

2.12 Statistical analyses

All the statistical analyses were performed using the statistical analysis software SPSS 21.0. Data are expressed as the mean ± standard

error of the mean (S.E.M.). An unpaired Student's *t* test was used to evaluate the significant differences among the groups. Statistical significance was established at a *p* value of less than 0.05.

3 Results

3.1 Identification of one novel rare variant in *ANXA4* in one patient with RSA using WES

In total, 325 patients with RSA and 941 local women without RSA as controls underwent WES. WES performed on BGISEQ-500 sequencers achieved, on average, 21,266.42 Mb of raw bases. After removing low-quality reads, we obtained an average of 212,622,806 clean reads (21,255.71 Mb). The clean reads of each sample had high Q20 and Q30 values, indicating high sequencing quality. The average GC content was 52.14%. The chromosomal positions of the SNPs were based on the UCSC GRCh37/hg19 genome.

We obtained a total of 79,534 variants, including synonymous, missense, stop-loss, stop-gain, start-loss and splicing variants. The SnpEff tool[5] was used to perform variant annotation and prediction. After excluding variants with MAFs greater than or equal to 1% in publicly available databases, e.g., dbSNP, 1,000 Genomes Project, ExAC and BGI's in-house databases, 262 variants were retained. Using the *in silico* prediction algorithm SIFT, we selected candidate variants with potentially damaging impacts. The results showed that the *ANXA4* gene was prominent on the basis of its functional annotation related to cell migration and invasion and the pathogenicity predictions of its mutations.

A novel mutation (c.23G > A, p.G8D) in the *ANXA4* gene (NM_001153.5) was detected in 1 of the 325 patients with RSA (RSA-219) but was not detected in either the 941 controls or public databases (dbSNP, 1,000 Genomes Project, ExAC and BGI in-house databases). Primers were designed for PCR to amplify the *ANXA4* fragment containing the rare variant, and Sanger sequencing was performed to validate this mutation (Figure 1A). The patient (RSA-219) was confirmed to harbor this mutation by Sanger sequencing. The *ANXA4* gene is located on chromosome 2p13.3 and consists of 22 exons. The mutation is positioned in the first exon.

3.2 Evolutionary conservation analysis and protein structural modeling

Evolutionary conservation analysis revealed that the amino acid residue affected by the p.G8D mutation was highly conserved among 13 vertebrate species (Figure 1B). Furthermore, we analyzed the changes in the 3D structure of *ANXA4* between the wild-type and mutant proteins using SWISS-MODEL. Protein structural modeling revealed that the p.G8D *ANXA4* mutation caused structural changes in the *ANXA4* protein (Figure 2).

5 http://snpeff.sourceforge.net/SnpEff_manual.html

3.3 A novel mutation in *ANXA4* (p.G8D) inhibits the migration, invasion and adhesion of THESCs

Annexins have been suggested to promote cellular migration, invasion and adhesion, and adhesion is an essential step required for blastocysts to adhere to the endometrium (31). Next, we examined whether the p.G8D mutation in *ANXA4* affects protein function.

THESCs were cultured and transfected with *ANXA4* wild-type (WT) or mutant-type (MT) plasmids, and pcDNA3.1 was used as a control plasmid (Con). Twenty-four hours after transfection, G418 was used to screen stably transfected cells, and one individual clone was separated by dilution. The expression of *ANXA4* in stably transfected cells was confirmed by RT–qPCR and WB (Figures 3A–C).

Transwell assays were used to detect the effects of *ANXA4* WT/MT on migration and invasion. The results showed that MT *ANXA4* could inhibit both cell migration ($p = 0.0018$) (Figures 3D,F) and invasion ($p = 0.0148$) (Figures 3E,G) of stable transfected cells, compared to WT *ANXA4*.

We examined the effect of the *ANXA4* mutation on cell adhesion with a Matrigel adhesion assay. Compared with WT *ANXA4*, MT *ANXA4* inhibited cell adhesion, especially at 40 min ($p = 0.0147$) and 60 min ($p = 0.0137$) (Figures 3H,I).

These results indicated that the novel *ANXA4* mutation inhibited cell migration, invasion and adhesion.

4 Discussion

The main cause of abortion is implantation failure. Embryo implantation is a dynamic development process that requires a series of interactions between the blastocyst and the endometrium. The three prerequisites for successful implantation are an embryo with implantation ability, a receptive endometrium and simultaneous development of the embryo-endometrium. Many existing animal studies have confirmed that the uterus can keep the embryo in a viable but dormant state until the uterus is in an acceptable state for implantation, indicating that the proper regulation of the uterus by ovarian hormones mainly determines the success of implantation (32). Decidualization of the endometrium leads to corresponding vascular changes preparing for the invasion of human endometrial trophoblasts and the formation of a functional blood placenta; in other words, decidualization contributes to the functional transition of the endometrium from a nonreceptive state to a receptive state (33). An increasing amount of experimental and clinical data indicates that impaired or disrupted decidualization plays a role in the development of an unsuitable maternal–fetal interface. This situation has important clinical implications, including repeated implantation failures and recurrent pregnancy losses in the early stages of pregnancy, as well as several serious complications during late gestation (34). Studies have shown that in the presence of trophoblasts, the motility and invasive ability of decidualized endometrial stromal cells are enhanced (35), and trophoblast cell-derived CXCL12 has been shown to promote the expression and invasiveness of CXCR4 (the receptor of CXCL12) in decidual stromal cells isolated from early pregnancy (36). Many studies have shown that the coordinated migration and invasion of decidualized endometrial stromal cells in response to signals from embryo and trophoblast cells is the key to successful implantation. The

10.3389/fmed.2024.1462649

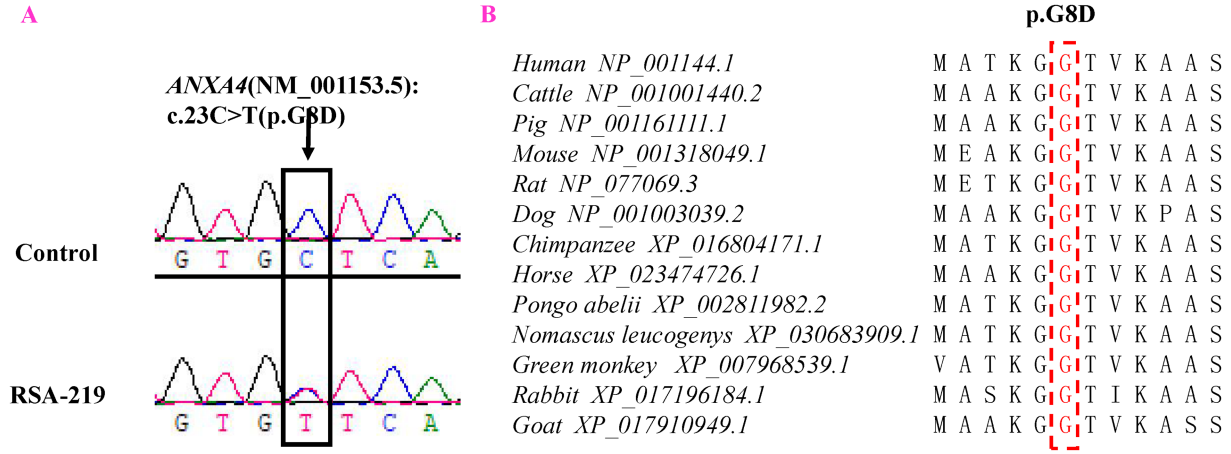

FIGURE 1
Representative sequencing electropherogram of the *ANXA4* variant (p.G8D). The arrow indicates the location of the mutation, and patient "RSA-219" harbored the *ANXA4* p.G8D variant **(A)**. Evolutionary conservation analysis of *ANXA4* p.G8D in vertebrate species **(B)**.

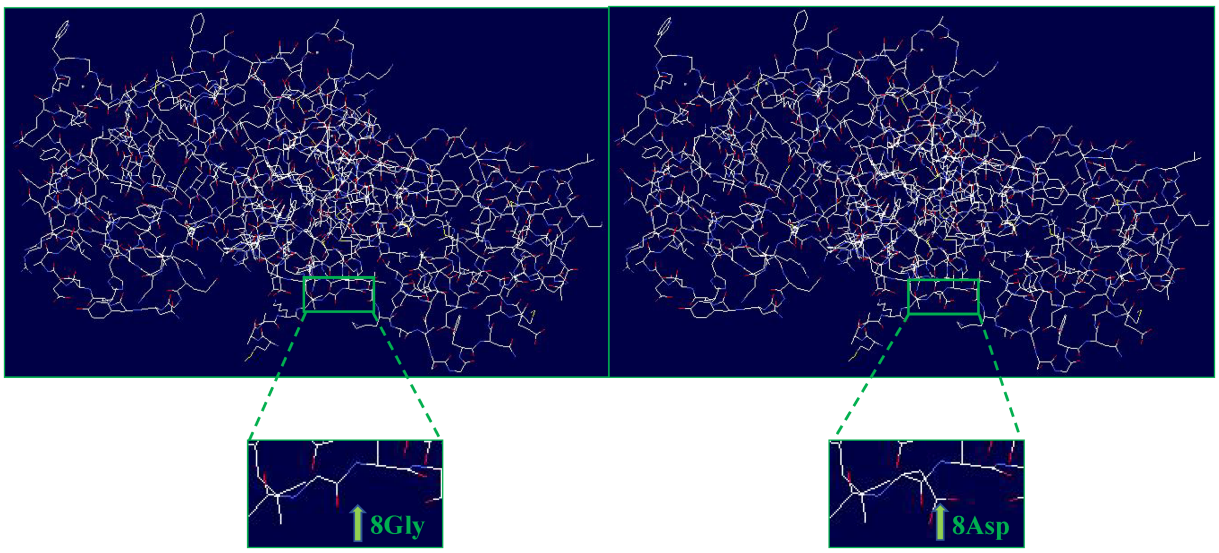

FIGURE 2
Structural differences between wild-type *ANXA4* (p.G8) and mutated *ANXA4* (p.D8) proteins. The protein structures of the wild-type and mutated *ANXA4* proteins were modeled based on the crystal model of the human *ANXA4* protein.

ability of mature endometrial stromal cells to migrate and invade is increasingly recognized as the basis of intense tissue remodeling associated with endometrial regeneration, decidualization, embryo implantation, and trophoblast invasion (37, 38).

The annexin family is a class of proteins that are widely present in cell membranes and intracellular organs and are known mainly for their ability to bind phospholipids and their dependence on calcium ions. Given that the annexin family is highly expressed at the maternal–fetal interface and that both the endometrium and blastocyst undergo swift growth and differentiation throughout pregnancy, annexin is believed to perform various functions in both the placental tissues originating from the fetus and the decidual cells derived from the mother (39). Defective expression of endometrial ANXA2 may impair decidualization of endometrial stromal cells *in vitro* and *in vivo*, leading to the development of preeclampsia (40). Previous studies have demonstrated that the binding of antibodies and the apoptosis of syncytiotrophoblasts, can hinder the secretion of trophoblast gonadotropin. These factors may be crucial mechanisms through which anti-annexin V antibodies influence embryo implantation upon binding and the results of pregnancy (41). Annexin A4 (*ANXA4*) is a member of the annexin family that belongs to the multigene family of calcium ion (Ca^{2+}) and phospholipid-binding proteins. The gene encoding human *ANXA4* is located on chromosome 2q13.3. The C-terminal conserved domain contains four annexin repeats, each of which has a Ca^{2+}-binding site and five α-helical structures, which can bind phospholipids in a calcium-dependent manner. The N-terminal domain is 12 amino acids long and determines the function of *ANXA4 in vivo*. *ANXA4* can regulate

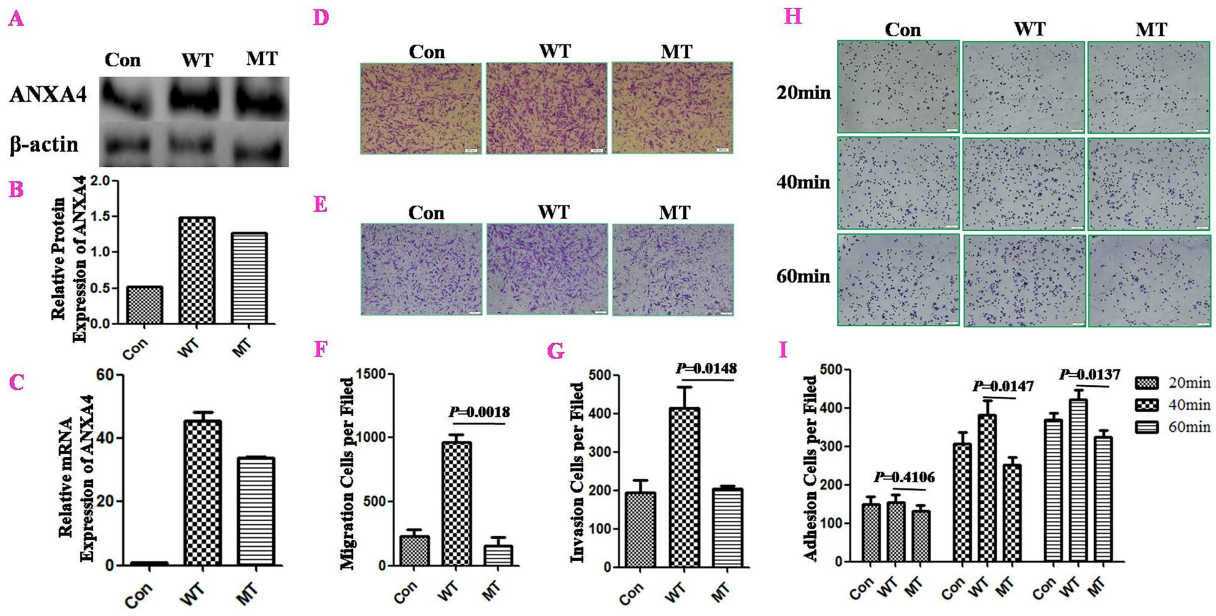

FIGURE 3
The *ANXA4* p.G8D mutation inhibited cell migration, invasion and adhesion in stably transfected THESCs. THESCs were cultured and transfected with *ANXA4* wild-type (WT) or mutant-type (MT) plasmids, while pcDNA3.1 was used as a control plasmid (Con). Twenty-four hours after transfection, G418 was used to screen stably transfected cells, and one individual clone was separated by dilution. *ANXA4* expression in stably transfected cells was confirmed by WB **(A,B)** and RT–qPCR **(C)**. The effects of *ANXA4* mutation on cell migration **(D,F)** and invasion **(E,G)** were determined by Transwell assays, and cell adhesion **(H,I)** was determined by Matrigel adhesion assays. These assays revealed that *ANXA4* p.G8D mutation inhibited cell migration, invasion and adhesion in stably transfected cells compared with *ANXA4* WT. Scale bars, 100 μm. The data are shown as the means ± SDs. $p < 0.05$ was considered to indicate a significant difference.

membrane permeability and membrane transport, participate in cell growth and apoptosis, increase tumor invasion and promote antitumor drug resistance (24). An increasing number of studies have shown that *ANXA4* is highly expressed in various clinical tumors and is an indicator of tumor development, invasion, chemoresistance, and poor outcomes in cancer patients (42–45). For example, *ANXA4* is upregulated and translocated to the nucleus in ovarian clear cell carcinoma and colorectal cancer (44, 45), and the knockdown of *ANXA4* weakens the migration and invasion of ovarian cancer and breast cancer cells (43, 46). In addition, several studies have shown that *ANXA4* expression is significantly increased in the oviductal fluid of pregnant mares and bovines, and in the uterine endometrium during early pregnancy in pigs, *ANXA4* has been identified as an embryo-interacting protein originating from oviductal fluid (47–49). Additionally, the *ANXA4* transcript in the human endometrium is significantly upregulated during the menstrual secretory phase compared with the proliferative phase (50). Previous studies reported that elevated *ANXA4* mRNA levels are correlated with an increasing level of progesterone (51), and might play a crucial role in the receptive process (52). Proportionately, proteomic *PLCD4* (p.L696P) analyses of human placental tissues have shown that *ANXA4* expression is downregulated in preeclampsia (PE) placentas and PE placenta-derived extravillous cytotrophoblasts compared with the expression in normal placentas (53). Additionally, *ANXA4* overexpression alleviated rat PE progression, accompanied by increases in the expression of PI3K, p-Akt, and p-eNOS in rat placentas, which indicated that *ANXA4* may promote trophoblast invasion through the PI3K/Akt/eNOS pathway (53). Furthermore, the *ANXA4* gene may play an important role in pregnancy, but there are currently no reports on whether it is involved in the occurrence of miscarriage.

RSA is a multifactorial disease, the exact causes of which are still unknown. Environmental, maternal, and genetic factors have been shown to contribute to this condition (54). WES is widely used to detect genetic variations associated with human diseases, including RSA. In recent years, using WES, studies in families with inherited recurrent pregnancy loss (RPL) confirmed the associations of a number of gene variations with RPL, including *IFT122* (p.V553G, p.S373F, p.W7C, p.G546R), *ASIC5* (p.R227I), *DYNC2H1* (p.Y2016C, p.D2184V), *ALOX15* (p.Y139C, p.T560M), *FKBP4* (p.A16E, p.N125S, p.Q381L, p.R399Q), *PLCD4* (p.L696P) and *OSBPL5* (p.G385R) (55–59). The compromised expression of FKBP4, resulting from the increased expression of miR-29c, leads to impaired progesterone signaling and defects in decidualization (60). This may contribute to the development of endometriosis and infertility. In addition, studies in unrelated RSA patients and controls revealed potentially harmful gene variants, including *APP* (p.K510N), *FN1* (p.M1874T), *KDR* (p.D814N), *POLR2B* (p.G136C), *ITGB1* (p.Y219H), *PLK1* (p.A404S), *COL4A2* (p.H1603P), *LAMA4* (p.D1053G), *FOXA2* (p.Y420X), *FGA* (p.A762V), *F13A1* (p.Q401X, p.R612C), *KHDC3L* (p.146_156del), *ANXA5* (p.G317R), *DNMT1* (p.G876R), *THBS1* (p.N700S), and *MSH2* (p.L390F) (11, 61). Earlier research has offered significant understanding regarding the function of human KHDC3L in the modulation of homologous recombination (HR) repair, the activation of PARP1, and the maintenance of genome stability, while also recognizing human KHDC3L as an emerging risk gene associated with RPL (62). Previous studies have demonstrated that reduced expression of ANXA5, associated with the M2 haplotype, increases the risk of vascular

thrombosis during pregnancy, thereby increasing the occurrence of RPL, primary maternal peripheral complications (PMPC), and recurrent implantation failure (RIF) (63). In the present study, a total of 325 patients with RSA and 941 local women without RSA as controls were included, and we identified specific gene variants that are pathogenic or possibly pathogenic through WES analysis. Among those variants, a novel *ANXA4* mutation (p.G8D) in 1 of the 325 patients with RSA (RSA-219) was detected. This mutation was not detected in either the 941 controls or public databases (dbSNP, 1,000 Genomes Project, ExAC and BGI in-house databases). The SIFT program prediction and structural modeling analysis show that this mutation is harmful. Furthermore, functional analyses revealed that this mutation could inhibit cell migration, invasion and adhesion. Moreover, despite appropriate fertility preservation treatment, the patient experienced another recurrent miscarriage in 2020 during the follow-up period. These results indicate that the *ANXA4* mutation p.G8D plays an important role in the pathogenesis of RSA. To the best of our knowledge, this is the first report revealing a novel *ANXA4* mutation that may be associated with RSA. Insights from these studies can help us better identify recurrent miscarriages. By discovering new mutations in the *ANXA4* gene that are associated with recurrent miscarriage, geneticists can work with clinicians to identify patients with deleterious variants in orthologous genes, thereby playing an active role in personalized medicine. However, the underlying mechanisms by which this mutation affects RSA need further investigation.

5 Conclusion

In conclusion, for the first time, we detected a novel *ANXA4* mutation (p.G8D) in 325 samples from women with RSA using WES analysis, this mutation was not detected in either 941 local control women or public databases. Furthermore, functional analyses revealed that this mutation could inhibit cell migration, invasion and adhesion. Taken together, our findings suggest that the *ANXA4* mutation p.G8D plays an important role in the pathogenesis of RSA and may contribute to its genetic diagnosis.

Data availability statement

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

Ethics statement

The studies involving humans were approved by the Ethics Committee of Jiangxi Maternal and Child Health 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. Ethical approval was not required for the studies on animals in accordance with the local legislation and institutional requirements because only commercially available established cell lines were used. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

QY: Data curation, Writing – original draft, Validation. F-YL: Data curation, Validation, Writing – review & editing. X-JX: Data curation, Investigation, Resources, Writing – original draft. X-YC: Data curation, Investigation, Writing – original draft. LZ: Investigation, Writing – original draft. H-MW: Validation, Writing – original draft. D-DL: Investigation, Writing – original draft. C-NX: Validation, Writing – original draft. TH: Investigation, Writing – original draft. YC: Data curation, Methodology, Writing – review & editing, Writing – original draft. YZ: Data curation, Methodology, Resources, Writing – review & editing, Writing – original draft.

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 Key Research and Development Program of Jiangxi Province (20203BBGL73135) and the Natural Science Foundation of Jiangxi Province (20181BBG70018).

Acknowledgments

We thank all participants involved in this study.

Conflict of interest

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

Publisher's note

References

1. Practice Committee of the American Society for Reproductive Medicine. Electronic address aao. Definitions of infertility and recurrent pregnancy loss: a committee opinion. *Fertil Steril*. (2020) 113:533–5. doi: 10.1016/j.fertnstert.2019.11.025

2. Zhang LM, Yang YN, Zhang RX, Luo L, Tan JF, Zhou L, et al. Comparison of the etiological constitution of two and three or more recurrent miscarriage. *Zhonghua Fu Chan Ke Za Zhi*. (2018) 53:855–9. doi: 10.3760/cma.j.issn.0529-567x.2018.12.010

3. Youssef A, Vermeulen N, Lashley E, Goddijn M, van der Hoorn MLP. Comparison and appraisal of (inter) national recurrent pregnancy loss guidelines. *Reprod Biomed Online*. (2019) 39:497–503. doi: 10.1016/j.rbmo.2019.04.008

4. Practice Committee of the American Society for Reproductive Medicine. Evaluation and treatment of recurrent pregnancy loss: a committee opinion. *Fertil Steril*. (2012) 98:1103–11. doi: 10.1016/j.fertnstert.2012.06.048

5. Wu M, Zhu Y, Zhao J, Ai H, Gong Q, Zhang J, et al. Soluble costimulatory molecule sTim3 regulates the differentiation of Th1 and Th2 in patients with unexplained recurrent spontaneous abortion. *Int J Clin Exp Med*. (2015) 8:8812–9.

6. Popescu F, Jaslow CR, Kutteh WH. Recurrent pregnancy loss evaluation combined with 24-chromosome microarray of miscarriage tissue provides a probable or definite cause of pregnancy loss in over 90% of patients. *Hum Reprod*. (2018) 33:579–87. doi: 10.1093/humrep/dey021

7. Rodrigues VO, Soligo A, Pannain GD. Antiphospholipid antibody syndrome and infertility. *Rev Bras Ginecol Obstet*. (2019) 41:621–7. doi: 10.1055/s-0039-1697982

8. Ticconi C, Pietropolli A, Di Simone N, Piccione E, Fazleabas A. Endometrial immune dysfunction in recurrent pregnancy loss. *Int J Mol Sci*. (2019) 20:5332. doi: 10.3390/ijms20215332

9. di Simone N, Castellani R, Raschi E, Borghi MO, Meroni PL, Caruso A. Anti-beta-2 glycoprotein I antibodies affect Bcl-2 and Bax trophoblast expression without evidence of apoptosis. *Ann N Y Acad Sci*. (2006) 1069:364–76. doi: 10.1196/annals.1351.034

10. D'Ippolito S, Ticconi C, Tersigni C, Garofalo S, Martino C, Lanzone A, et al. The pathogenic role of autoantibodies in recurrent pregnancy loss. *Am J Reprod Immunol*. (2020) 83:e13200. doi: 10.1111/aji.13200

11. Mou JT, Huang SX, Yu LL, Xu J, Deng QL, Xie YS, et al. Identification of genetic polymorphisms in unexplained recurrent spontaneous abortion based on whole exome sequencing. *Ann Transl Med*. (2022) 10:603. doi: 10.21037/atm-22-2179

12. Claussnitzer M, Cho JH, Collins R, Cox NJ, Dermitzakis ET, Hurles ME, et al. A brief history of human disease genetics. *Nature*. (2020) 577:179–89. doi: 10.1038/s41586-019-1879-7

13. Erdenee S, Akhatayeva Z, Pan C, Cai Y, Xu H, Chen H, et al. An insertion/deletion within the CREB1 gene identified using the RNA-sequencing is associated with sheep body morphometric traits. *Gene*. (2021) 775:145444. doi: 10.1016/j.gene.2021.145444

14. Savoldi IR, Ibelli AMG, Cantao ME, Peixoto JO, Pires MP, Mores MAZ, et al. A joint analysis using exome and transcriptome data identifiescandidate polymorphisms and genes involved with umbilical hernia in pigs. *BMC Genomics*. (2021) 22:818. doi: 10.1186/s12864-021-08138-4

15. Li C, Ou R, Hou Y, Chen Y, Wei Q, Zhang L, et al. Rare variant analysis of PTRHD1 in Parkinson's disease in the Chinese population. *J Parkinsons Dis*. (2022) 12:1917–20. doi: 10.3233/JPD-223337

16. Kolte AM, Nielsen HS, Moltke I, Degn B, Pedersen B, Sunde L, et al. A genome-wide scan in affected sibling pairs with idiopathic recurrent miscarriage suggests genetic linkage. *Mol Hum Reprod*. (2011) 17:379–85. doi: 10.1093/molehr/gar003

17. Karim S, Jamal HS, Rouzi A, Ardawi MSM, Schulten HJ, Mirza Z, et al. Genomic answers for recurrent spontaneous abortion in Saudi Arabia: an array comparative genomic hybridization approach. *Reprod Biol*. (2017) 17:133–43. doi: 10.1016/j.repbio.2017.03.003

18. Pereza N, Ostojic S, Kapovic M, Peterlin B. Systematic review and meta-analysis of genetic association studies in idiopathic recurrent spontaneous abortion. *Fertil Steril*. (2017) 107:150–159.e2. doi: 10.1016/j.fertnstert.2016.10.007

19. Loizidou EM, Kucherenko A, Tatarskyy P, Chernushyn S, Livshyts G, Gulkovskyi R, et al. Risk of recurrent pregnancy loss in the Ukrainian population using a combined effect of genetic variants: a case-control study. *Genes (Basel)*. (2021) 12:64. doi: 10.3390/genes12010064

20. Povysil G, Petrovski S, Hostyk J, Aggarwal V, Allen AS, Goldstein DB. Rare-variant collapsing analyses for complex traits: guidelines and applications. *Nat Rev Genet*. (2019) 20:747–59. doi: 10.1038/s41576-019-0177-4

21. Gourhant L, Bocher O, De Saint ML, Ludwig TE, Boland A, Deleuze JF, et al. Whole exome sequencing, a hypothesis-free approach to investigate recurrent early miscarriage. *Reprod Biomed Online*. (2021) 42:789–98. doi: 10.1016/j.rbmo.2021.01.008

22. Smail C, Ferraro NM, Hui Q, Durrant MG, Aguirre M, Tanigawa Y, et al. Integration of rare expression outlier-associated variants improves polygenic risk prediction. *Am J Hum Genet*. (2022) 109:1055–64. doi: 10.1016/j.ajhg.2022.04.015

23. Arii Y, Butsusihta K, Fukuoka S. Role of calcium-binding sites in calcium-dependent membrane association of annexin A4. *Biosci Biotechnol Biochem*. (2015) 79:978–85. doi: 10.1080/09168451.2014.1003131

24. Yao H, Sun C, Hu Z, Wang W. The role of annexin A4 in cancer. *Front Biosci (Landmark Ed)*. (2016) 21:949–57. doi: 10.2741/4432

25. Ponnampalam AP, Rogers PA. Cyclic changes and hormonal regulation of annexin IV mRNA and protein in human endometrium. *Mol Hum Reprod*. (2006) 12:661–9. doi: 10.1093/molehr/gal075

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. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From fast Q data to high confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr Protoc Bioinformatics*. (2013) 43:11.10.1–11.10.33. doi: 10.1002/0471250953.bi1110s43

28. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat Protoc*. (2015) 10:1556–66. doi: 10.1038/nprot.2015.105

29. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. Db SNP: the NCBI database of genetic variation. *Nucleic Acids Res*. (2001) 29:308–11. doi: 10.1093/nar/29.1.308

30. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60, 706 humans. *Nature*. (2016) 536:285–91. doi: 10.1038/nature19057

31. Greening DW, Nguyen HP, Evans J, Simpson RJ, Salamonsen LA. Modulating the endometrial epithelial proteome and secretome in preparation for pregnancy: the role of ovarian steroid and pregnancy hormones. *J Proteome*. (2016) 144:99–112. doi: 10.1016/j.jprot.2016.05.026

32. Zhang S, Lin H, Kong S, Wang S, Wang H, Wang H, et al. Physiological and molecular determinants of embryo implantation. *Mol Asp Med*. (2013) 34:939–80. doi: 10.1016/j.mam.2012.12.011

33. Brosens JJ, Pijnenborg R, Brosens IA. The myometrial junctional zone spiral arteries in normal and abnormal pregnancies: a review of the literature. *Am J Obstet Gynecol*. (2002) 187:1416–23. doi: 10.1067/mob.2002.127305

34. Ticconi C, Di Simone N, Campagnolo L, Fazleabas A. Clinical consequences of defective decidualization. *Tissue Cell*. (2021) 72:101586. doi: 10.1016/j.tice.2021.101586

35. Gellersen B, Reimann K, Samalecos A, Aupers S, Bamberger AM. Invasiveness of human endometrial stromal cells is promoted by decidualization and by trophoblast-derived signals. *Hum Reprod*. (2010) 25:862–73. doi: 10.1093/humrep/dep468

36. Ren L, Liu YQ, Zhou WH, Zhang YZ. Trophoblast-derived chemokine CXCL12 promotes CXCR4 expression and invasion of human first-trimester decidual stromal cells. *Hum Reprod*. (2012) 27:366–74. doi: 10.1093/humrep/der395

37. Weimar CH, Macklon NS, Post Uiterweer ED, Brosens JJ, Gellersen B. The motile and invasive capacity of human endometrial stromal cells: implications for normal and impaired reproductive function. *Hum Reprod Update*. (2013) 19:542–57. doi: 10.1093/humupd/dmt025

38. Murata H, Tanaka S, Okada H. The regulators of human endometrial stromal cell Decidualization. *Biomol Ther*. (2022) 12:1275. doi: 10.3390/biom12091275

39. Hu J, Chen L, Ruan J, Chen X. The role of the annexin a protein family at the maternal-fetal interface. *Front Endocrinol (Lausanne)*. (2024) 15:1314214. doi: 10.3389/fendo.2024.1314214

40. Garrido-Gomez T, Quiñonero A, Dominguez F, Rubert L, Perales A, Hajjar KA, et al. Preeclampsia: a defect in decidualization is associated with deficiency of Annexin A2. *Am J Obstet Gynecol*. (2020) 222:376.e1–e17. doi: 10.1016/j.ajog.2019.11.1250

41. Di Simone N, Castellani R, Caliandro D, Caruso A. Monoclonal anti-annexin V antibody inhibits trophoblast gonadotropin secretion and induces syncytiotrophoblast apoptosis. *Biol Reprod*. (2001) 65:1766–70. doi: 10.1095/biolreprod65.6.1766

42. Konopka-Postupolska D, Clark G, Hofmann A. Structure, function and membrane interactions of plant annexins: an update. *Plant Sci*. (2011) 181:230–41. doi: 10.1016/j.plantsci.2011.05.013

43. Mogami T, Yokota N, Asai-Sato M, Yamada R, Koizume S, Sakuma Y, et al. Annexin A4 is involved in proliferation, chemo-resistance and migration and invasion in ovarian clear cell adenocarcinoma cells. *PLoS One*. (2013) 8:e80359. doi: 10.1371/journal.pone.0080359

44. Liu J, Wang H, Zheng M, Deng L, Zhang X, Lin B. P 53 and ANXA4/NF-kappaB p50 complexes regulate cell proliferation, apoptosis and tumor progression in ovarian clear cell carcinoma. *Int J Mol Med*. (2020) 46:2102–14. doi: 10.3892/ijmm.2020.4757

45. Peng Y, Zhang Z, Zhang A, Liu C, Sun Y, Peng Z, et al. Membrane-cytoplasm translocation of annexin A4 is involved in the metastasis of colorectal carcinoma. *Aging (Albany NY)*. (2021) 13:10312–25. doi: 10.18632/aging.202793

46. Li L, Zhang R, Liu Y, Zhang G. ANXA4 activates JAK-STAT3 signaling by interacting with ANXA1 in basal-like breast Cancer. *DNA Cell Biol*. (2020) 39:1649–56. doi: 10.1089/dna.2020.5570

47. Smits K, Willems S, Van Steendam K, Van De Velde M, De Lange V, Ververs C, et al. Proteins involved in embryo-maternal interaction around the signalling of maternal recognition of pregnancy in the horse. *Sci Rep*. (2018) 8:5249. doi: 10.1038/s41598-018-23537-6

48. Banliat C, Tsikis G, Labas V, Teixeira-Gomes AP, Com E, Lavigne R, et al. Identification of 56 proteins involved in embryo-maternal interactions in the bovine oviduct. *Int J Mol Sci*. (2020) 21:466. doi: 10.3390/ijms21020466

49. Pierzchala D, Liput K, Korwin-Kossakowska A, Ogluszka M, Polawska E, Nawrocka A, et al. Molecular characterisation of uterine endometrial proteins during early stages of pregnancy in pigs by MALDI TOF/TOF. *Int J Mol Sci*. (2021) 22:6720. doi: 10.3390/ijms22136720

50. Ponnampalam AP, Weston GC, Trajstman AC, Susil B, Rogers PA. Molecular classification of human endometrial cycle stages by transcriptional profiling. *Mol Hum Reprod*. (2004) 10:879–93. doi: 10.1093/molehr/gah121

51. Talbi S, Hamilton AE, Vo KC, Tulac S, Overgaard MT, Dosiou C, et al. Molecular phenotyping of human endometrium distinguishes menstrual cycle phases and underlying biological processes in normo-ovulatory women. *Endocrinology*. (2006) 147:1097–121. doi: 10.1210/en.2005-1076

52. Mirkin S, Arslan M, Churikov D, Corica A, Diaz JI, Williams S, et al. In search of candidate genes critically expressed in the human endometrium during the window of implantation. *Hum Reprod*. (2005) 20:2104–17. doi: 10.1093/humrep/dei051

53. Xu Y, Sui L, Qiu B, Yin X, Liu J, Zhang X. ANXA4 promotes trophoblast invasion via the PI3K/Akt/eNOS pathway in preeclampsia. *Am J Physiol Cell Physiol*. (2019) 316:C481–91. doi: 10.1152/ajpcell.00404.2018

54. Guo W, Zhu X, Yan L, Qiao J. The present and future of whole-exome sequencing in studying and treating human reproductive disorders. *J Genet Genomics*. (2018) 45:517–25. doi: 10.1016/j.jgg.2018.08.004

55. Tsurusaki Y, Yonezawa R, Furuya M, Nishimura G, Pooh RK, Nakashima M, et al. Whole exome sequencing revealed biallelic IFT122 mutations in a family with CED1 and recurrent pregnancy loss. *Clin Genet*. (2014) 85:592–4. doi: 10.1111/cge.12215

56. Filges I, Manokhina I, Penaherrera MS, McFadden DE, Louie K, Nosova E, et al. Recurrent triploidy due to a failure to complete maternal meiosis II: whole-exome sequencing reveals candidate variants. *Mol Hum Reprod*. (2015) 21:339–46. doi: 10.1093/molehr/gau112

57. Qiao Y, Wen J, Tang F, Martell S, Shomer N, Leung PC, et al. Whole exome sequencing in recurrent early pregnancy loss. *Mol Hum Reprod*. (2016) 22:364–72. doi: 10.1093/molehr/gaw008

58. Demetriou C, Chanudet E, GOSgeneJoseph A, Topf M, Thomas AC, et al. Exome sequencing identifies variants in FKBP4 that are associated with recurrent fetal loss in humans. *Hum Mol Genet*. (2019) 28:3466–74. doi: 10.1093/hmg/ddz203

59. Al Qahtani NH, AbdulAzeez S, Almandil NB, Fahad Alhur N, Alsuwat HS, Al Taifi HA, et al. Whole-genome sequencing reveals Exonic variation of ASIC5 gene results in recurrent pregnancy loss. *Front Med (Lausanne)*. (2021) 8:699672. doi: 10.3389/fmed.2021.699672

60. Joshi NR, Miyadahira EH, Afshar Y, Jeong JW, Young SL, Lessey BA, et al. Progesterone resistance in endometriosis is modulated by the altered expression of Micro RNA-29c and FKBP4. *J Clin Endocrinol Metab*. (2017) 102:141–9. doi: 10.1210/jc.2016-2076

61. Xiang H, Wang C, Pan H, Hu Q, Wang R, Xu Z, et al. Exome-sequencing identifies novel genes associated with recurrent pregnancy loss in a Chinese cohort. *Front Genet*. (2021) 12:746082. doi: 10.3389/fgene.2021.746082

62. Zhang W, Chen Z, Zhang D, Zhao B, Liu L, Xie Z, et al. KHDC3L mutation causes recurrent pregnancy loss by inducing genomic instability of human early embryonic cells. *PLoS Biol*. (2019) 17:e3000468. doi: 10.1371/journal.pbio.3000468

63. Peng L, Yang W, Deng X, Bao S. Research progress on ANXA5 in recurrent pregnancy loss. *J Reprod Immunol*. (2022) 153:103679. doi: 10.1016/j.jri.2022.103679

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