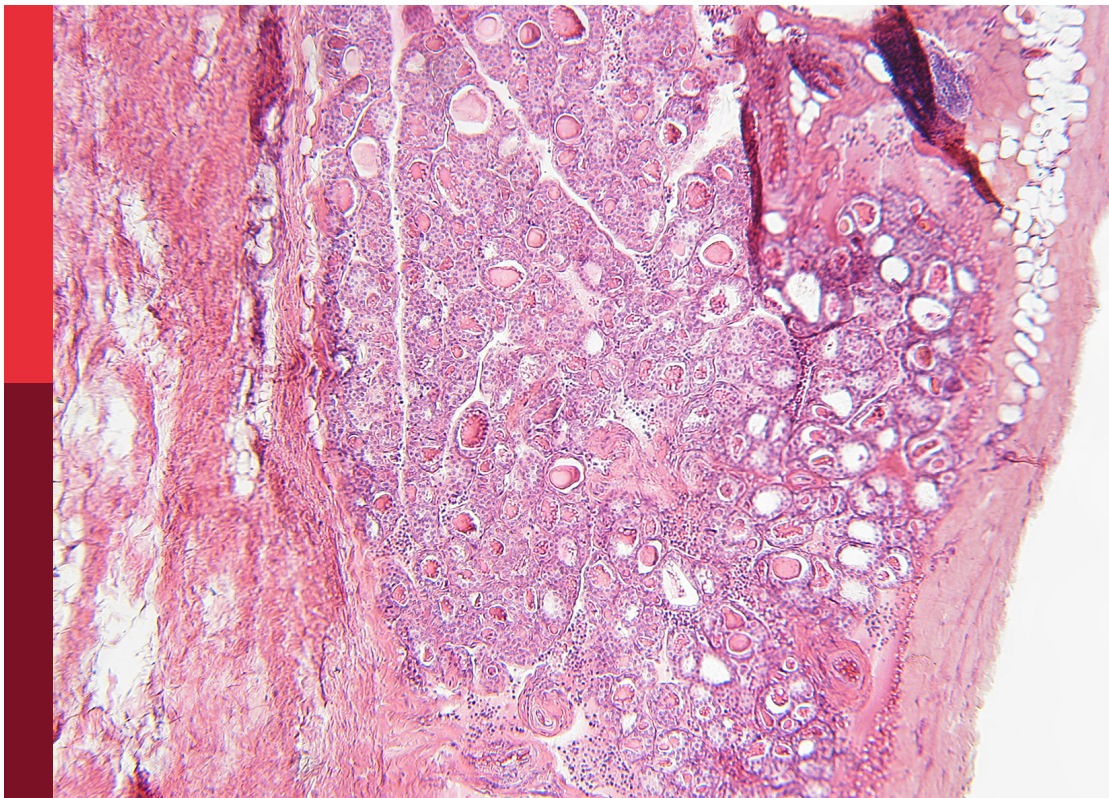
frontiers | Research Topics

Integrative multi-modal, multi-omics analytics for the better understanding of metabolic diseases

Edited by
Animesh Acharjee, Prasoon Agarwal
and Georgios Gkoutos

Published in
Frontiers in Endocrinology



ISSN 1664-8714
ISBN 978-2-8325-3550-9
DOI 10.3389/978-2-8325-3550-9

About Frontiers

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

Frontiers journal series

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

Dedication to quality

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

What are Frontiers Research Topics?

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

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

Integrative multi-modal, multi-omics analytics for the better understanding of metabolic diseases

Topic editors

Animesh Acharjee — University of Birmingham, United Kingdom
Prasoon Agarwal — National Bioinformatics Infrastructure Sweden, Sweden
Georgios Gkoutos — University of Birmingham, United Kingdom

Citation

Acharjee, A., Agarwal, P., Gkoutos, G., eds. (2023). *Integrative multi-modal, multi-omics analytics for the better understanding of metabolic diseases*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3550-9

Table of contents

frontiers | Frontiers in Endocrinology

TYPE Editorial
PUBLISHED 08 September 2023
DOI 10.3389/fendo.2023.1266557

OPEN ACCESS

EDITED AND REVIEWED BY
Tom Michoel,
University of Bergen, Norway

*CORRESPONDENCE
Animesh Acharjee
a.acharjee@bham.ac.uk

RECEIVED 25 July 2023
ACCEPTED 21 August 2023
PUBLISHED 08 September 2023

CITATION
Acharjee A, Agarwal P and Gkoutos GV (2023) Editorial: Integrative multi-modal, multi-omics analytics for the better understanding of metabolic diseases. *Front. Endocrinol.* 14:1266557. doi: 10.3389/fendo.2023.1266557

Editorial: Integrative multi-modal, multi-omics analytics for the better understanding of metabolic diseases

Animesh Acharjee[1,2,3,4]*, Prasoon Agarwal[5] and Georgios V. Gkoutos[1,2,3,4,6]

[1]College of Medical and Dental Sciences, Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, United Kingdom, [2]Institute of Translational Medicine, University Hospitals Birmingham National Health Service (NHS) Foundation Trust, Birmingham, United Kingdom, [3]Medical Research Council (MRC) Health Data Research United Kingdom (UK) (HDR), Midlands Site, Birmingham, United Kingdom, [4]Centre for Health Data Research , University of Birmingham, Birmingham, United Kingdom, [5]National Bioinformatics Infrastructure Sweden (NBIS), Science for Life Laboratory, Division of Occupational and Environmental Medicine, Department of Laboratory Medicine, Lund University, Lund, Sweden, [6]Cancer Research, National Institute for Health and Care Research (NIHR) Experimental Cancer Medicine Centre, Birmingham, United Kingdom

KEYWORDS

omics, diagnostic, biomarker, therapeutic, metabolic disease

Editorial on the Research Topic
Integrative multi-modal, multi-omics analytics for the better understanding of metabolic diseases

In the past few years, large-scale, high-throughput multi-omics experiments and improved clinical measurements have led to the generation of a plethora of multi-modal datasets related to various metabolic diseases (MetS), for example, type 1 diabetes (T1D), obesity, non-alcoholic fatty liver disease (NAFLD), etc. Many integration strategies have been discussed in the literature, such as early, intermediate, and late integration (1). The process of early integration involves merging multiple omics information sources into a unified matrix, whereas intermediate integration involves transforming the source datasets into representations that are both common and specific to omics. Late integration consists of the individual analysis of each omics dataset, followed by the combination of their respective predictions to obtain a result (1). Figure 1 shows an example of late integration described in the context of MetS.

This editorial summarizes the contributions to the Research Topic of Frontiers in Endocrinology, "*Integrative Multi-Modal, Multi-Omics Analytics for the Better Understanding of Metabolic Diseases*", between November 2021 and July 2023. This Research Topic aims to provide a platform for researchers in multi-omics and multi-modal analysis of metabolic diseases to identify targets for therapy and diagnosis. A total of 10 original articles were selected for publication from the submissions received. A summary of each manuscript is detailed below.

We have grouped the papers based on the different types of diseases they address, starting with hyperlipidemia, which is characterized by elevated levels of lipids, for example, cholesterol and triglycerides. The study performed by Zhai et al. investigated

10.3389/fendo.2023.1266557

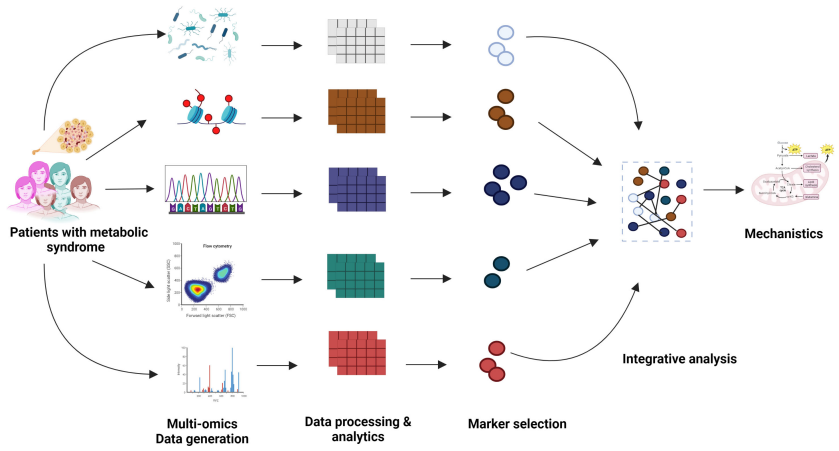


FIGURE 1
Graphical representation of the multi-modal integration to understand the mechanistic aspects of the metabolic disease.

the effects of policosanol on the control of hyperlipidemia, gut microbiome composition, and metabolic status using a C57BL/6 mouse model. Serum triglycerides, total cholesterol, and brown adipose tissue weight were all shown to be dramatically reduced after policosanol administration. Overall, policosanol demonstrated to the potential to alter the makeup of the gut microbiota, speed up fat breakdown, and differentiate thermogenesis-related gene activity.

It has been noted that vitamin D also affects patients with metabolic syndrome, and Amirkhizi et al. conducted a case-control study to investigate the relationship between vitamin D deficiency and adipokines, atherogenesis indicators, and metabolic syndrome factors in 195 patients with metabolic syndrome. The study demonstrated that patients with MetS and vitamin D insufficiency (cases) had higher AIP and LAP than controls. Moreover, vitamin D deficiency was correlated with some of the cardiometabolic risk factors in patients with MetS.

T1D is characterized by multiple factors, i.e., genetics, lifestyle, and metabolism. Clos-Garcia et al. investigated the gut microbiome and blood metabolome in individuals with T1D and healthy controls. The study stratified T1D cases based on albuminuria levels, identifying 51 species with and without albuminuria. Plasma metabolomics analysis identified differences in steroidogenesis, glucose metabolism, and circulating sphingolipids in subjects with T1D. Furthermore, the analysis revealed reduced interactions between the gut microbiome and plasma metabolome profiles, while polar metabolite, lipid, and bacteriome compositions contributed to the variance in albuminuria levels among T1D individuals. Zhang et al. analyzed the serum metabolic profiles of children with T1D diabetes and healthy controls and identified many differential metabolites, such as carbohydrates, indoles, unsaturated fatty acids, amino acids, and organic acids, consistently across pediatric patients.

The transcriptome and proteome also play a role in MetS, for example, in patients with non-alcoholic steatohepatitis (NASH). Pyo and Choi found that upregulated genes were associated with inflammation, steatosis, apoptosis, and extracellular matrix organization, while downregulated genes were associated with the response to metal ions and lipid and amino acid metabolism. Functional enrichment analysis revealed amino acid metabolism as the most significant hepatic perturbation in both human and mouse NASH.

Wang et al. developed a new obesity measurement index for metabolically associated fatty liver disease (MAFLD), which includes traditional BMI. The results showed a strong correlation between MAFLD, traditional BMI, and the new index.

Epithelial-mesenchymal transition (EMT) is a critical event in the migration and invasion of endometriosis, involving immune and stromal cells. Quan et al. investigated the potential use of EMT-based classification for the precise diagnosis and treatment of peritoneal endometriosis. A total of 76 peritoneal endometriosis samples were classified into two clusters based on EMT hallmark genes. EMT scores and abundances were compared, and a diagnostic model was constructed based on 9 markers related to immune and stromal scores.

Pan et al. developed a targeted method for the accurate quantification of 80 bile acids in gastric cancer patients with the aim of developing diagnostics for the early screening of GC. The panel of six bile acids (ratio), which included HCA, TLCA, NorCA, DCA-3G, TLCA-3S, and HDCA/LCA, showed high accuracy for the diagnosis of gastric cancer.

Yuan et al. investigated the causal relationship between rheumatoid arthritis (RA) and atlantoaxial subluxation (AAS), identifying and quantifying the potential involvement of C-reactive protein as a mediator. The researchers utilized a genome-wide association study, a two-sample Mendelian randomization (MR) analysis of genetically predicted rheumatoid arthritis, and an AAS. The MR analysis revealed a higher genetically predicted risk of rheumatoid arthritis and an increased risk of AAS. There was no convincing evidence that genetically predicted AAS affected the risk of rheumatoid arthritis.

Pirzada et al. used machine learning to discover potential inhibitors of GSK3, a protein involved in the replication and assembly of the nucleocapsid protein of SARS-CoV-2 and other coronaviruses. They used a dataset of FDA-approved and investigational pharmaceuticals from the ChEMBL database and a variety of molecular descriptors to define the inhibitors. Based on

their predicted activity, selinexor and ruboxistaurin were identified as the two most promising candidates. The study demonstrated that this virtual high-throughput screening approach based on artificial intelligence can accelerate drug discovery and identify novel targets.

In summary, this Research Topic presented multiple MetS-related studies. Multi-omics and multi-modal datasets, such as clinical data (Clos-Garcia et al.), transcriptomics (Pyo and Choi), metabolomics (Clos-Garcia et al.), and gut microbiome datasets (Clos-Garcia et al.), have been utilized to understand metabolic diseases. Thus, the results of these studies improve the prospects for future therapeutic and translational studies (2).

Author contributions

AA: Conceptualization, Funding acquisition, Investigation, Supervision, Visualization, Writing – original draft, Writing – review & editing, Formal Analysis, Methodology, Validation. PA: Conceptualization, Data curation, Investigation, Writing – original draft, Writing – review & editing. GG: Conceptualization, Funding acquisition, Investigation, Writing – original draft, Writing – review & editing.

Funding

The authors acknowledge support from the NIHR Birmingham ECMC, NIHR Birmingham SRMRC, Nanocommons H2020-EU (731032), MAESTRIA (Grant agreement ID 965286), HYPERMARKER (Grant agreement ID 101095480), PARC (Grant Agreement No. 101057014), and the MRC Heath Data Research UK (HDRUK/CFC/01), an initiative funded by UK Research and Innovation, the Department of Health and Social Care (England) and the devolved administrations, and leading medical research charities. The views expressed in this publication are those of the authors and do not necessarily represent those of the NHS, the National Institute for Health Research, the Medical Research Council, or the Department of Health.

Conflict of interest

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

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

Publisher's note

References

1. Picard M, Scott-Boyer MP, Bodein A, Périn O, Droit A. Integration strategies of multi-omics data for machine learning analysis. *Comput Struct Biotechnol J* (2021) 19:3735–46. doi: 10.1016/j.csbj.2021.06.030

2. Bravo-Merodio L, Acharjee A, Russ D, Bisht V, Williams JA, Tsaprouni LG, et al. Translational biomarkers in the era of precision medicine. *Adv Clin Chem* (2021) 102:191–232. doi: 10.1016/bs.acc.2020.08.002

frontiers in Endocrinology

ORIGINAL RESEARCH
published: 11 October 2021
doi: 10.3389/fendo.2021.722055

Integrated Metagenomics and Metabolomics to Reveal the Effects of Policosanol on Modulating the Gut Microbiota and Lipid Metabolism in Hyperlipidemic C57BL/6 Mice

Zhenya Zhai[1,2†], Jianping Liu[1†], Kai-Min Niu[1,2], Chong Lin[1], Yue Tu[1], Yichun Liu[1], Lichuang Cai[1], Huiping Liu[3] and Kexian Ouyang[1*]*

[1] Jiangxi Functional Feed Additive Engineering Laboratory, Institute of Biological Resource, Jiangxi Academy of Sciences, Nanchang, China, [2] Key Laboratory of Agro-Ecological Processes in Subtropical Region, Institute of Subtropical Agriculture, Chinese Academy of Sciences, National Engineering Laboratory for Pollution Control and Waste Utilization in Livestock and Poultry Production, Changsha, China, [3] Era Biotechnology (Shenzhen) Co., Ltd., Shenzhen, China

The aim of the study was to investigate the regulatory effects of policosanol on hyperlipidemia, gut microbiota and metabolic status in a C57BL/6 mouse model. A total of 35 C57BL/6 mice were assigned to 3 groups, chow (n=12), high fat diet (HFD, n=12) and HFD+policosanol (n=11), then treated for 18 weeks. Policosanol supplementation significantly reduced serum triglycerides and total cholesterol, as well as the weight of brown adipose tissue (BAT) ($p<0.05$), without affecting body weight in HFD-fed mice ($p>0.05$). Combined 16S rRNA gene sequencing and untargeted metabolomic analysis demonstrated that policosanol had regulatory effects on gut microbiota and serum metabolism in mice. In obese mice, policosanol increased the proportion of *Bacteroides*, decreased the proportion of *Firmicutes*, and increased the ratio of *Bacteroides* to *Firmicutes* ($p<0.05$). Policosanol promoted lipolysis and thermogenesis process, including tricarboxylic acid (TCA) cycle and pyruvate cycle, correlated with the increasing level of *Bacteroides*, *Parasutterella*, and decreasing level of *Lactobacillus* and *Candidatus_Saccharimonas.* Moreover, policosanol decreased fatty acid synthase (FAS) in the iWAT of obese mice. Policosanol also increased peroxisome proliferators-activated receptor-γ (PPARγ), uncoupling Protein-1 (UCP-1), peroxisome proliferator-activated receptor gamma coactivator-1α (PGC-1α) and PR domain containing 16 (PRDM16) in brown adipose tissue (BAT) obese mice ($p<0.05$). This study presents the new insight that policosanol may inhibit the synthesis of fatty acids, and promote lipolysis, thermogenesis related gene expression and regulate gut microbiota constituents, which provides potential for policosanol as an antihyperlipidemia functional food additive and provide new evidence for whole grain food to replace refined food.

Keywords: policosanol, gut microbiota, metabolomics, C57BL/6 mouse, antihyperlipidemia

OPEN ACCESS

Edited by:
Claire Joanne Stocker, Aston University, United Kingdom

Reviewed by:
Zhipeng Liu, Purdue University, United States
Edward Wargent, University of Buckingham, United Kingdom

****Correspondence:***
Kexian Ouyang
ziyuanhuaxuelab@163.com
Huiping Liu
546033730@qq.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Obesity, a section of the journal Frontiers in Endocrinology

Received: *09 June 2021*
Accepted: *14 September 2021*
Published: *11 October 2021*

Citation:
Zhai Z, Liu J, Niu K-M, Lin C, Tu Y, Liu Y, Cai L, Liu H and Ouyang K (2021) Integrated Metagenomics and Metabolomics to Reveal the Effects of Policosanol on Modulating the Gut Microbiota and Lipid Metabolism in Hyperlipidemic C57BL/6 Mice. Front. Endocrinol. 12:722055. doi: 10.3389/fendo.2021.722055

INTRODUCTION

With the gradual improvement of living standards of people in most countries, obesity induced by high-fat, high-sugar and refined diets is becoming a global problem (1). According to the World Health Organization, the number of obese people in the world nearly doubled between 1998 and 2008, with 9 million obese adolescents in the world (2). According to the data of the China Centers for Disease Control and Prevention, during 2013-2014, approximately 14% of Chinese people were also suffering from obesity (3). Hyperlipidemia, which mainly reflects the high cholesterol, high triglyceride content and lipid metabolism disorder in blood, is a major complication of obesity (4, 5). Long-term hyperlipidemia can induce nonalcoholic fatty liver and other diseases, which are harmful to human health.

The microbiome was found regulated by diets and has essential roles on impacting obesity by influencing the caloric absorption and energy expenditure (6). In germ-free mouse model, the bacteria transplantation dramatically increased the body weight and decreased the feed intake (7). However, a higher proportion of *Firmicutes* relative to *Bacteroidetes* was found in obese people and when obese people accepted a low-calorie diet and lose weight, this phenomenon was reversed (8, 9). Microbiome-metabolic axis plays important role in obesity (10). The metabolites of intestinal microbiota, such as bile acids, butyric acid, monosaccharide and vitamins, have been found to promote or ameliorate obesity (11, 12). For example, butyrate, a short-chain fatty acid produced by microbial degradation of carbohydrates, was found to upregulate downstream genes such as UCP1, PGC-1α and PRDM16 related to lipolysis and thermogenesis (13, 14). Secondary bile acids, metabolites of gut microbiota from primary bile acids, have been found to enhance the expression of downstream lipolysis and thermogenesis genes by activating bile acid receptors. There is growing consensus that the lipid metabolism and energy expenditure of the host can be regulated and improved by interfering the structure of intestinal microbiota.

Policosanol is a kind of long-chain fatty acid alcohols (LFAs) including octacosanol ($C_{28}H_{57}OH$, the most abundant component), tetracosanol ($C_{24}H_{49}OH$), triacontanol ($C_{30}H_{61}OH$) and tetratriacontanol ($C_{34}H_{69}OH$), that has been considered a functional food with hypolipidemic, antiobesity, antihypercholesterolemia, and antihypolipidemic activity and has been used as a potential adjuvant drug for type 2 diabetes mellitus in the past two decades (15). Policosanol exists widely in rice bran, beeswax and other natural products (15). However, compared with whole grain foods, policosanol in refined grains or related foods is almost completely lost (16). Policosanol has been proven to be absorbed easily, and metabolized by animals and to exert activity on blood lipids. In a mouse model, octacosanol can be absorbed after oral administration, the plasma peak was reached in 30-60 min and existed in the body for more than 3 hours. is mainly concentrated in the liver, adipose tissue and digestive tract (17). Furthermore, octacosanol, the main component of policosanol, has been reported to reduce blood cholesterol levels, insulin resistance, and low-density lipoprotein cholesterol (LDL-C) levels after oral administration in humans (18). Similarly, in children with hypercholesterolemia, oral administration of policosanol can significantly reduce the levels of TC, LDL and apolipoprotein B (19). Furthermore, in a rat model and in Hep G2 cell models, octacosanol has been found to inhibit cholesterol synthesis (20). These studies provide clues that policosanol can regulate blood lipid and cholesterol levels and is targeted mainly to the liver and fat tissues, however, whether policosanol has a regulating effect on gut microbiota and serum metabolism is still no clear.

To clarify this situation, we attempted to investigate the regulatory effects of policosanol on hyperlipidemia, gut microbiota and serum metabolic status in mice by integrated microbiome-metabolomic methods, coupled with serum biochemical and gene expression analyses in the present study.

MATERIALS AND METHODS

Ethics Statement

All the experimental design, procedures and experimental operations in the present study were approved and in accordance with the guidelines of the Committee of the Institute of Subtropical Agriculture at the Chinese Academy of Science (No. ISA-2020-18).

Preparation of Policosanol

Policosanol was prepared by a distillation method. Briefly, we constructed a vacuum high-temperature distillation column. The saponified crude alcohol used as raw material was added into the material tank, high-temperature distillation was carried out in the distillation column, and policosanol with different purities was collected in different collectors. The Analyzing and Testing Center of Guangzhou Institute of Chemistry, Chinese Academy of Sciences, was entrusted to test the components of policosanol (report number: YS160503-03). The policosanol used in this study was composed of docosanol ($C_{22}H_{45}OH$, 2.48%), tetradecanol ($C_{24}H_{49}OH$, 4.19%), hexacosanol ($C_{26}H_{53}OH$, 4.33%), octacosanol ($C_{28}H_{57}OH$, 64.16%) and triacontanol ($C_{30}H_{61}OH$, 15.51%). Because policosanol is a waxy solid, it is necessary to use a vehicle to improve its mixing uniformity in the diet. The vehicle includes mainly cyclodextrin, isomaltooligosaccharide and Arabic gum. The ratio of vehicle and policosanol was 1:9 (weight: weight).

Animals

A total of 35 male C57BL/6 specific pathogen-free mice (3-4 weeks of age) were purchased from STJ Laboratory Animal Co., Ltd. (Hunan, China). All mice were housed in cages and raised under the same controlled conditions (temperature 25 ± 2°C, light/dark 12 h:12 h, humidity of 60 ± 10%). After 1 week of adaptation, the mice were randomly assigned to 4 groups with equally adjusted initial body weights as follows: the chow group was fed a chow diet containing vehicle (17.81 ± 0.16 kJ/g, n=12), the HFD group was fed an high-fat diet containing vehicle (24.02 ± 0.09 kJ/g, n=12), and (4) the HFD+policosanol group was fed HFD containing 0.5% policosanol (23.97 ± 0.06 kJ/g, n=11). All the diets were customized by Jiangsu Xietong Pharmaceutical Bio-Engineering

Co., Ltd. The normal and HFD formulas are shown in **Supporting Information Tables S1**, **S2**. All the mice were raised with free access to feed and water and body weight and food intake of the mice were measured once a week for 18 weeks.

Sample Preparation

At the end of this experiment, all the mice were fasted for 6 h, and anesthesia was induced by intraperitoneal injection of 2% pentobarbital sodium intraperitoneally injection (45 mg/kg body weight). Then, blood samples were collected after enucleation of the eyeball. Then, the inguinal white adipose tissue (iWAT), epididymal white adipose tissue (eWAT) and brown adipose tissue were separately collected and weighed, as described in our previous study (12, 21). iWAT was completely separated from the subcutaneous of the abdomen. Further open the abdominal cavity and completely separate eWAT from around the epididymis. Finally, the skin of the scapula was cut to completely separate the brown fat. The iWAT samples were placed into 4% paraformaldehyde and embedded in paraffin and then subjected to hematoxylin and eosin (H&E) staining. Adipocyte size was measured using VistarImage (Olympus, Japan) following the instructions. The blood, adipose tissue and cecal content samples for metabolomics analysis, qPCR or 16s rRNA were collected with centrifuges (germ-, RNase- and DNase- free) tube, frozen in liquid nitrogen and stored at - 80 °C.

Lipid Parameters in Serum

The triglyceride (TG), total cholesterol (TC), high -density lipoprotein cholesterol (HDL-C) and LDL-C in serum were determined according to the instructions of the commercial kits (Nanjing Jiancheng, Bioengineering Institute, Nanjing, China).

16S rRNA Gene Sequencing

Detailed sequencing methods were described in our previous research (22) and performed in a commercial company (Novogene, Beijing, China). Briefly, total DNA was extracted from the cecum contents by using Qiangen QIAamp DNA Stool Mini Kit according to the protocol. Then the V3-V4 region of the bacterial 16S ribosomal RNA gene amplified by PCR, the following primers (5'-3') 180 were used: 341F CCTAYGG GRBGCASCAG and 806R 181 GGACTACNNGGGTATC TAAT. After the PCR, and the amplicons were electrophoresed and extracted. The gene library was constructed by using the Ion Plus Fragment Library Kit 48 rxns 188 (Thermo Scientific, Waltham, MA, USA). Then the sequencing of genes was performed on the Ion S5^{TM}XL platform to obtain the raw data. The raw data were treated and measured to clean data. The clean data were clustered with 97% identity to identify the operational taxonomy units (OTUs).

Linear discriminant analysis (LDA) effect size (LEfSe) was used to elucidate the differences in bacterial taxa, which were described in our previously study (22). An LDA score ≥3 was considered to be an important contributor to the model. Spearman analysis was used to measure the correlation between gut microbiota and adipose tissue weight or serum metabolites using R software.

Untargeted Metabolomics in Serum

Untargeted metabolomics was conducted using a commercial service from Biotree (Shanghai, China). Briefly, a 50 μL serum sample was mixed with 200 μL extracting solution (50% methanol: 50% acetonitrile) and internal standards [L-leucine-5,5,5-d3 (CAS:87828-86-2), trimethylamine-d9-N-oxidein (CAS: 1161070-49-0)] in a 1.5 mL centrifuge tube. The mixture was sonicated in an ice-water bath for 10 min and placed at -40°C for 1 h. Then, the samples were centrifuged at 4°C and 12000 rpm for 5 min, and the supernatant was subjected to liquid chromatography-mass spectrometry (LC-MS).

The LC was performed using a Vanquish ultra-performance liquid chromatography instrument (Thermo Fisher Scientific, MA, US) with a Waters ACQUITY UPLC BEH Amide (2.1 mm × 100 mm, 1.7 μm) chromatographic column to separate the metabolites. A phase (ultrapure water containing 25 mmol/L ammonium acetate and 25 mmoL/L ammonia) and B phase (acetonitrile) were used for elution. The elution gradient is shown in **Supporting Information (Table S2)**. The column temperature was 30°C, and the injection volume was 3 μL.

A QE HFX mass spectrometer was used to obtain MS/MS spectra in information-dependent acquisition (IDA) mode under the control of the acquisition software (Xcalibur, Thermo Fisher Scientific). In this mode, the acquisition software continuously evaluates the full scan MS spectrum. The ESI source conditions were set as follows: sheath gas flow rate of 50 Arb, Aux gas flow rate of 10 Arb, capillary temperature of 320°C, full MS resolution of 60000, MS/MS resolution of 7500, collision energy of 10/30/60 in NCE mode, and spray voltage of 3.5 kV (positive) or -3.2 kV (negative).

The raw data were converted to an mzXML format, and R software was used for peak detection, extraction alignment and integration. The metabolites were annotated using an in-house database. The peaks were normalized using internal standard.

Bioinformatic Analysis

To analyze the differences in gut microbiota and serum metabolites in different groups, principal component analysis (PCA), and orthogonal partial least squares discriminant analysis (OPLS-DA) were performed and the over fitting of the model is verified by cross-validation (permutation test) using SIMCA-P (16.0.2, Sartorius Stedim Data Analytics AB, Umea, Sweden) (23). Then, univariate statistical analysis was used to screen metabolic markers with significant differences in accordance with our previous study (24, 25). Briefly, based on OPLS-DA, the variable importance in the projection (VIP)>1, the relative abundance fold change of the metabolites of <0.5 or >2, and a p-value <0.05 were defined as differences. Furthermore, the significantly changed metabolite pathways were analyzed using the KEGG database (Kyoto Encyclopedia of Genes and Genomes).

RNA Extraction and Quantitative Real-Time Polymerase Chain Reaction

The iWAT and BAT were homogenized and total mRNA was extracted using column RNA extraction kits (Magen,

Guangzhou, China, R4121). The total RNA concentration was determined using a NanoDrop 2000C spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA), and the mRNA was reverse transcribed to cDNA using cDNA synthesis kits (CWBIO, Jiangsu, China, CW2582M). Information on the primers is shown in **Supporting Information (Table S6)** and synthesized (Sangon Biotech, Shanghai, China). qPCR was performed in an Applied Biosystems by Life Technologies QuantStudio 7 (Thermo Fisher Scientific). The relative abundance of the genes was normalized to GAPDH using the 2-ΔΔCT method and the relative expression level were shown as fold changes relative to chow group.

Statistical Analysis

Data are expressed as the mean ± SEM. One-way ANOVA and least significance difference (LSD) method as a suitable *post-hoc* test was used to determine the differences among the groups by using SPSS 20.0 (IBM, SPSS, USA) and a p-value of <0.05 was considered statistically significant. GraphPad Prism 7 (GraphPad Software Inc., San Diego, CA, USA) was used to generate statistical plots.

RESULTS

Effects of Policosanol on Hyperlipidemia in HFD-Fed Mice

C57BL/6 male mice aged 3-4 weeks were fed 0.5% policosanol in a chow diet or HFD for 18 weeks. The data showed that the body weight of HFD-fed was mice significantly higher than the body weight of chow-fed mice ($p<0.05$). However, policosanol treatment did not significantly affect changes in body weight, or total energy intake in HFD-fed mice ($p>0.05$) (**Figures 1A, B**). In serum, the TG and TC contents were both significantly reduced in the HFD-fed mice ($p<0.05$, **Figures 1C, D**), while decreased LDL-C content and increased HDL-C content were observed in HFD-fed mice ($p<0.05$, **Figures 1E, F**) after policosanol treatment.

Effects of Policosanol on Body Fat in HFD-Fed Mice

The accumulation of adipose tissue in mice was evaluated. Policosanol treatment showed no effect on eWAT weight ($p>0.05$, **Figure 2A**), tended to reduce the weight of iWAT ($p=0.056$, **Figure 2B**), and significantly decreased the BAT weight ($p<0.05$) in the HFD-fed mice (**Figure 2C**). H&E staining-based histological results for both groups showed that policosanol reduced the iWAT adipocyte size ($p<0.05$, **Figures 2D, E**).

Effect of Policosanol on Gut Bacterial Communities in Mice

The16S rRNA gene sequencing data are shown in **Figure 3**. The rarefaction and observed OTU results showed that the selective sequences in clean data were sufficient to determine most of the bacterial species (**Supporting Information Figures S1 A, B**). The petal diagram showed that there were 544 core OUTs found in all the groups, while 57, 74 and 38 OTUs were detected in chow, HFD and HFD+policosanol groups (**Figures 3A** and **S1C**). Policosanol increased *Bacteroides*, decreased the *Firmicutes* relative abundance, and increased the B/F ratio in obese mice ($p<0.05$, **Figures 3B, C**). To better understand the effect of policosanol on microbiota constituents, LEfse was conducted (**Figure 3D**). The data showed that *Bacteroidales* and *Muribaculaceae* belonging to *Bacteroidota* phylum were the feature taxa in the chow fed mice. In HFD group, *Lactobacillaceae*, *Actinobacteria* and *Erysipelotrichales* were the feature taxa. In HFD+policosanol group, *Akkermansiaceae* belonging to *Verucomicrobiales*, *Bacteroidaceae* and *Desulfovibrionia* were the feature taxa. Based on this, a spearman correlation analysis was conducted between microbiota composition and adipose tissue weight. The data showed that *Lactobacillus*, *Dubosiella*, *Candidatus: Sarccharinonasm* and *Lachnospiaceae_UGG.006* showed positive correlation with iWAT and BAT weight ($R^2>0.4$), while *Alloprevotella*, *Parasutterella*, *Oscillibacter* and *Faecalibacterium* showed negative correlation ($R^2<-0.4$). Furthermore, compared to HFD group, policosanol was found had lowering-effect on the relative abundance of *Lactobacillus* and *Candidatus:Sarccharinonasm* compared to HFD group (**Figures 3F, G**, $p<0.05$), while up-regulating effect on the relative abundance of *Bacteroides* and *Parasutterella* ($p<0.05$, **Figures 3I, J**). It is also found that, compared to HFD group, policosanol treatment tend to increase the relative abundance of *Allpprevotella* and *Akkmermansia* ($0.05<P<0.1$, **Figures 3J, K**).

Effects of Policosanol on Serum Metabolites in HFD-Fed Mice Uncovered by Untargeted Metabolomics

Untargeted metabolomics with both positive and negative models was subsequently conducted to investigate the effects of policosanol on influencing serum metabolite changes related to lipid and energy metabolism. The PCA and OPLS-DA results showed discriminately distinguished serum metabolites in HFD-fed mice (**Figures 4A, B**). The permutation test showed that all OPLS-DA models were reliable without overfitting (**Figure 4C**). The volcano plot (**Figure 4D**) presented significantly changed serum metabolites after treatment with policosanol. In HFD-fed mice, we found that 483 metabolites were decreased and 359 metabolites were increased after policosanol treatment.

Effects of Policosanol on Changing the Metabolic Pathways in HFD-Fed Mice

Based on the PCA, the relative abundance of metabolites between HFD and HFD + policosanol mice were shown to be remarkably different. Thus, relevant metabolic pathways were further analyzed (**Figure 5**). In HFD-fed mice, policosanol was found to regulate mainly the TCA cycle, butanoate metabolism, alanine, aspartate and glutamate metabolism, pyruvate metabolism, pantothenate and CoA biosynthesis and glycerophospholipid metabolism pathways under the negative scan model, while the

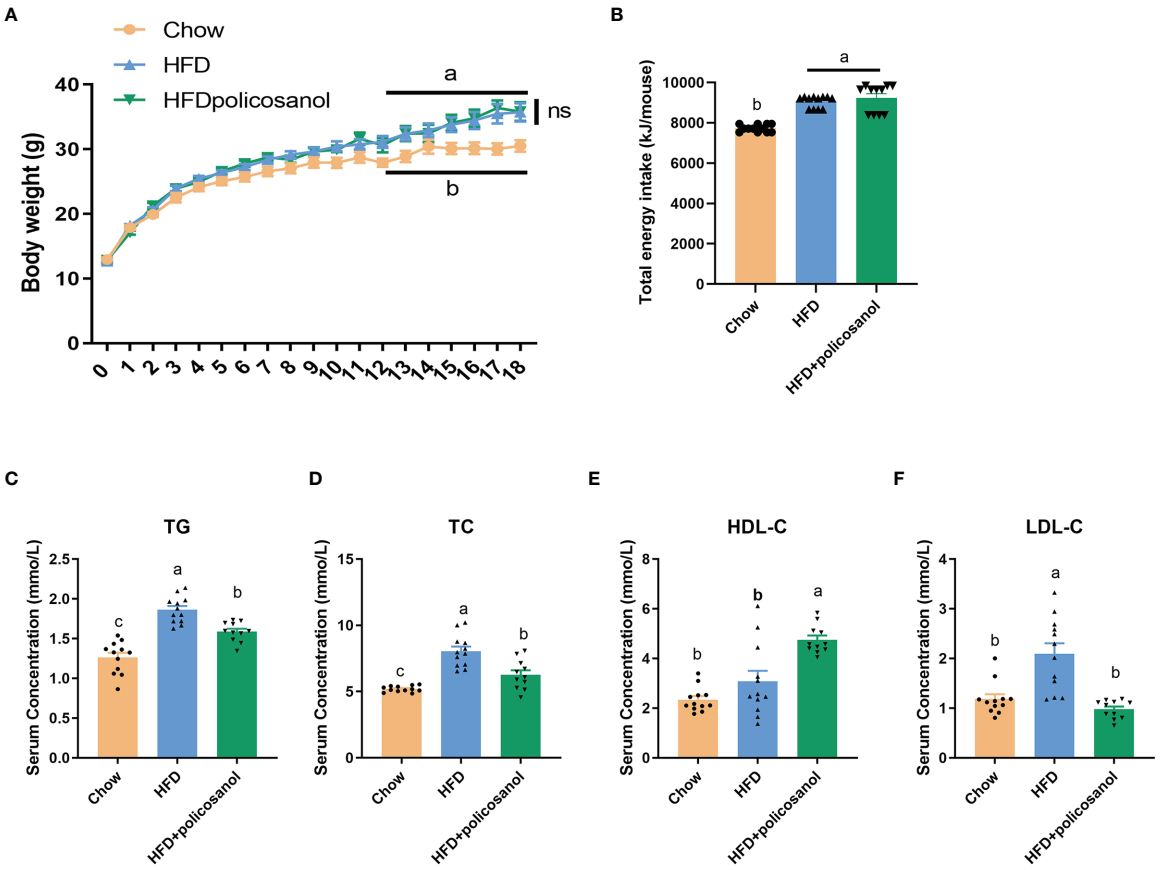


FIGURE 1 | Policosanol did not affect body weight, but showed the lipid lowering effect in the serum of mice. **(A)** Policosanol had no effect on the body weight of obese mice for 18 weeks. **(B)** The total energy intake (per mice) of chow diet or HFD fed mice after 18 weeks. **(C–F)** Policosanol reduced serum TG, TC, LDL-C, HDL-C in HFD induce obese mice for 18 weeks. Chow: The mice fed with chow diet (n=12), HFD: the mice fed with high fat diet (n=12), HFD+policosanol: The mice fed with high fat diet containing 0.5% policosanol (n=11). TG, Total glyceride; TC, total cholesterol; LDL-C, low density lipoprotein cholesterol; HDLC, high density lipoprotein cholesterol. Data were expressed as mean ± SEM. Differences of data in mice subjects were assessed by one-way ANOVA. For all pictures, columns indicated with different letters (a- c) have significant difference, $p < 0.05$. ns, not significant.

regulation related to glycerophospholipid metabolism, arginine and proline metabolism, and cysteine and methionine metabolism-related pathways was determined under the positive the scan model. Small molecule metabolites in mouse cecal contents were also detected. The data showed that policosanol treatment mainly affect the linoleic acid metabolism, biosynthesis of unsaturated fatty acids, propanoate metabolism and amino acid metabolism (**Supplementary Figure S2**).

Effects of Policosanol on the Regulation of Representative Metabolites in Chow- and HFD-Fed Mice

Further, we combined OPLS-DA, t-test and serum content to screen key small molecule metabolites that may be regulated by policosanol. A total of 30 metabolites were primarily identified (**Figure 6**) and all the significantly changed metabolites found in this study were shown in **Supporting Information Tables S5**.

In HFD-fed mice, metabolites involved in nucleotides, amino acids, fatty acids, and their intermediates that are associated with mitochondrial oxidation, such as 3−methyl−2−oxovaleric acid, citraconic acid, trans−aconitic acid, and pyruvic acid, were significantly enriched with policosanol treatment compared to the control. Conversely, glycerophospholipid metabolism-related metabolites such as 2,3−dihydroxybutanedioic acid, pantothenic acid, L−carnitine, choline, phosphorylcholine, and trimethylamine oxide (TMAO), were significantly downregulated. The data showed that, policosanol significantly decreased the content of lipids and lipid-like molecules (e.g., 12-HEPE and 12-HETE, the derivates of eicosapentaenoic acid) and increased the content of linoleic acid and ricinoleic acid ($p<0.05$, **Supporting Information Figure S3** and **Supplementary Table S6**) in the cecal content of mice.

Correlation of Key Microbiota Communities and Differential Serum Metabolites

To better reveal how altered microbiota regulate host metabolic processes in mice, correlation analysis of microbiome-

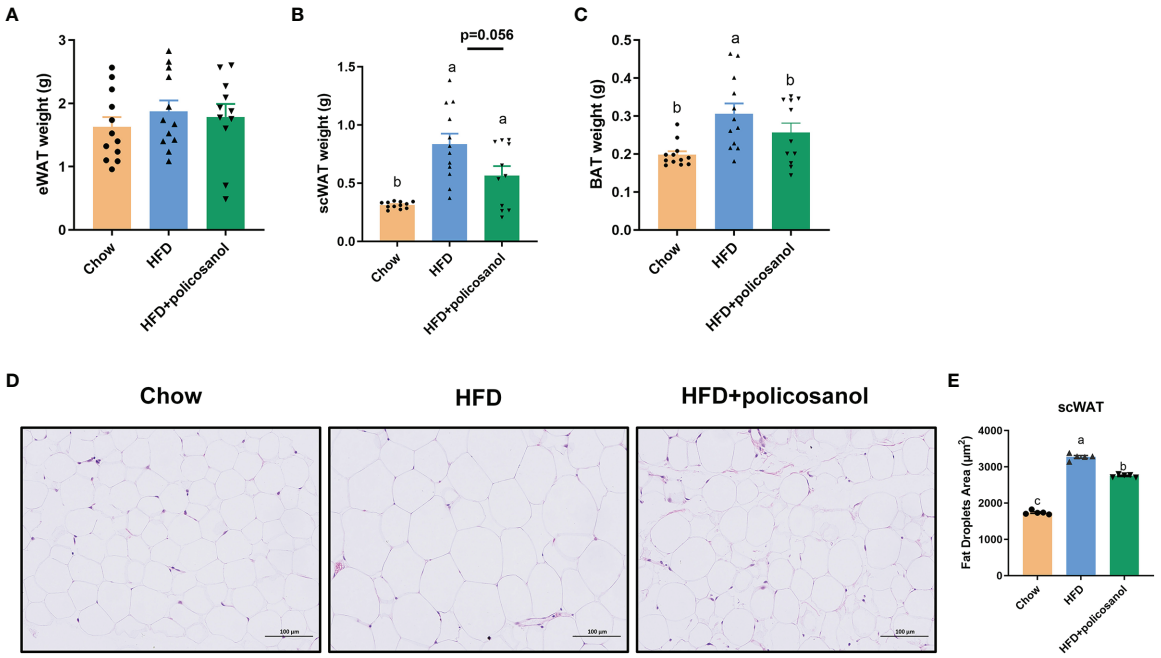


FIGURE 2 | Policosanol reduced lipid accumulation in the iWAT and BAT. **(A–C)** The tissue weight of eWAT, iWAT and BAT. Chow: The mice fed with chow diet (n = 12), HFD: the mice fed with high fat diet (n = 12), HFD + policosanol: The mice fed with high fat diet containing 0.5% policosanol (n = 11). **(D)** Representative H&E staining picture of adipocyte size of iWAT. **(E)** Statistical plot of fat droplet area in scWAT, n = 5 mice/group. Differences of data in mice subjects were assessed by one-way ANOVA. For all pictures, columns indicated with different letters (a-c) have significant difference, $p < 0.05$. ns, not significant.

metabolomic studies were conducted in obese mice (**Figure 7**). In HFD induced obese mice, *Allobactulum* is the feature taxa in HFD + policosanol, was positively correlated with the metabolites involved in energy metabolism, such as succinic acid, trans-aconitic acid and citraconic acid. *Bacteroides* were positively correlated with pyruvic acid and uridine (R^2>0.4, $p<0.05$) but negatively correlated with trigonelline, phoshorycholine, pantothenic acid, L-cartnitine (R^2<-0.4, $p<0.05$). In contrast, *Candidatus_Sacchanimonas* and *Turicibacter*, the feature communities in the HFD group, were negatively correlated with uridine, trans-aconitic acid, succinic acid, citraconic acid (R^2<-0.4, $p<0.05$), while positively correlated with choline, pantothenic acid and 2,3-dihydroxybutanedioic acid (R^2>0.4, $p<0.05$). Inductive information on the sources of all small molecule metabolites is presented in **Supporting Information Table S7**.

Effects of Policosanol on the Regulation of Lipid Metabolism-Related mRNA Expression in Adipose Tissue

In iWAT, policosanol treatment decreased the gene expression of FAS, TNF-α and IL-6, while increased PPARγ, PGC-1α and UCP-1 in HFD-fed mice ($p<0.05$, **Figures 8A–I**). Similarly, in BAT, policosanol showed no significant effect on FAS, ATGL, HSL, ($p>0.05$), but upregulation of PPARγ, UCP1, PGC-1 and PRDM16 in obese mice after policosanol treatment ($p<0.05$, **Figures 8J–P**).

DISCUSSION

Obesity is a chronic metabolic disease that has been found to be closely related to cardiovascular disease, diabetes and gut disease since the early 20th century. To date, the causes of obesity are still unclear, but controlling energy intake, increasing energy consumption and promoting fat decomposition and oxidation have been deemed to alleviate obesity (26). In general, pharmacotherapy intervention, bariatric surgery treatment and lifestyle intervention are mainly used to treat obesity (27). However, these interventions can cause side effects such as vitamin deficiency, hyperthyroidism, headache, vomiting and hypoglycemia (28, 29). Among these interventions, lifestyle interventions, including adjustment of dietary structure and exercise, are considered safe ways to improve obesity. A large amount of evidence has shown that policosanol improves hypertriglyceridemia and hypercholesterolemia in humans and animals and has the potential as a functional food to improve obesity. For example, a double-blind experiment on volunteers with prehypertension showed that orally taking 10 mg policosanol tablets (containing 60-70% octacosanol) each day can reduce TC and TG by 20% and 14%, respectively, and the activity of cholesteryl ester transfer protein in the serum (30). Policosanol (mainly containing 28% octacosanol, 21% triacontanol, and 36% tetratriacontanol) also significantly decreased the weight of iWAT and BAT (31). The present study was conducted to investigate the effects of policosanol on

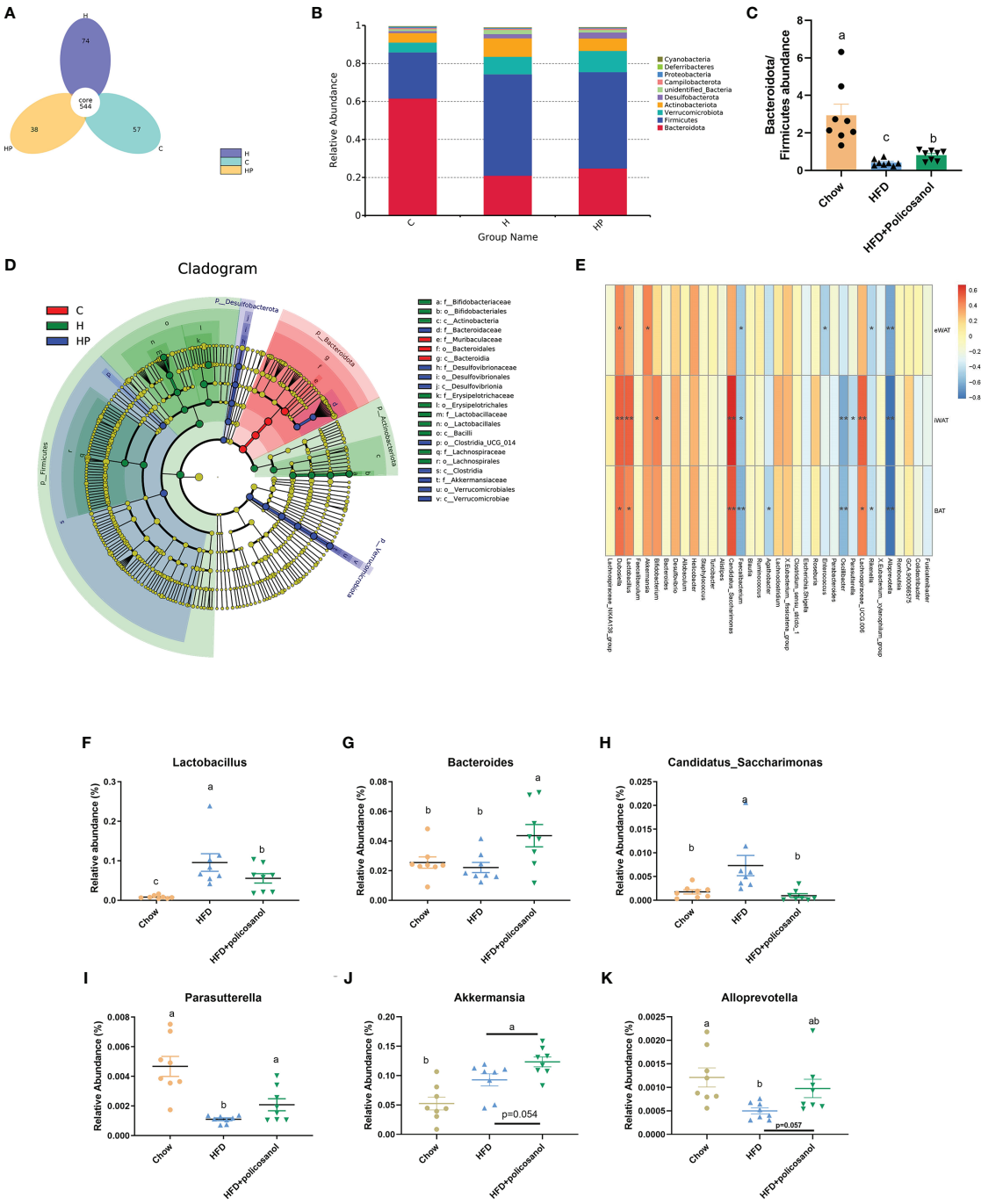


FIGURE 3 | Policosanol reshaped the gut microbiota in mice. **(A)** The Venn plot of OTUs. **(B)** Relative abundance of gut microbiota at phylum level. **(C)** The of *Bacteroides*/*Firmicutes* ratio in the cecal contents of mice. **(D)** LEfse analysis between HFD and HFD+policosanol groups. **(E)** The correlation among microbiota and adipose tissue weight. ** : | R^2| > 0.5, *0.4 < | R^2| < 0.5 **(F–K)** The key microbiota with significant difference at genus level. **(C)** mice fed with chow diet. **(H)** mice fed with HFD. HP: mice fed with HFD+0.5% policosanol diet, n=8/mice per group. Differences of data in mice subjects were assessed by one-way ANOVA. For all pictures, columns indicated with different letters have significant difference (a-c), $p < 0.05$. ns, not significant.

metabolism *in vivo* and to explore the potential mechanisms for improving obesity. Our results showed that dietary supplementation with 0.5% policosanol (containing approximately 64% octacosanol) did not change the body weight or total energy intake of mice. This is consistent with a previous study that showed that policosanol orally administered for 6-12 months does not affect the body weight of Sprague-Dawley rats (32, 33). Similarly, octacosanol, the main component

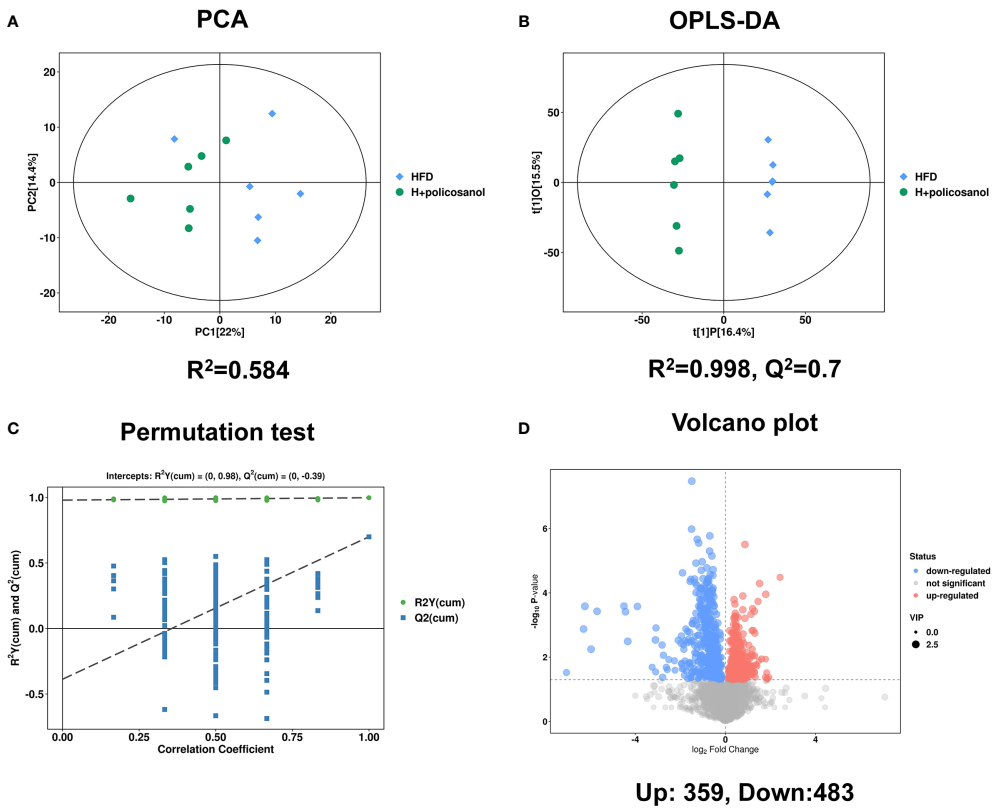


FIGURE 4 | Policosanol affects the metabolites in mice. **(A)** The principal component analysis plot based on the metabolites matrix in the serum. **(B)** The orthogonal partial least squares discrimination analysis (OPLS-DA) plot based on the metabolites matrix in the serum, **(C)** permutation test plot based on OPLS-DA methods. **(D)** The volcano plot based on the changed metabolites the HFD group compared to the HFD+policosanol group.

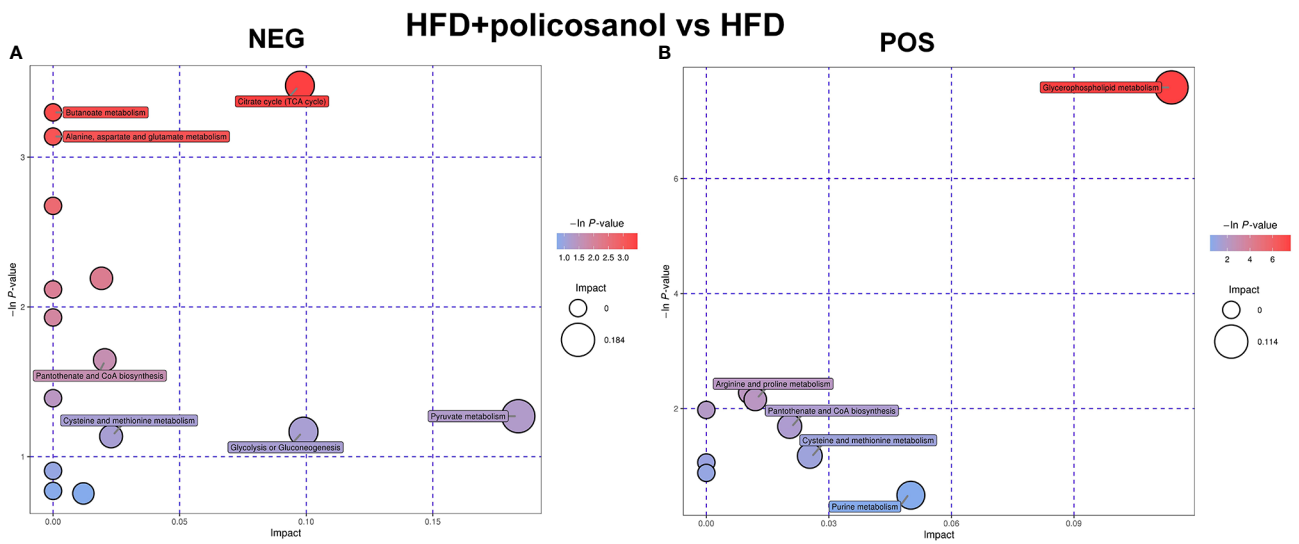


FIGURE 5 | Policosanol regulates the metabolic pathways in obese mice. **(A)** Enrichment analysis of KEGG pathways in NEG mode. **(B)** Enrichment analysis of KEGG pathways in POS mode. NEG, negative scanning mode; POS, positive scanning mode.

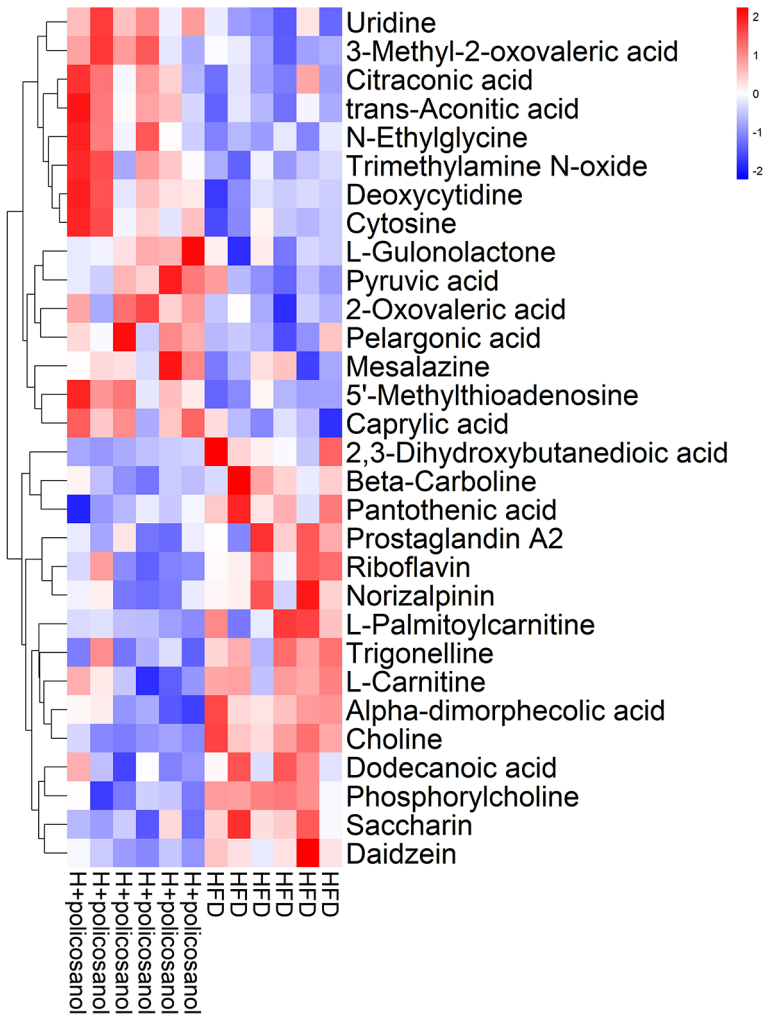


FIGURE 6 | Hierarchical clustering analysis of top the 30 metabolites in mice. Heat map of top the 30 metabolites in HFD and HFD +policosanol groups. The red and blue indicate upregulation and downregulation, respectively.

of policosanol, has no effect on mouse body weight after treatment for 12 weeks (34). However, the present study found that policosanol can effectively reduce the content of TG, TC, LDL-C and increase HDL-C content in HFD-fed mice, suggesting its potential to ameliorate hyperlipidemia.

The diversity, richness and structure of gut microbiota play an important role in the regulation of host obesity, diabetes and metabolic diseases. The individuals who were marked as overall adiposity, were found with low bacterial richness compared to the non-obese individuals (35). In the present study, we analyzed the effects of policosanol on gut microbiota in obese mice. Policosanol reduced the weight of iWAT, BAT but not eWAT. Then, the effect of policosanol was further confirmed by histological analyses showing reduced adipocyte size in iWAT. As expected, the gut microbiota of mice fed with diet (containing 0.5% policosanol) for 18 weeks changed significantly and showed change manners in obese mice. Policosanol increased *Bacteroides*/*Firmicutes* ratio only in obese mice. In addition to the proportional changes of *Bacteroidea* and *Firmicutes*, the significant change of *Parasutterella* and *Canaidatus_Saccharimonas* after policosanol intervention was highly correlated with the weight change of iWAT and BAT. *Akkermansia* and *Alloprevotella* tended to recover to a level similar to that of the control group after policosanol treatment. Researchers have found that the gut microbiota can regulate the metabolic state and health of the host through pathways such as energy metabolism regulation and release of signaling molecules. According to the previous studies, the balance between *Bacteroides* and *Firmicutes* is considered to be the key factor affecting obesity (35). Furthermore, the altered microbiota can regulate the metabolism of host by regulating the changes of various metabolites. The proportion of *Firmicutes* and *Bacteroidetes* increases, so that more plant derived polysaccharides are decomposed into short fatty acids, which affects the metabolism of carbohydrates and promotes the host's calories absorption and adipogenesis (6). *Parasutterell* can occupy a specific niche in the intestine and affect the metabolism of

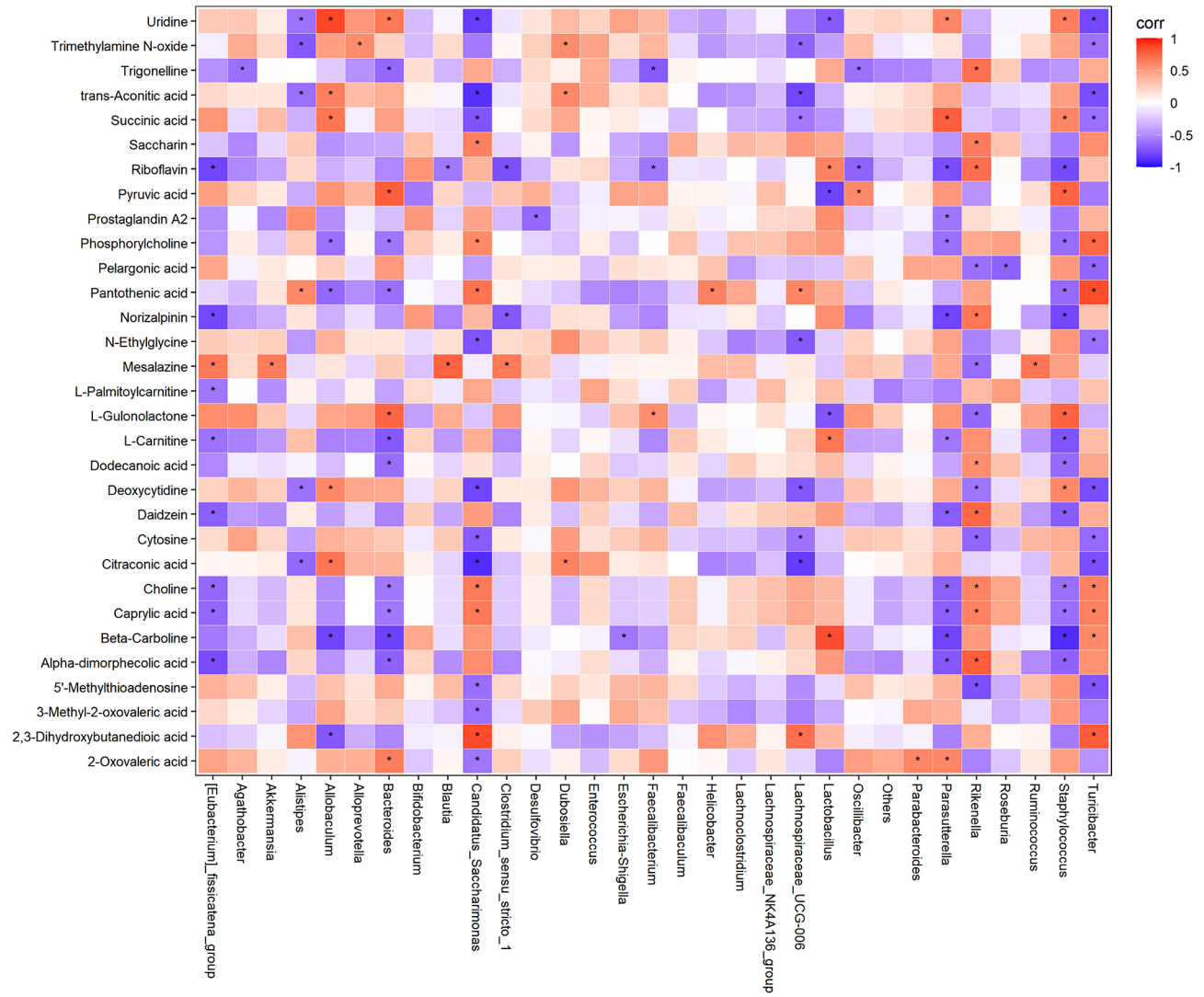


FIGURE 7 | Correlation analysis of gut microbiota and serum metabolites in HFD and HFD+policosanol mice. Red and blue indicate positive and negative correlations, respectively. *p < 0.05.

aromatic amino acids, bilirubin, succinate and bile acids in the intestine, so as to regulate the metabolism of the host. Therefore, *Parasutterell* is considered to be a core flora but has not been fully studied in recent years (36). In the present study, a combined metabolomics and microbiomics analysis was used to reveal the correlation of metabolites and microbiota. Our results clearly showed that policosanol could significantly influence metabolism in HFD-fed mice. Policosanol significantly increased metabolites associated with the TCA cycle, pyruvate metabolism, glycolysis or gluconeogenesis, and pantothenate and CoA metabolism and significantly inhibited the synthesis of glycerophospholipids. Furthermore, our results clearly revealed that *Bacteroides*, *Allobaculum*, *Staphylococcus*, and *Parasutterella* were participated in succinic acid, pyruvic acid, and critraconic acid regulation. In this study, the results also showed that *Akkmermansia* tended to be increased in HFD + policosanol group, but there was no high correlation in the regulation of metabolites in serum and intestinal contents. Lack of *Akkmermansia* is a key factor in inducing obesity and even diabetes (37). However, studies have shown that compared with living *Akkmermansia*, dead *Akkmermansia* can better regulate host fat content, insulin resistance and hyperlipidemia by releasing Amuc-1100 protein from cell membrane (38, 39).

Lipolysis and thermogenesis process in adipose tissue are important factors regulating fat deposition in mammals. In this study, *Bacteroidetes*, *Firmicutes* in phylum and *Parasutterell* and *Canaidatus_Saccharimonas* in genus level showed high correlation with iWAT and BAT weight and metabolites contents correlated with TCA cycle. Fat in scWAT is more easily metabolized than visceral WAT, which including eWAT and the adipose tissue around liver and heart (40). Compared with visceral WAT, scWAT is more likely to be mobilized,

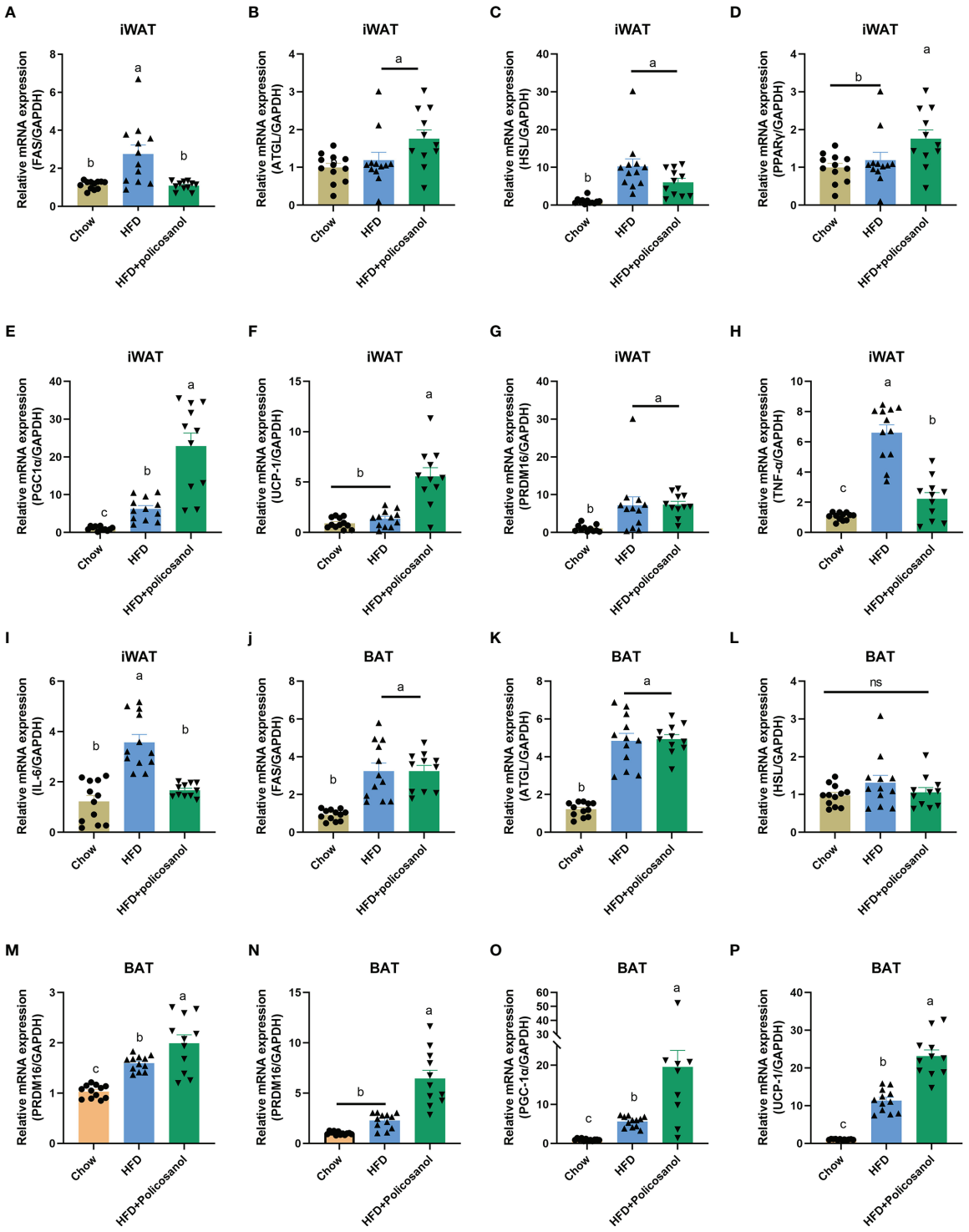


FIGURE 8 | Policosanol activates AMPK and downstream gene expression related to fatty acid synthesis, inflammation, lipolysis and thermogenesis in adipose tissue. **(A–I)** Relative mRNA expression in iWAT. **(J–P)** Relative mRNA expression in the BAT. Chow: The mice fed with chow diet (n=12), HFD: the mice fed with high fat diet (n=12), HFD+policosanol: The mice fed with high fat diet containing 0.5% policosanol (n=11). Differences of data in mice subjects were assessed by one-way ANOVA. For all pictures, columns indicated with different letters have significant difference, $p < 0.05$. ns, not significant.

browned, hydrolyzed, oxidized and thermogenerated (41). Therefore, we examined the gene expression associated with lipogenesis and lipolysis in adipose tissue. In obese mice, policosanol did not significantly change the expression of FAS and HSL genes but remarkably increased ATGL gene expression in iWAT. In addition to lipolysis related genes, the genes involved in thermogenesis and adipose "browning" process such as PPARs, UCP-1 and PRDM also play important roles in anti-obesity (14, 27). PPARs, including PPAR α and PPAR γ, regulate the metabolism of liver and adipose tissue, respectively. PPAR α in the liver has the effect of decreasing the level of SREBP-1c, promoting fat thermogenesis and alleviating nonalcoholic fatty liver (42, 43). In addition, we found that PPAR γ regulates fat metabolism in adipose tissue, which was also elevated in scWAT. PPARs are key proteins in thermogenesis. PRDM16 and PGC-1α can combine with PPARs to drive the differentiation of brown adipocytes and promote the expression of UCP-1 (12, 43). In this study, in the HFD+policosanol group, we found that PGC-1α and UCP-1 were increased in iWAT and BAT while PRDM16 was only increased in BAT. These results suggest that in obese mice, policosanol mainly plays a role by increasing the decomposition, oxidation and thermogenesis of lipids.

Based on the findings of this research and predecessors, there are still some limitations to be further studied in the regulation of host metabolism by policosanol through intestinal microbiota. How policosanol regulates intestinal microbiota and how it is metabolized and utilized by intestinal microbiota still remains unclear. In addition, whether the changes of small molecule metabolites in host intestine and serum after treatment with policosanol, such as the increase of polyunsaturated fatty acids and the decrease of TMAO contents, are caused by the changes of intestinal microbiota, which may need to be verified by sterile mouse model. In the future, in-depth research on these aspects may provide a new perspective for the promotion of whole grain food or policosanol as a functional food additive.

CONCLUSION

In summary, our findings elucidated that policosanol reduces fat accumulation in obese mice. Furthermore, we found that in obese mice, the metabolites involved in TCA cycle and thermogenesis were highly correlated with *Firmicutes, Bacteroidetes*, and *Candidatus:Sarccharinonasm*, *and Parasutterella*. These findings suggest that policosanol has the potential to be applied as an antihyperlipidemia supplement to improve well-being and health in humans.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repositories and accession(s) number can be found below: National Center for Biotechnology Information (NCBI) BioProject, https://www.ncbi.nlm.nih.gov/bioproject/, PRJNA729899 and Figshare, DOI: https://doi.org/10.6084/m9.figshare.14456595.v1.

ETHICS STATEMENT

The animal study was reviewed and approved by Committee of the Institute of Subtropical Agriculture at the Chinese Academy of Science (No. ISA-2020-18).

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

The project was funded by General Projects of Key Research and Development Plan in Jiangxi Province (20203BBFL63054), Project funded by China Postdoctoral Science Foundation (2020M682108), Guangdong Province enterprise special person special plan project (GDKTP2020054600), Pilot demonstration project for overall rationing system of Jiangxi Academy of Sciences (2021YSBG22008, 2021YSBG50009) and Research and Development Project of Jiangxi Academy of Sciences - Doctoral Fund Project (2020-YYB-01).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fendo.2021.722055/full#supplementary-material

REFERENCES

1. Jaacks LM, Vandevijvere S, Pan A, McGowan CJ, Wallace C, Imamura F, et al. The Obesity Transition: Stages of the Global Epidemic. *Lancet Diabetes Endocrinol* (2019) 7:231–40. doi: 10.1016/S2213-8587(19)30026-9
2. Cardel MI, Jastreboff AM, Kelly AS. Treatment of Adolescent Obesity in 2020. *JAMA* (2019) 322:1707–8. doi: 10.1001/jama.2019.14725
3. Zhang X, Zhang M, Zhao Z, Huang Z, Deng Q, Li Y, et al. Geographic Variation in Prevalence of Adult Obesity in China: Results From the 2013-2014 National Chronic Disease and Risk Factor Surveillance. *Ann Intern Med* (2020) 172:291. doi: 10.7326/M19-0477
4. Loffredo L, Martino F, Carnevale R, Pignatelli P, Catasca E, Perri L, et al. Obesity and Hypercholesterolemia Are Associated With NOX2 Generated Oxidative Stress and Arterial Dysfunction. *J Pediatr* (2012) 161:1004–9. doi: 10.1016/j.jpeds.2012.05.042
5. Marengo A, Rosso C, Bugianesi E. Liver Cancer: Connections With Obesity, Fatty Liver, and Cirrhosis. *Annu Rev Med* (2016) 67:103–17. doi: 10.1146/annurev-med-090514-013832
6. Komaroff AL. The Microbiome and Risk for Obesity and Diabetes. *Jama-J Am Med Assoc* (2017) 317:355–6. doi: 10.1001/jama.2016.20099
7. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An Obesity-Associated Gut Microbiome With Increased Capacity for Energy Harvest. *Nature* (2006) 444:1027–31. doi: 10.1038/nature05414

8. Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, Kau AL, et al. Gut Microbiota From Twins Discordant for Obesity Modulate Metabolism in Mice. *Science* (2013) 341:1049–79. doi: 10.1126/science.1241214
9. Liu R, Hong J, Xu X, Feng Q, Zhang D, Gu Y, et al. Gut Microbiome and Serum Metabolome Alterations in Obesity and After Weight-Loss Intervention. *Nat Med* (2017) 23:859–68. doi: 10.1038/nm.4358
10. Kincaid HJ, Nagpal R, Yadav H. Microbiome-Immune-Metabolic Axis in the Epidemic of Childhood Obesity: Evidence and Opportunities. *Obes Rev* (2020) 2:1–13. doi: 10.1111/obr.12963
11. Zheng X, Chen T, Jiang R, Zhao A, Wu Q, Kuang J, et al. Hyocholic Acid Species Improve Glucose Homeostasis Through a Distinct TGR5 and FXR Signaling Mechanism. *Cell Metab* (2020) 33:791. doi: 10.1016/j.cmet.2020.11.017
12. Wang D, Liu CD, Li HF, Tian ML, Pan JQ, Shu G, et al. LSD1 Mediates Microbial Metabolite Butyrate-Induced Thermogenesis in Brown and White Adipose Tissue. *Metabolism* (2020) 102:154011. doi: 10.1016/j.metabol.2019.154011
13. Chavez-Talavera O, Haas J, Grzych G, Tailleux A, Staels B. Bile Acid Alterations in Nonalcoholic Fatty Liver Disease, Obesity, Insulin Resistance and Type 2 Diabetes: What do the Human Studies Tell? *Curr Opin Lipidol* (2019) 30:244–54. doi: 10.1097/MOL.0000000000000597
14. Guo J, Han X, Tan H, Huang W, You Y, Zhan J. Blueberry Extract Improves Obesity Through Regulation of the Gut Microbiota and Bile Acids *via* Pathways Involving FXR and TGR5. *iScience* (2019) 19:676–90. doi: 10.1016/j.isci.2019.08.020
15. Shen J, Luo F, Lin Q. Policosanol: Extraction and Biological Functions. *J Funct Foods* (2019) 57:351–60. doi: 10.1016/j.jff.2019.04.024
16. Irmak S, Dunford NT, Milligan J. Policosanol Contents of Beeswax, Sugar Cane and Wheat Extracts. *Food Chem* (2006) 95:312–8. doi: 10.1016/j.foodchem.2005.01.009
17. Kabir Y, Kimura S. Tissue Distribution of (8-14C)-Octacosanol in Liver and Muscle of Rats After Serial Administration. *Ann Nutr Metab* (1995) 39:279–84. doi: 10.1159/000177873
18. Mazza A, Lenti S, Schiavon L, Zuin M, D'Avino M, Ramazzina E, et al. Nutraceuticals for Serum Lipid and Blood Pressure Control in Hypertensive and Hypercholesterolemic Subjects at Low Cardiovascular Risk. *Adv Ther* (2015) 32:680–90. doi: 10.1007/s12325-015-0229-x
19. Guardamagna O, Abello F, Baracco V, Stasiowska B, Martino F. The Treatment of Hypercholesterolemic Children: Efficacy and Safety of a Combination of Red Yeast Rice Extract and Policosanols. *Nutr Metab Cardiovasc Dis* (2011) 21:424–9. doi: 10.1016/j.numecd.2009.10.015
20. Nam DE, Yun JM, Kim D, Kim OK. Policosanol Attenuates Cholesterol Synthesis *via* AMPK Activation in Hypercholesterolemic Rats. *J Med Food* (2019) 22:1110–7. doi: 10.1089/jmf.2019.4491
21. Liu Y, Xie C, Zhai Z, Deng ZY, De Jonge HR, Wu X, et al. Uridine Attenuates Obesity, Ameliorates Hepatic Lipid Accumulation and Modifies the Gut Microbiota Composition in Mice Fed With a High-Fat Diet. *Food Funct* (2021) 12:1829–40. doi: 10.1039/d0fo02533j
22. Zhai Z, Zhang F, Cao R, Ni X, Xin Z, Deng J, et al. Cecropin a Alleviates Inflammation Through Modulating the Gut Microbiota of C57BL/6 Mice With DSS-Induced IBD. *Front Microbiol* (2019) 10:1595. doi: 10.3389/fmicb.2019.01595
23. Saccenti E, Hoefsloot HCJ, Smilde AK, Westerhuis JA, Hendriks MMWB. Reflections on Univariate and Multivariate Analysis of Metabolomics Data. *Metabolomics* (2014) 10:361–74. doi: 10.1007/s11306-013-0598-6
24. Xin Z, Zhai Z, Long H, Zhang F, Ni X, Deng J, et al. Metabolic Profiling by UPLC-Orbitrap-MS/MS of Liver From C57BL/6 Mice With DSS-Induced Inflammatory Bowel Disease. *Mediators Inflamm* (2020) 2020:6020247. doi: 10.1155/2020/6020247
25. Tan C, Zhai Z, Ni X, Wang H, Ji Y, Tang T, et al. Metabolomic Profiles Reveal Potential Factors That Correlate With Lactation Performance in Sow Milk. *Sci Rep* (2018) 8:10712. doi: 10.1038/s41598-018-28793-0
26. Gonzalez-Muniesa P, Martinez-Gonzalez MA, Hu FB, Despres JP, Matsuzawa Y, Loos R, et al. Obesity. *Nat Rev Dis Primers* (2017) 3:17034. doi: 10.1038/nrdp.2017.34
27. Nguyen NT, Varela JE. Bariatric Surgery for Obesity and Metabolic Disorders: State of the Art. *Nat Rev Gastroenterol Hepatol* (2017) 14:160–9. doi: 10.1038/nrgastro.2016.170
28. Heymsfield SB, Wadden TA. Mechanisms, Pathophysiology, and Management of Obesity. *N Engl J Med* (2017) 376:254–66. doi: 10.1056/NEJMra1514009
29. Inge TH, Jenkins TM, Xanthakos SA, Dixon JB, Daniels SR, Zeller MH, et al. Long-Term Outcomes of Bariatric Surgery in Adolescents With Severe Obesity (FABS-5+): A Prospective Follow-Up Analysis. *Lancet Diabetes Endocrinol* (2017) 5:165–73. doi: 10.1016/S2213-8587(16)30315-1
30. Cho K, Kim S. Consumption of Cuban Policosanol Improves Blood Pressure and Lipid Profile *via* Enhancement of HDL Functionality in Healthy Women Subjects: Randomized, Double-Blinded, and Placebo-Controlled Study. *J Am Coll Cardiol* (2018) 72S:C84. doi: 10.1016/j.jacc.2018.08.458
31. Sharma R, Matsuzaka T, Kaushik MK, Sugasawa T, Ohno H, Wang Y, et al. Octacosanol and Policosanol Prevent High-Fat Diet-Induced Obesity and Metabolic Disorders by Activating Brown Adipose Tissue and Improving Liver Metabolism. *Sci Rep-Uk* (2019) 9:5169. doi: 10.1038/s41598-019-41631-1
32. Gamez R, Aleman CL, Mas R, Noa M, Rodeiro I, Garcia H, et al. A 6-Month Study on the Toxicity of High Doses of Policosanol Orally Administered to Sprague-Dawley Rats. *J Med Food* (2001) 4:57–65. doi: 10.1089/109662001300341707
33. Aleman CL, Mas R, Hernandez C, Rodeiro I, Cerejido E, Noa M, et al. A 12-Month Study of Policosanol Oral Toxicity in Sprague Dawley Rats. *Toxicol Lett* (1994) 70:77–87. doi: 10.1016/0378-4274(94)90147-3
34. Othman RA, Xu Z, Fitz E, Riediger N, Moghadasian MH. The Impact of Dietary Octacosanol on Plasma Lipids and Atherosclerotic Lesion Development in Apo E-KO Mice. *FASEB J* (2007) 21:A1087.
35. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, et al. Richness of Human Gut Microbiome Correlates With Metabolic Markers. *Nature* (2013) 500:541–6. doi: 10.1038/nature12506
36. Ju T, Kong JY, Stothard P, Willing BP. Defining the Role of Parasutterella, a Previously Uncharacterized Member of the Core Gut Microbiota. *Isme J* (2019) 13:1520–34. doi: 10.1038/s41396-019-0364-5
37. Zhang J, Ni Y, Qian L, Fang Q, Zheng T, Zhang M, et al. Decreased Abundance of Akkermansia Muciniphila Leads to the Impairment of Insulin Secretion and Glucose Homeostasis in Lean Type 2 Diabetes. *Adv Sci* (2021). doi: 10.1002/advs.202100536
38. Wang L, Tang L, Feng Y, Zhao S, Han M, Zhang C, et al. A Purified Membrane Protein From Akkermansia Muciniphila or the Pasteurised Bacterium Blunts Colitis Associated Tumourigenesis by Modulation of CD8 (+) T Cells in Mice. *Gut* (2020) 69:1988–97. doi: 10.1136/gutjnl-2019-320105
39. Plovier H, Everard A, Druart C, Depommier C, Van Hul M, Geurts L, et al. A Purified Membrane Protein From Akkermansia Muciniphila or the Pasteurized Bacterium Improves Metabolism in Obese and Diabetic Mice. *Nat Med* (2017) 23:107–13. doi: 10.1038/nm.4236
40. Hsiao WY, Jung SM, Tang Y, Haley JA, Li R, Li H, et al. The Lipid Handling Capacity of Subcutaneous Fat Is Programmed by Mtorc2 During Development. *Cell Rep* (2020) 33:108223. doi: 10.1016/j.celrep.2020.108223
41. Pollard AE, Martins L, Muckett PJ, Khadayate S, Bornot A, Clausen M, et al. AMPK Activation Protects Against Diet Induced Obesity Through Ucp1-Independent Thermogenesis in Subcutaneous White Adipose Tissue. *Nat Metab* (2019) 1:340–9. doi: 10.1038/s42255-019-0036-9
42. Sun B, Jia Y, Hong J, Sun Q, Gao S, Hu Y, et al. Sodium Butyrate Ameliorates High-Fat-Diet-Induced Non-Alcoholic Fatty Liver Disease Through Peroxisome Proliferator-Activated Receptor Alpha-Mediated Activation of Beta Oxidation and Suppression of Inflammation. *J Agric Food Chem* (2018) 66:7633–42. doi: 10.1021/acs.jafc.8b01189
43. Jordan S, Tung N, Casanova-Acebes M, Chang C, Cantoni C, Zhang D, et al. Dietary Intake Regulates the Circulating Inflammatory Monocyte Pool. *Eur J Immunol* (2019) 491:178. doi: 10.1016/j.cell.2019.07.050

Conflict of Interest: Author HL is employed by Era Biotechnology (Shenzhen) Co., Ltd, China.

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.

this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Endocrinology

TYPE Systematic Review
PUBLISHED 04 October 2022
DOI 10.3389/fendo.2022.934847

OPEN ACCESS

EDITED BY
Animesh Acharjee,
University of Birmingham,
United Kingdom

REVIEWED BY
Alina Bazarova,
Jülich Supercomputing Center,
(HZ), Germany
Tiago L. Duarte,
Universidade do Porto, Portugal

*CORRESPONDENCE
Yongsoo Choi
yongsoo.choi@kist.re.kr

This article was submitted to
Systems Endocrinology,
a section of the journal
Frontiers in Endocrinology

SPECIALTY SECTION
RECEIVED 03 May 2022
ACCEPTED 14 September 2022
PUBLISHED 04 October 2022

CITATION
Pyo JJ and Choi Y (2022) Key hepatic
signatures of human and mouse
nonalcoholic steatohepatitis:
A transcriptome–proteome
data meta-analysis.
Front. Endocrinol. 13:934847.
doi: 10.3389/fendo.2022.934847

Key hepatic signatures of human and mouse nonalcoholic steatohepatitis: A transcriptome–proteome data meta-analysis

Jeong Joo Pyo[1] and Yongsoo Choi[1,2]*

[1]Natural Product Research Center, Korea Institute of Science and Technology (KIST), Gangneung, South Korea, [2]Division of Bio-Medical Science & Technology, KIST School, Korea University of Science and Technology, Seoul, South Korea

Background: Despite the global prevalence of nonalcoholic fatty liver disease (NAFLD), its pathophysiology remains unclear. In this study, we established highly confident nonalcoholic steatohepatitis (NASH) gene signatures and evaluated the pathological mechanisms underlying NASH through a systematic meta-analysis of transcriptome and proteome datasets obtained from NASH patients and mouse models.

Methods: We analyzed NASH transcriptome datasets from 539 patients and 99 mice. A whole-liver tissue proteome dataset was used to confirm the protein level dysregulation of NASH signatures significant in both humans and mice.

Results: In total, 254 human and 1,917 mouse NASH gene signatures were established. Up-regulated genes of 254 human signatures were associated with inflammation, steatosis, apoptosis, and extracellular matrix organization, whereas down-regulated genes were associated with response to metal ions and lipid and amino acid metabolism. When different mouse models were compared against humans, models with high fat and high fructose diet most closely resembled the genetic features of human NAFLD. Cross-species analysis revealed 66 genes that were concordantly dysregulated between human and mouse NASH. Among these, 14 genes were further validated to be dysregulated at the protein level. The resulting 14 genes included some of the well-established NASH associated genes and a promising NASH drug target. Functional enrichment analysis revealed that dysregulation of amino acid metabolism was the most significant hepatic perturbation in both human and mouse NASH.

Conclusions: We established the most comprehensive hepatic gene signatures for NASH in humans and mice to date. To the best of our knowledge, this is the

first study to collectively analyze the common signatures between human and mouse NASH on a transcriptome–proteome scale.

KEYWORDS

nonalcoholic fatty liver disease, nonalcoholic steatohepatitis, transcriptomics, proteomics, cross-species analysis

Introduction

Nonalcoholic fatty liver disease (NAFLD), defined by the presence of abnormal liver fat (steatosis in ≥ 5% of hepatocytes) in the absence of secondary causes of fatty liver, is the most common form of chronic liver disease worldwide (1). NAFLD encompasses a spectrum of conditions, from the simple nonalcoholic fatty liver (NAFL) to the more severe nonalcoholic steatohepatitis (NASH), which is hallmarked by inflammation and hepatocyte ballooning (1). The global prevalence of NAFLD is rapidly increasing, concurrent to the global epidemics of obesity and type 2 diabetes mellitus (T2DM) (2). It is estimated that 25% and 5% of the general population have NAFLD and NASH, respectively (2). Due to the high propensity of NASH to further develop into cirrhosis and hepatocellular carcinoma (HCC), NASH is becoming the leading cause of liver transplantation, but approved therapies have not yet been developed (3).

The pathogenesis of NASH is a complex multi-etiological process associated with genetic, epigenetic, metabolic, and environmental factors (4). Two gene variants, patatin-like phospholipase domain-containing protein 3 (*PNPLA3*) and transmembrane 6 superfamily member 2 (*TM6SF2*), have been validated to be strongly associated with NASH (4). Obesity and insulin resistance are key pathogenic factors in NAFLD, and T2DM is a well-established risk factor for the rapid progression of NAFL to NASH, cirrhosis, or HCC (5). Recent studies revealed an inverse correlation between diet quality and NAFLD prevalence (6), and regular consumption of fructose promoted hepatic *de novo* lipogenesis in a double-blind, randomized clinical trial (7). Despite many efforts to understand the disease, the complete pathophysiology of NASH remains unclear. Current options for managing NAFLD include bariatric surgery or lifestyle modification, such as exercise or diet control (1). These methods have proven to be effective in resolving NASH or even mild fibrosis, but are only applicable to a limited number of patients (1). Thus, additional studies are urgently needed to understand the pathogenesis of NASH for successful identification of therapeutic NASH targets and the development of corresponding drugs.

Animal models, particularly those used to study human diseases, offer valuable opportunities to researchers to experimentally investigate the pathophysiology of diseases *via* genetic or dietary interventions. NASH is regarded as a metabolic syndrome, and various mouse diet models have been developed to mimic human NASH. Although mouse models cannot represent the full spectrum of human NASH due to the genetic differences between the species, consistent dysregulation patterns of orthologous genes or proteins across the species could offer valuable insights into the disease pathogenesis. High-throughput technologies, such as transcriptomics and proteomics, are invaluable tools for providing a holistic view of biological systems. Transcriptomics, with its high-resolution and genome-wide capacity, has enabled researchers to query global gene expression patterns and infer protein abundances from mRNA abundances. Liquid chromatography-tandem mass spectrometry-based proteomics can be used to directly measure the abundance of proteins and identify post-translational modifications. Combining these two approaches can better depict biological dysregulations within cells or tissues (8). Although several previous studies have individually reported transcriptomic or proteomic changes in NASH, none of them have collectively analyzed the common signatures between human and mouse NASH on a transcriptome–proteome scale.

Therefore, in this study, we established comprehensive gene signatures of human and mouse NASH and profiled significantly aberrant genes to better understand the pathological mechanisms underlying NASH. To this end, we performed a systematic meta-analysis of publicly available transcriptome and proteome datasets from liver tissues of patients with NASH and mouse models.

Methods

This meta-analysis was performed according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (9). A detailed checklist is provided in the Supplementary Materials.

Data sources and search strategies

A systematic dataset search was conducted up to March 2021. To obtain transcriptome study datasets, publicly available

databases ArrayExpress (https://www.ebi.ac.uk/arrayexpress) and Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo) were utilized. The following keywords were used to search human NASH transcriptome datasets: "Homo sapiens," "RNA assay," "Expression profiling," "Nonalcoholic fatty liver disease," "Nonalcoholic steatohepatitis," "NAFLD," and "NASH." After removing duplicates, 9 datasets from ArrayExpress and 78 datasets from GEO were identified. Two additional datasets were identified through a reference search. When searching mouse NASH model datasets, the following keywords were used: "Mus musculus," "RNA assay," "Expression profiling," "Nonalcoholic fatty liver disease," "Nonalcoholic steatohepatitis," "NAFLD," and "NASH." After removing duplicates, 7 datasets from ArrayExpress and 187 datasets from GEO were identified.

Proteome study datasets were searched using the publicly available ProteomeXchange database (http://www.proteomexchange.org). The keywords "NAFLD" or "NASH" were first used to search the datasets, and the results were subsequently filtered according to species. After removing duplicates, 8 human datasets and 10 mouse model datasets were identified.

Dataset selection process and eligibility criteria

After systematic dataset identification, the datasets were further screened and excluded according to the following criteria. For human NASH studies, datasets were excluded when they corresponded to: (a) *in vitro*/cell line studies; (b) ambiguous disease classification; (c) studies involving other liver diseases (such as hepatitis, cirrhosis or HCC); (d) unavailability of raw data; (e) unclear sample annotation; and (f) total sample size ≤ 15. For the NASH mouse model studies, datasets were excluded when they corresponded to: (a) genetically intervened mouse studies; (b) *in vitro*/cell line studies; (c) NASH irrelevant model studies; (e) unavailability of raw data; (f) studies with mouse strains other than C57BL/6; and (g) single-cell transcriptome studies (Figure 1).

For human NASH proteome studies, datasets were excluded when they corresponded to: (a) *in vitro*/cell line studies; (b) studies with samples other than liver tissue (e.g., plasma); and (c) enriched peptide studies. NASH mouse model proteome study datasets were excluded when they corresponded to: (a) genetically intervened mouse studies; (b) *in vitro*/cell line studies; (c) studies with samples other than liver tissue; and (d) enriched peptide studies (Supplementary Figure 1).

Transcriptome dataset processing

Nine human (GSE33814 (10), GSE37031 (11), GSE48452 (12), GSE49541 (13), GSE61260 (14), GSE63067 (15), GSE66676 (16), GSE126848 (17), E-MEXP-3291 (18)) and 10 murine (GSE35961 (19), GSE43106 (20), GSE52748 (21), GSE93819 (22), GSE94593 (23), GSE119340 (24), GSE120977 (25), GSE137449 (26), GSE145665 (27), GSE148849 (28)) transcriptome datasets were subjected to further analysis. The raw data or processed data from each study were downloaded from the respective data repositories. The analysis platform for each dataset was heterogeneous and consisted of microarray and RNA sequencing data from multiple vendors. Due to the variability of platforms and discrepancies in sample handling or disease assessment of each study cohort, each dataset was analyzed separately. Affymetrix microarray raw data (.CEL files) were downloaded, and gene expression values were processed using Affymetrix Expression Console software (Affymetrix,

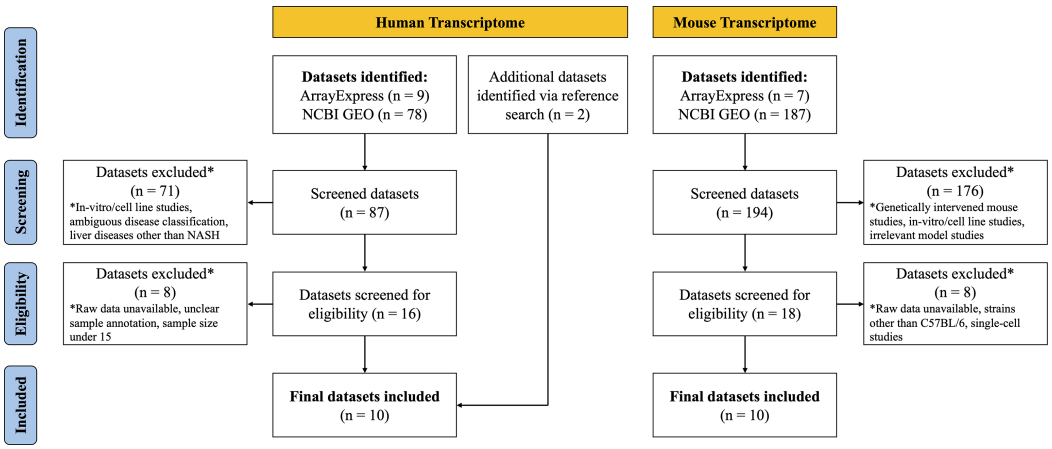


FIGURE 1
Flow diagram on the selection process of human and mouse NASH transcriptome datasets.

Santa Clara, CA, USA) with robust multiarray averaging (RMA) or microarray analysis suite 5.0 (MAS5) normalization. Processed Illumina BeadChip data were downloaded from NCBI GEO and probeset intensity normalization was performed using the ACTB and GAPDH genes as internal controls. In most cases, RNA sequencing read counts raw data were provided by NCBI GEO or the authors who published the studies. Normalization of the read counts and differentially expressed gene (DEG) selection were performed as downstream analyses if necessary. If raw counts were not available, RAW data were retrieved from Sequence Read Archive and further processed using the Galaxy platform. Detailed information on the acquisition, processing, normalization, and analysis method for each dataset is shown in Supplementary Table 1.

DEGs from each transcriptome dataset were individually selected as described by Pyo et al. (29). Briefly, after normalization of each gene expression value, the fold change and Benjamini–Hochberg adjusted *p*-value were calculated between the normal and NASH groups. Up-regulated DEGs were determined if the fold change (NASH/Normal) was greater than 1.7 and *p*-value was lower than 0.05. Down-regulated DEGs were determined if the fold change was lower than 0.7 and *p*-value was lower than 0.05. In the case of GSE49541, the normal group cohort was not available; hence, the fold change and *p*-value were determined by comparing the advanced NAFLD group with the mild NAFLD group. Although raw data were not available for the study published by Dali-Youcef et al. (30), a complete list of up-regulated and down-regulated DEGs was available in the manuscript. The DEGs were extracted and used for further analysis, affording a total of 10 human NASH transcriptome datasets that were used in this study.

Proteome dataset processing

Raw data (.RAW files) of the NASH mouse model liver proteome study PXD013423 (31) were downloaded from ProteomeXchange. Proteome Discoverer 2.3 (Thermo Fisher Scientific, Hanover Park, IL, USA) was used for the precursor quantification and label-free quantitative analyses. The UniProt mouse proteome reference (https://www.uniprot.org, March 2021 version) was used for MS2 peptide spectral matching. The assignment of MS2 spectra was carried out using the SEQUEST algorithm, and peptide hits were filtered at a maximum of 1% FDR using the Percolator algorithm. Carbamidomethylation of cysteine was set as static modification, whereas methionine oxidation and N-terminal acetylation were set as dynamic modifications. Full trypsin specificity with up to two missed cleavage sites was applied. Mass tolerance for precursor and fragment ions were set at 10 ppm and 0.02 Da, respectively.

Differentially expressed proteins (DEPs) were determined using the SEQUEST search parameters. Valid up-regulated DEPs were determined when SEQUEST HT score ≥ 30, protein coverage ≥ 20, and protein abundance of NASH/control ≥ 1.7. Down-regulated DEPs were determined when SEQUEST HT score ≥ 30, protein coverage ≥ 20, and protein abundance of NASH/control ≤ 0.7.

Establishing key common human and mouse hepatic NASH signatures

First, the selected up-regulated and down-regulated DEGs from the 10 human NASH datasets were merged based on the gene symbols. In this process, DEGs showing different dysregulation pattern between studies were excluded. Then, the frequency score (how many times a certain gene was listed as a DEG among the 10 datasets) of each DEG was calculated. Only the DEGs with frequency ≥ 3 constituted the human NASH signature (described in Section 2.6). Next, to compare the human NASH signature with that of mice, dysregulated DEGs from 10 NASH mouse model datasets were merged as described for the human datasets, and the frequency of each DEG was calculated. DEGs with a frequency ≥ 4 comprised the mouse NASH signature. Finally, NASH signatures of both humans and mice were integrated using BioMart orthologous gene annotation. Genes showing a consistent dysregulation pattern between the two species constituted a common hepatic NASH signature. To validate the NASH signatures at the protein level, the UniProt accession numbers of selected DEPs from PXD013423 were converted into gene symbols, and both profiles were combined. The final common NASH signature comprised the DEG frequency from human and mouse transcriptome datasets and fold change values from a mouse proteome dataset (Supplementary Data).

Degree of confidence of DEGs in multiple transcriptome dataset meta-analysis

The selection of DEGs from individual transcriptome datasets involved a fold change cutoff and statistical significance validation. Nevertheless, DEG lists from different cohorts of the same disease showed significant discrepancies. This may be due to heterogeneous patient groups (e.g., ethnicity), variability in disease diagnosis, assay platform differences, technical bias, instrumental error, or inherent limitations in statistical assumptions. Such problems can be partly overcome by identifying consistent gene dysregulation patterns across multiple independent cohort studies. Furthermore, when analyzing multiple datasets, the DEG

frequency score can be added as statistical dimension to evaluate the confidence of the DEGs.

The previous transcriptome datasets of humans and mice were used to evaluate the degree of confidence of the DEGs. However, rather than selecting DEGs based on their fold changes and *p*-values in individual datasets, genes were randomly selected (number of DEGs and randomly selected genes (RSGs) was matched for each dataset). Subsequently, RSGs were merged based on the gene symbols, and the frequency score of each gene was calculated. This process was bootstrapped 1,000 times, and the results were recorded each time. Finally, the median value of 1,000 times bootstrap resampling was calculated. (Supplementary Figure 2). The degree of confidence of DEGs per frequency score was calculated using the following formula:

$$Degree\ of\ Confidence = 100 - \frac{\text{Number of RSGs}/\text{Number of total RSGs}}{\text{Number of DEGs}/\text{Number of total DEGs}} \times 100$$

DEGs with a degree of confidence ≤ 50 were classified as "Not significant," confidence between 100 and 50 were classified as "Fair confident," and confidence of 100 were classified as "High confident" DEGs.

Gene Ontology biological process enrichment analyses

Gene Ontology (GO) biological process (BP) enrichment analysis was performed as described by Pyo et al. (28). Enriched GO BP for corresponding DEGs was determined by comparing the frequency of genes annotated by GO BP terms in a group of DEGs with those in the entire set of genes in the human reference list. The GO annotation files were downloaded from the Gene Ontology Consortium webpage (http://www.geneontology.org), and the March 2021 version of the GO BP terms was used for the analysis. A 2 × 2 contingency table was constructed to compare the frequency of DEGs annotated by the GO BP terms, with the number of genes annotated by these terms in the total 20,595 human gene reference. The 2 × 2 contingency table was analyzed for the calculation of *p*-values using the X^2 test (frequency ≥ 5) or Fisher's exact test (frequency< 5).

Software and statistical analyses

Statistical analyses were performed using the GraphPad Prism 8.0 software (GraphPad Software Inc., San Diego, CA, USA). Filtration, classification, and integration of transcriptome and proteome datasets were conducted using R software (version 4.0.3), specifically the packages included in "tidyverse". The transcriptome profile datasets were visualized in a heatmap using the heatmap.2 function in "gplots" package, and the Venn diagram was produced using the "VennDiagram" package. The heatmap and volcano plot for the proteome dataset were drawn using Proteome Discoverer 2.3.

Results

Description of included datasets

Details on the selection process of human and mouse NASH transcriptome datasets can be found in the flow diagram (Figure 1). A total of 10 human (539 patients) and 10 mouse model (99 mice) studies were included in this study. All the 10 human NASH datasets have been published and represent patients of multiple ethnicities. All included participants underwent liver biopsy, and NAFLD was diagnosed histologically (Table 1). The 10 NASH mouse model datasets were also previously published, used the wild-type C57BL/6 strain, and comprised multiple dietary intervention methods including a methionine- and choline-deficient diet with high-fat diet (MCDHFD), high-fat diet (HFD), NASH-inducing diet (ND), high-cholesterol high-cholate high-fat diet (CLD), high-fat high-sugar diet (HFHSD), fat fructose cholesterol diet (FFCD), choline-deficient L-amino acid-defined high-fat diet (CDAHFD), and fast food diet (FFD) with varying degrees of duration. Details of the diet composition and experimental design are summarized in Table 2.

Hepatic gene signature of human NASH

First, to establish the hepatic gene signature of human NASH, the selected up-regulated and down-regulated DEGs from 10 individual NASH studies were merged based on the gene symbols. Next, 535 DEGs showing heterogeneous dysregulation patterns between the studies were excluded. Consequently, a total of 7,070 (2,629 up-regulated and 4,441 down-regulated) DEGs was compiled. The frequency score for each DEG was calculated. One or two overlapping DEGs were regarded as statistically insignificant, and only DEGs with a frequency ≥ 3 were considered as valid, resulting in 254 human NASH gene signatures (Figure 2A). These genes were further divided into "Fair confident" and "High confident" gene groups according to the degree of confidence per each frequency score. (Supplementary Figure 2A). Accordingly, 28 genes were identified as dysregulated in human NASH with high confidence (Supplementary Table 2).

Functional enrichment analysis on 254 human NASH signatures

To gain a holistic view of the biological processes associated with the 254 human NASH signatures, GO BP enrichment

TABLE 1 Study characteristics of human NASH transcriptome datasets used in this study.

Study	Country	Patient group	Disease assessment	Platform	Data availability
Starmann et al., 2012 (10)	Germany	Normal 13 Steatosis 19 NASH 12	Biopsy	Illumina Human WG-6 v3.0 expression beadchip	GEO GSE33814
López-Vicario et al., 2014 (11)	Spain	Normal 7 NASH 8	Biopsy	Affymetrix Human Genome U133 Plus 2.0 Array	GEO GSE37031
Ahrens et al., 2013 (12)	Germany	Normal 14 Obese 27 Steatosis 14 NASH 18	Biopsy	Affymetrix Human Gene 1.1 ST Array	GEO GSE48452
Moylan et al., 2014 (13)	USA	Mild NAFLD 40 Advanced NAFLD 33	Biopsy	Affymetrix Human Genome U133 Plus 2.0 Array	GEO GSE49541
Horvath et al., 2014 (14)	USA	Normal 38 Obese 24 NAFLD 23 NASH 24	Biopsy	Affymetrix Human Gene 1.1 ST Array	GEO GSE61260
Frades et al., 2015 (15)	Sweden	Normal 7 Steatosis 2 NASH 9	Biopsy	Affymetrix Human Genome U133 Plus 2.0 Array	GEO GSE63067
Xanthakos et al., 2015 (16)	USA	Normal 34 Steatosis 26 Borderline NASH 5 NASH 2	Biopsy	Affymetrix Human Gene 1.0 ST Array	GEO GSE66676
Suppli et al., 2019 (17)	Denmark	Normal 14 Obese 12 NAFLD 15 NASH 16	US, Biopsy	Illumina NextSeq 500	GEO GSE126848
Lake et al., 2015 (18)	USA	Normal 19 Steatosis 10 NASH 16	Biopsy	Affymetrix Human Gene 1.0 ST Array	ArrayExpress E-MEXP-3291
Dali-Youcef et al., 2019 (30)	France	Normal 10 Obese 10 Steatosis 10 NASH 8	Biopsy	Agilent Human GE 8x60K	N/A

NASH, nonalcoholic steatohepatitis; NAFLD, nonalcoholic fatty liver disease; US, ultrasonography; N/A, not available.

analysis was performed. As a result, 125 up-regulated genes were significantly associated with inflammation (GO:0007155; 0050900; 0006954; 0034097), steatosis (GO:0071396), apoptosis (GO:1903034), and extracellular matrix (ECM) organization (GO:0030198) (Figure 2B and Supplementary Table 3). Not surprisingly, hepatic steatosis, inflammation, and apoptosis are the three main histological hallmarks of NASH, and the upregulation of ECM synthesis is a sign of NASH with fibrosis. Thus, we concluded that the up-regulated signature of human NASH fairly represents the overall histological features of NASH. In the analysis of the 129 down-regulated signatures, biological processes related to cellular response to metal ions (GO:0071280; 0071276; 0071294), lipid metabolism (GO:0008202; 0006629), and amino acid metabolism (GO:0009063; 1901605) were significantly enriched (Figure 2B and Supplementary Table 4).

Comparative analysis of gene expression profiles of different NASH mouse models

Next, to establish the hepatic gene signature of mouse NASH, up-regulated and down-regulated DEGs from 10 mouse model studies were merged, and the frequency of each DEG was calculated. For the mouse datasets, DEGs with a frequency ≥ 4 were statistically significant, resulting in 1,917 mouse NASH gene signatures (Supplementary Figure 3). These genes were divided into "Fair confident" and "High confident" genes as well (Supplementary Figure 2B). Since NASH model studies consisted of multiple dietary intervention methods of varying durations, we first compared the gene expression profiles of different dietary model studies. As expected, we observed considerable disparities between different NASH models (Figure 3). Two HFD models from independent studies

TABLE 2 Study characteristics of mouse NASH model transcriptome datasets used in this study.

Study	Diet model	Diet composition	Group	Platform	Data availability
Kita et al., 2012 (19)	MCDHFD, 8 weeks	Methionine- and choline-deficient diet with 60% fat	Normal 4 NASH 4	Affymetrix Mouse Genome 430 2.0 Array	GEO GSE35961
Kahle et al., 2013 (20)	HFD, 3 weeks	15% calories from casein, 27% calories from starch, maltose dextrin, cellulose 58% calories from soybean oil, safflower oil	Normal 8 NASH 7	Affymetrix Mouse Gene 1.0 ST Array	GEO GSE43106
Dorn et al., 2014 (21)	ND, 12 weeks	15% pork lard, 15% beef tallow, 4% palmitic acid, 4% stearic acid, 0.2% cholesterol, 30% sucrose	Normal 4 NASH 4	Affymetrix Mouse Gene 1.1 ST Array	GEO GSE52748
Kobori et al., 2017 (22)	CLD, 12 weeks	60% calories from fat, 1.25% cholesterol, 0.5% sodium cholate	Normal 5 NASH 5	Affymetrix Mouse Genome 430 2.0 Array	GEO GSE93819
Maradana et al., 2018 (23)	HFHSD, 14 weeks	N/A	Normal 3 NASH 3	Illumina HiSeq 4000	GEO GSE94593
Xiong et al., 2019 (24)	FFCD, 25 weeks	40% fat, 22% fructose, 2% cholesterol	Normal 3 NASH 3	Illumina HiSeq 2500	GEO GSE119340
Min-DeBartolo et al., 2019 (25)	CDAHFD, 12 weeks	L-amino acid diet with 45 kcal%fat with 0.1% methionine and no added choline	Normal 5 NASH 5	Illumina HiSeq 4000	GEO GSE120977
Heintz et al., 2020 (26)	CDAHFD, 8 weeks	18% protein, 62% fat, 20% carbohydrates, 0.1% methionine	Normal 4 NASH 4	Illumina NovaSeq 6000	GEO GSE137449
Lu et al., 2020 (27)	HFD, 24 weeks	60% fat, 20% carbohydrate, 20% protein	Normal 5 NASH 5	Illumina HiSeq 4000	GEO GSE145665
Bates et al., 2020 (28)	FFD, 21 weeks	17% kcal protein, 40% kcal fat, 43% kcal carbohydrate	Normal 8 NASH 10	Illumina HiSeq 2500	GEO GSE148849

NASH, non-alcoholic steatohepatitis; MCDHFD, Methionine- and choline-deficient diet with high-fat diet; HFD, high-fat diet; ND, NASH-inducing diet; CLD, high-cholesterol high-cholate high-fat diet; HFHSD, high-fat high-sugar diet; FFCD, fat fructose cholesterol diet; CDAHFD, choline-deficient L-amino acid defined high fat diet; FFD, fast food diet; N/A, not available.

(GSE43106 and GSE145665) showed noticeable differences in the DEG patterns, which may be attributed to the difference in dietary intervention duration (3 weeks vs. 24 weeks) or slightly different diet compositions (Table 2). However, two CDAHFD model studies (GSE120977 and GSE137449) showed relatively similar DEG patterns and were clustered together on the heatmap, indicating that upon comparable dietary intervention methods, correlative gene dysregulation patterns can be observed.

Comparative analysis of hepatic gene signatures of human and mouse NASH

We coalesced the NASH signatures of humans and mice using BioMart orthologous gene annotation. Combining the NASH signatures of humans and mice yielded a pattern that is similar to that observed by Teufel et al. (32). Profound differences were observed between the two signatures. Amongst the 254 human and 1,917 mouse NASH signatures, 66 genes were concordantly dysregulated between the two species (Figure 4). While comparing the NASH signatures, we prioritized genes that were significantly dysregulated in humans to derive clinically relevant results. Among the 28 "High confident" human NASH signatures, 8 genes were shown to be co-dysregulated in mice, whereas 20 genes did not match or showed no statistical significance in the mouse signature. In detail, CYP2C19 was excluded from the combined signature because humans and mice have different isoform variations. In mice, 10 genes (CDH23, CMYA5, EFHD1, ENO3, GPR88, PI6K3, P4HA1, PDE11A, STMN2, and VIL1) showed no evidence of dysregulation, and 7 genes (ABCB11, ACSL4, APOF, FAT1, LEPR, SOCS2, and TMEM154) showed some evidence of dysregulation, but were not statistically significant. Interestingly, ME1 and TSPAN13 showed distinctively opposite dysregulation patterns in the two species. (Supplementary Figure 4).

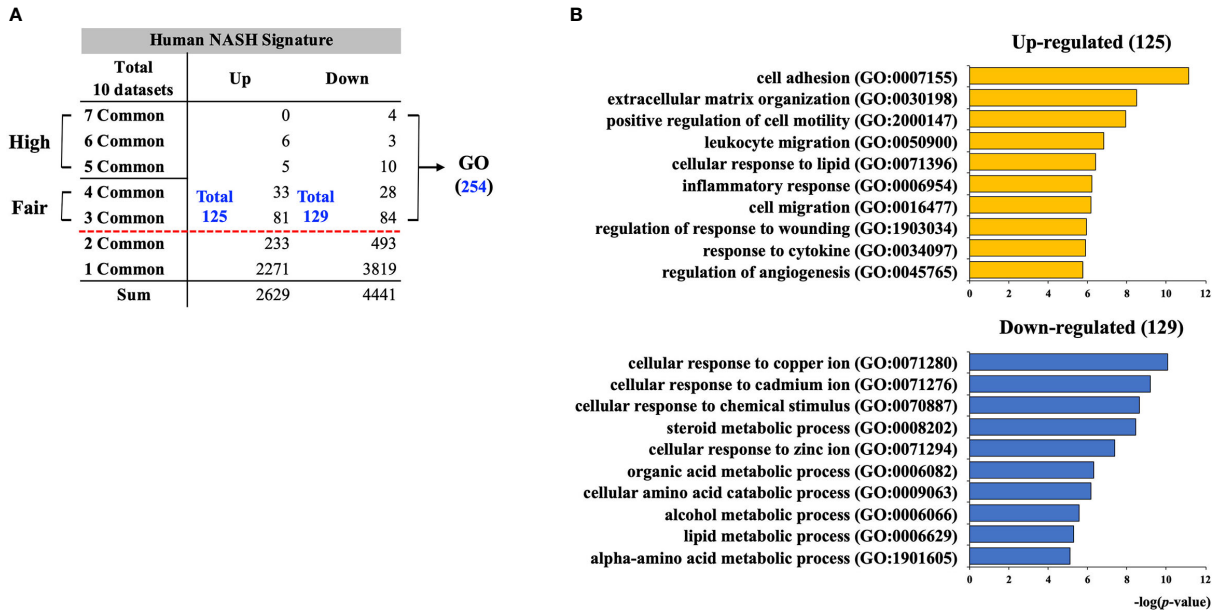


FIGURE 2
Establishment of human NASH signature and functional enrichment analysis. **(A)** Individually selected DEGs from 10 human NASH datasets were merged based on the gene symbols and the frequency score of each DEG was calculated. Total 254 genes (125 up-regulated and 129 down-regulated) were consistently dysregulated in at least three datasets. DEGs with frequency scores of 3 and 4 were classified as "Fair confident", and frequency scores of 5 to 7 were classified as "High confident" genes. **(B)** GO BP enrichment analysis result for 254 human NASH signatures. NASH, nonalcoholic steatohepatitis; GO BP, gene ontology biological process.

Key common NASH signatures and protein level validation

A list of 66 genes co-dysregulated in human and mouse NASH is provided in the Supplementary Data. Notably, the gene corresponding to glycine-N-methyltransferase (GNMT) was classified as a "High confident" gene in both the human and mouse signatures. Subsequently, to validate the 66 NASH signatures at the protein expression level, we performed a meta-analysis of proteome studies of human and mouse NASH. A systematic dataset search revealed one dataset, PXD013423, which analyzed whole-liver proteome expressions from a mouse fed with fructose palmitate cholesterol (FPC) diet (31). Using the SEQUEST algorithm for MS2 level peptide spectral matching, we identified 4,679 proteins with FDR ≤ 1%. A total of 617 proteins (232 up-regulated and 385 down-regulated) was selected as the NASH DEPs (Supplementary Figure 5). Finally, when 66 genes were validated using the DEPs, 14 genes were confirmed to be dysregulated at the protein level (Table 3). At the biological process level, these 14 validated NASH signatures were strongly associated with amino acid metabolism (Table 4). Particularly, six enzymes involved in amino acid metabolism were significantly down-regulated in both human and mouse NASH (Figure 5). This dysregulation was more prominent in the mouse datasets, and NAFLD progression-dependent down-regulation was observed in human datasets.

Discussion

Principal findings and corresponding interpretations

In this study, we performed a systematic meta-analysis of transcriptome studies of liver tissues from NASH patients and mouse models. As a result, we constructed 254 human and 1,917 mouse NASH gene signatures. According to the GO BP enrichment analysis, the up-regulated genes of the 254 human NASH signatures were associated with inflammation, steatosis, apoptosis, and ECM organization, which are the most significant hepatic perturbations in the pathogenesis of NASH. Down-regulated NASH signatures were associated with cellular response to metal ions, lipid metabolism, and amino acid metabolism. Interestingly, cellular responses to copper, cadmium, and zinc ions were found to be the most significantly enriched biological processes among the down-regulated gene signatures. This result was due to the consistent down-regulation of metallothionein genes– MT1E, MT1F, MT1M, MT1X, and MT2A. Metallothionein is a family of cysteine-rich, low molecular weight proteins with metal-binding capacity, thus protecting the cells from metal toxicity and oxidative stress (33). Its expression is believed to be dependent on the cellular environment of stress and mineral availability. Accordingly, accumulated evidence suggests a zinc

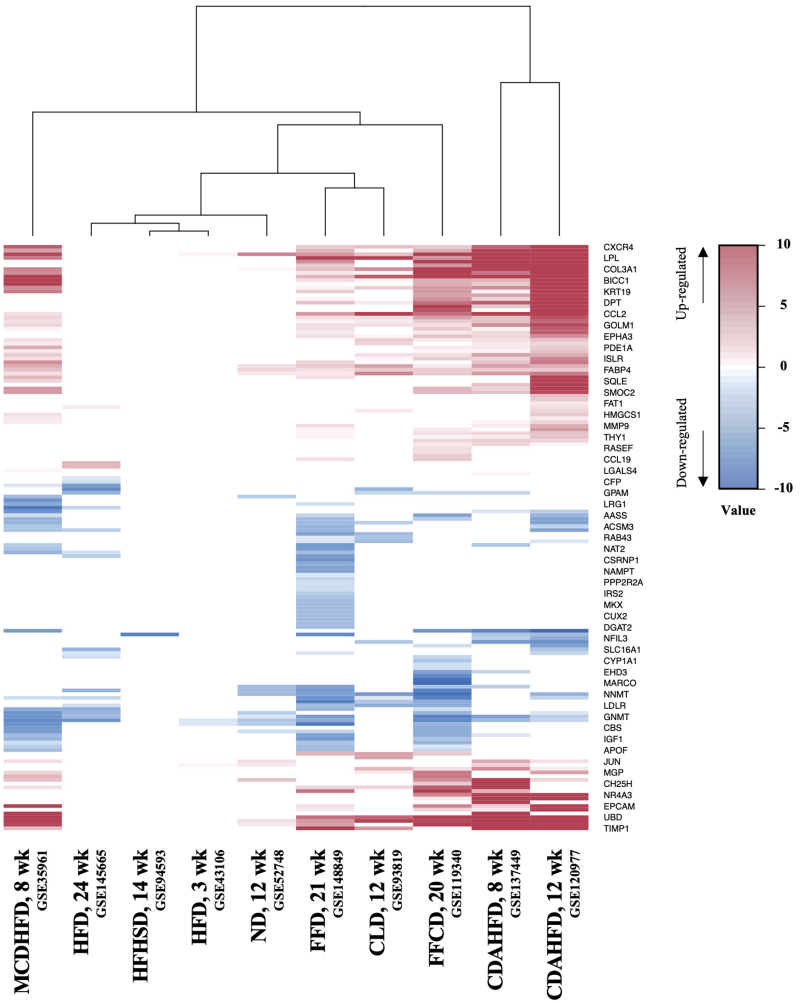


FIGURE 3
Heatmap of gene expression profiles of different NASH mouse models. DEGs from 10 mouse NASH datasets were merged based on the gene symbols, and the fold change values of dysregulated genes were normalized. Gradient of red color represents high expression of genes, whereas gradient of blue color represents low expression of genes in mouse NASH group. MCDHFD, methionine- and choline-deficient diet with high-fat diet; HFD, high-fat diet; HFHSD, high-fat high-sugar diet; ND, NASH-inducing diet; FFD, fast food diet; CLD, high-cholesterol high-cholate high-fat diet; FFCD, fat fructose cholesterol diet; CDAHFD, choline-deficient L-amino acid defined high fat diet.

and copper deficiency in the serum and/or hepatic tissue of NAFLD patients (34, 35). Mechanistic studies using rats also showed the causative role of zinc and copper deficiency in the pathogenesis of NAFLD (35, 36). However, despite these reports, metallothionein gene (mouse isoforms Mt1 and Mt2) dysregulation was not statistically significant in our mouse NASH signature. This may be explained by the fact that almost all experimental NASH models supply the same amounts of micronutrients (e.g., minerals) to both control and NASH-inducing groups. Although the intake of individual dietary minerals varies in clinical circumstances, mineral intake in NASH mouse models is tightly controlled. To confirm the effect of mineral deficiency and metallothionein dysregulation in the pathogenesis of NASH, dietary models with varying amounts of minerals should be used (36). Although mineral deficiencies may not be the primary cause of NAFLD, our results and those of others strongly suggest that these deficiencies are involved in the disease pathogenesis.

Mouse models are integral to the studies of NASH pathogenesis, and many different dietary methods have been developed. However, no single model has been established to represent the full spectrum of human NASH, and each model reflects different aspects of the disease. Since the transcriptome profile is believed to represent the overall biological status of cells or tissues, we hypothesized that a mouse model exhibiting the most similar gene expression profile to the 254 human NASH signatures would best represent the hepatic conditions of human NASH. Upon comparison, models with FFCD and FFD showed

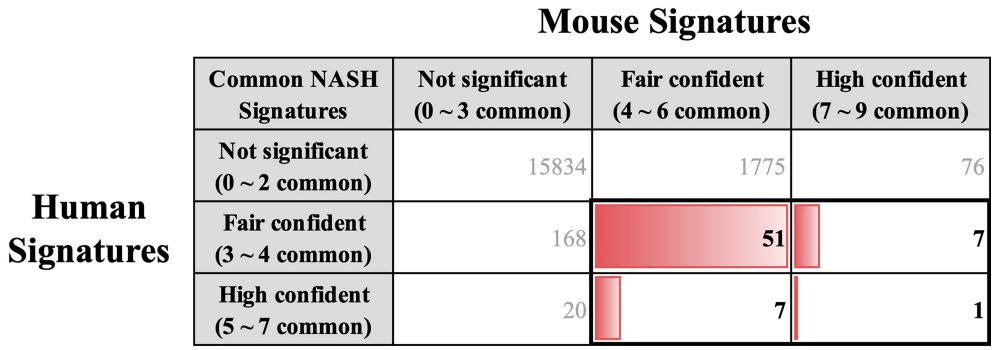


FIGURE 4
Key common hepatic signatures of human and mouse NASH. NASH signatures of humans and mice were coalesced using the BioMart orthologous gene annotation. Amongst the 254 human and 1,917 mouse NASH signatures, 66 genes (boxed with bold lines) were concordantly dysregulated between the two species.

the strongest resemblance to the human NASH signature, each reflecting 90 of the 254 genes. Im et al. recently carried out a systematic literature review of 3,920 NAFLD mouse models and concluded that dietary models with high fat and high fructose most closely resembled the metabolic and histological features of human NAFLD (37). This is in line with our results, considering that FFCD and FFD both have large amounts of fat and fructose/sucrose as the main diet composition. These two models were followed by the models, CDAHFD-12 week (85 genes), CDAHFD-8 week (73 genes), and MCDHFD (73 genes), in terms of human NASH resemblance. Choline-deficient diet models showed the highest mean liver histology scores in a systematic review study, but had relatively poor metabolic features compared to high-fat high-fructose diet models (37). Taken together, these results indicate that our transcriptome signature is in good correlation with the metabolic and histological features of NASH. On the contrary, high-fat diets with no added fructose exhibited the least gene expression features of human NASH in our study, which also highlights the importance of fructose in the pathogenesis of NASH.

Transcriptomics is a powerful tool that can measure genome-scale mRNA expression levels with high accuracy, and thus infer protein abundances within a biological system. However, microarray or RNA-seq experimental results still

TABLE 3 List of 14 common NASH signatures validated at the protein expression level.

Gene symbol	Gene title	Dysregulation	Human frequency score (# out of 10)	Mouse frequency score (# out of 10)	Mouse protein fold change
AASS	Alpha-aminoadipic semialdehyde synthase	Down	4	4	0.41
AMDHD1	Amidohydrolase Domain Containing 1	Down	4	5	0.56
ANXA2	Annexin A2	Up	3	8	3.55
CYP1A2	Cytochrome P450 Family 1 Subfamily A Member 2	Down	4	5	0.39
FABP4	Fatty Acid Binding Protein 4	Up	4	7	1.83
GCAT	Glycine C-Acetyltransferase	Down	3	4	0.35
GNMT	Glycine N-Methyltransferase	Down	6	7	0.12
GSN	Gelsolin	Up	3	6	1.76
HAL	Histidine Ammonia-Lyase	Down	3	8	0.24
KRT19	Keratin 19	Up	3	5	1.82
LGALS3	Galectin 3	Up	3	7	7.17
LUM	Lumican	Up	4	6	2.39
OAT	Ornithine Aminotransferase	Down	4	6	0.36
SDS	Serine Dehydratase	Down	4	4	0.34

#, number.

TABLE 4 GO BP enrichment analysis on 14 validated NASH signatures.

GO BP ID	Enriched GO BP terms in DEGs	*p*-value	# of DEGs in GO BP	# of genes in reference	DEGs
GO:0019557	histidine catabolic process to glutamate and formate	6.42E-06	2	4	AMDHD1, HAL
GO:0006567	threonine catabolic process	8.98E-06	2	5	GCAT, SDS
GO:1901607	alpha-amino acid biosynthetic process	1.22E-05	3	64	AASS, OAT, SDS

GO BP, gene ontology biological process; NASH, nonalcoholic steatohepatitis; DEGs, differentially expressed genes; #, number.

require follow-up validation. Hence, we performed a meta-analysis of proteomic studies of human and mouse NASH, and derived liver tissue DEPs from the FPC NASH mouse model. When 66 common NASH signatures were validated using the DEPs, 14 genes were confirmed as dysregulated at the protein level. Among the 14 genes, those such as FABP4, GNMT, and LUM are well-known to be associated with NASH (38–40), and galectin-3 encoded by LGALS3 is currently being tested as a therapeutic target in a phase 2b/3 clinical trial (41). Inclusion of these prominent drug targets motivated us to investigate other relatively less-studied genes.

Notably, among the 14 genes, down-regulation of AASS, AMDHD1, GCAT, HAL, OAT, and SDS was associated with amino acid metabolism in the GO BP enrichment analysis. In particular, histidine and threonine catabolic processes and α-amino acid anabolic processes were prominently enriched. Several previous studies have corroborated our findings. Lake et al. performed metabolome analyses on liver tissues from NAFLD patients and reported an increased level of lysine in NASH patients (18). Alpha-aminoadipic semialdehyde synthase encoded by AASS is involved in the first two steps of lysine degradation *via* the saccharopine pathway within the mitochondria, and down-regulation of AASS may lead to elevated levels of lysine in the liver. Eriksen et al. assessed the expression of urea cycle-related genes in NAFLD patients and reported the down-regulation of hepatic genes governing ureagenesis as well as impaired amino acid metabolism (42). Down-regulation of hepatic α-amino acid metabolism was

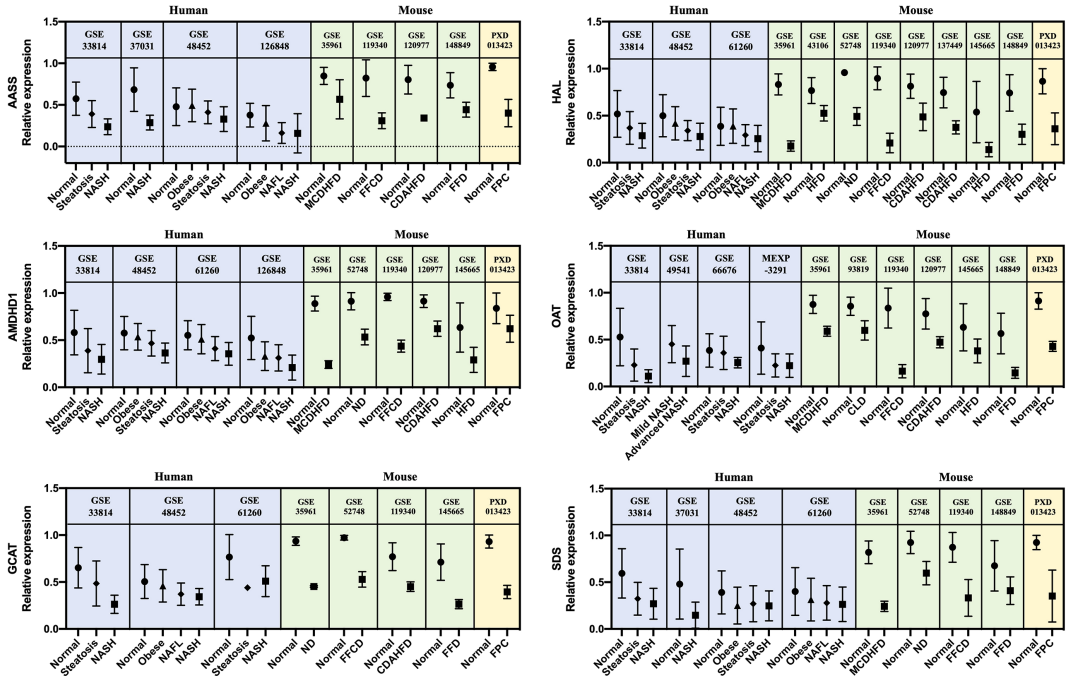


FIGURE 5
Relative gene and protein expression values of down-regulated amino acid metabolism enzymes across human and mouse datasets. Blue boxes indicate data from human transcriptome studies, green boxes indicate data from mouse transcriptome studies, and yellow boxes indicate data from mouse proteome studies. Graphs show the mean values with SD. For every dataset, *p*-value< 0.05 for normal versus NASH group. AASS, alpha-aminoadipic semialdehyde synthase; AMDHD1, amidohydrolase domain containing 1; GCAT, glycine C-acetyltransferase; HAL, histidine ammonia-lyase; OAT, ornithine aminotransferase; SDS, serine dehydratase.

manifested by higher blood concentration of α-amino nitrogen levels in NAFLD patients (42). In our study, down-regulation of AMDHD1 and HAL, which are involved in the catabolic pathway of histidine to glutamate conversion, and down-regulation of AASS, which also produces glutamate along the saccharopine pathway, may be presumed to result in glutamate level alteration in NASH. Accordingly, a metabonomic study by García-Cañaveras et al. reported decreased levels of glutamate in human NAFLD liver tissues (43). Glutamate plays a critical role in hepatic amino acid metabolism, acting as a key intermediate between the urea cycle and citric acid cycle, and it is a major substrate for glutathione (GSH) synthesis (44). Decreased level of glutamate in the liver can compromise the replenishment of GSH and make the liver more susceptible to pathological environment. In short, our results of gene dysregulation analysis reflect the metabolic status of the NAFLD liver. Interestingly, the down-regulation of amino acid metabolism enzymes was associated with NAFLD progression in human datasets. To the best of our knowledge, this is the first study to report six significant enzymes responsible for the dysregulation of amino acid metabolism in the pathogenesis of NASH.

Effect of the number of included datasets in meta-analysis

In this study, we used 10 human transcriptome datasets to establish NASH signatures. To determine whether the number of included datasets affects the overall NASH signatures, we randomly selected five human datasets (GSE37031, GSE49541, GSE63067, GSE126848, MEXP-3291) and repeated the meta-analysis. This time, the human NASH signatures consisted of 328 DEGs (DEGs with frequency ≥ 2). When 328 NASH signatures were compared with the previous 254 NASH signatures, 171 DEGs were found to be common (Supplementary Figure 6). Interestingly, there were some discrepancies between the two signatures. To explain this discrepancy, we provide examples of two DEGs, ABCC4 and AMDHD1, which were unique in each NASH signature. When we analyzed the transcriptional profile for ABCC4 and AMDHD1 across the 10 human datasets, ABCC4 was upregulated in the GSE126848 and MEXP-3291 datasets, whereas AMDHD1 was downregulated in the GSE33814, GSE48452, GSE61260, and GSE126848 datasets. Therefore, ABCC4 could be selected as NASH signatures when only five datasets were used (DEGs with frequency ≥ 2), but lacked statistical significance when all 10 datasets were used (DEGs with frequency ≥ 3). In contrast, AMDHD1 was selected as a NASH signature when 10 datasets were used, but the frequency score decreased to 1 when five datasets were randomly selected (GSE33814, GSE48452, and GSE61260 were not selected in this case). This result strongly supports the importance of using a comprehensive dataset when performing a meta-analysis. In our study, at least six datasets were required to obtain a threshold DEG frequency score of 3, but this result may vary according to the characteristics of each included dataset and the number of DEGs selected.

Comparison with other studies

We identified two studies that performed a large-scale systematic meta-analysis of human NASH transcription profiles and which they derived NASH gene signatures. The study by Ryaboshapkina and Hammar performed meta-analysis of seven microarray datasets and extracted 218 gene signature that are affected during the NAFLD progression (45). Jia and Zhai integrated six microarray datasets and established 96 significant DEGs between healthy people and NAFLD patients, using the robust rank aggregation method (46). In our study, we performed a comprehensive meta-analysis of 10 microarray and RNA sequencing datasets, and derived 254 statistically confident human NASH signatures. While comparing our NASH signature with those of the other two studies, we found five genes–FABP4, GNMT, IL32, TP53I3, and VIL1–to be common (Supplementary Figure 7). However, the overall signatures of each study were relatively different, and 181 genes (37.1%) were unique to our study. The differences in inclusion criteria during the dataset selection process, number of included datasets, types of included assay platforms, and different statistical approaches may all contribute to the diversity of individual gene signatures.

Limitations

Our study had several inherent limitations. First, while analyzing human transcriptome studies, disease progression variables were not considered. Although all study participants were diagnosed histologically, inter-observer variation exists in NASH diagnosis, and each study used different patient classifications. To minimize the effect of disparity on disease assessment between study cohorts and to systematically compare gene expression profiles from multifarious datasets, DEGs were selected by comparing normal subjects and well-defined NASH patient groups. Second, participant baseline characteristics and clinical biochemical data analyses were not possible due to limitations in data acquisition. Third, while combining human and mouse NASH signatures, genes with no reported orthologs and isoforms of genes that are unique to either species were excluded during the process. Finally, protein-level validation of NASH signatures was only performed using the mouse proteome dataset. Datasets of the whole liver proteome from NASH patients were not available, which strongly suggests a need for further studies.

Conclusions

We performed a meta-analysis of NASH transcriptome datasets and established the most comprehensive 254 human and 1,917 mouse NASH signatures to date. Based on comparison of different dietary models with the human NASH signature, our results add to the existing body of literature suggesting that dietary models with high fat and high fructose most closely resemble the genetic, metabolic, and histological features of human NAFLD. Cross-species analysis revealed 66 genes to be concordantly dysregulated between human and mouse NASH. Among these, 14 genes were further validated to be dysregulated at the protein level. The resulting 14 genes included some well-established NASH-associated genes and a promising NASH drug target. Functional enrichment analysis demonstrated that dysregulation of amino acid metabolism was the most significant hepatic perturbation in both human and mouse NASH. Moreover, down-regulation of six amino acid metabolism enzymes, AASS, AMDHD1, GCAT, HAL, OAT, and SDS, was associated with NAFLD progression in humans. Further studies are needed to unravel the link between dysregulation of hepatic amino acid metabolism and NAFLD pathogenesis.

Data availability statement

All datasets analyzed in this meta-analysis are publicly available from respective data repositories. The final meta-analysis result of human and mouse transcriptome–proteome profile data is provided in the Supplementary Data.

Author contributions

All authors listed have made substantial, direct, and intellectual contribution to the work and approved it for publication. All authors of this paper have read and approved the final version.

Funding

This work was financially supported by the Korea Institute of Science and Technology under grant (2Z06661) and (2Z06666).

Acknowledgments

We appreciate the authors of the transcriptomic and proteomic studies for making their data publicly available and all the patients who contributed to the research. We would like to offer special thanks to Yura Song for providing technical advice on transcriptome analyses and Moonyoung Lee for providing advice on statistical analyses.

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/fendo.2022.934847/full#supplementary-material

References

1. Wong VW, Adams LA, de Ledinghen V, Wong GL, Sookoian S. Noninvasive biomarkers in nafld and Nash - current progress and future promise. *Nat Rev Gastroenterol Hepatol* (2018) 15(8):461–78. doi: 10.1038/s41575-018-0014-9

2. Younossi ZM. Non-alcoholic fatty liver disease - A global public health perspective. *J Hepatol* (2019) 70(3):531–44. doi: 10.1016/j.jhep.2018.10.033

3. Vuppalanchi R, Noureddin M, Alkhouri N, Sanyal AJ. Therapeutic pipeline in nonalcoholic steatohepatitis. *Nat Rev Gastroenterol Hepatol* (2021) 18(6):373–92. doi: 10.1038/s41575-020-00408-y

4. Younossi Z, Anstee QM, Marietti M, Hardy T, Henry L, Eslam M, et al. Global burden of nafld and Nash: Trends, predictions, risk factors and prevention. *Nat Rev Gastroenterol Hepatol* (2018) 15(1):11–20. doi: 10.1038/nrgastro.2017.109

5. Targher G, Corey KE, Byrne CD, Roden M. The complex link between nafld and type 2 diabetes mellitus - mechanisms and treatments. *Nat Rev Gastroenterol Hepatol* (2021) 18(9):599–612. doi: 10.1038/s41575-021-00448-y

6. Vilar-Gomez E, Nephew LD, Vuppalanchi R, Gawrieh S, Mladenovic A, Pike F, et al. High quality diet, physical activity and college education are associated with low risk of nafld among the U.S. population. *Hepatology* (2021) 75(6):1491–1506. doi: 10.1002/hep.32207

7. Geidl-Flueck B, Hochuli M, Nemeth A, Eberl A, Derron N, Kofeler HC, et al. Fructose- and sucrose- but not glucose-sweetened beverages promote hepatic *De novo* lipogenesis: A randomized controlled trial. *J Hepatol* (2021) 75(1):46–54. doi: 10.1016/j.jhep.2021.02.027

8. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. *Genome Biol* (2017) 18(1):83. doi: 10.1186/s13059-017-1215-1

9. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The prisma 2020 statement: An updated guideline for reporting systematic reviews. *Bmj* (2021) 372:n71. doi: 10.1136/bmj.n71

10. Starmann J, Falth M, Spindelbock W, Lanz KL, Lackner C, Zatloukal K, et al. Gene expression profiling unravels cancer-related hepatic molecular signatures in steatohepatitis but not in steatosis. *PloS One* (2012) 7(10):e46584. doi: 10.1371/journal.pone.0046584

11. Lopez-Vicario C, Gonzalez-Periz A, Rius B, Moran-Salvador E, Garcia-Alonso V, Lozano JJ, et al. Molecular interplay between Delta5/Delta6 desaturases and long-chain fatty acids in the pathogenesis of non-alcoholic steatohepatitis. *Gut* (2014) 63(2):344–55. doi: 10.1136/gutjnl-2012-303179

12. Ahrens M, Ammerpohl O, von Schonfels W, Kolarova J, Bens S, Itzel T, et al. DNA Methylation analysis in nonalcoholic fatty liver disease suggests distinct disease-specific and remodeling signatures after bariatric surgery. *Cell Metab* (2013) 18(2):296–302. doi: 10.1016/j.cmet.2013.07.004

13. Moylan CA, Pang H, Dellinger A, Suzuki A, Garrett ME, Guy CD, et al. Hepatic gene expression profiles differentiate presymptomatic patients with mild versus severe nonalcoholic fatty liver disease. *Hepatology* (2014) 59(2):471–82. doi: 10.1002/hep.26661

14. Horvath S, Erhart W, Brosch M, Ammerpohl O, von Schonfels W, Ahrens M, et al. Obesity accelerates epigenetic aging of human liver. *Proc Natl Acad Sci U.S.A.* (2014) 111(43):15538–43. doi: 10.1073/pnas.1412759111

15. Frades I, Andreasson E, Mato JM, Alexandersson E, Matthiesen R, Martinez-Chantar ML. Integrative genomic signatures of hepatocellular carcinoma derived from nonalcoholic fatty liver disease. *PloS One* (2015) 10(5): e0124544. doi: 10.1371/journal.pone.0124544

16. Xanthakos SA, Jenkins TM, Kleiner DE, Boyce TW, Mourya R, Karns R, et al. High prevalence of nonalcoholic fatty liver disease in adolescents undergoing bariatric surgery. *Gastroenterology* (2015) 149(3):623–34 e8. doi: 10.1053/j.gastro.2015.05.039

17. Suppli MP, Rigbolt KTG, Veidal SS, Heeboll S, Eriksen PL, Demant M, et al. Hepatic transcriptome signatures in patients with varying degrees of nonalcoholic fatty liver disease compared with healthy normal-weight individuals. *Am J Physiol Gastrointest Liver Physiol* (2019) 316(4):G462–G72. doi: 10.1152/ajpgi.00358.2018

18. Lake AD, Novak P, Shipkova P, Aranibar N, Robertson DG, Reily MD, et al. Branched chain amino acid metabolism profiles in progressive human nonalcoholic fatty liver disease. *Amino Acids* (2015) 47(3):603–15. doi: 10.1007/s00726-014-1894-9

19. Kita Y, Takamura T, Misu H, Ota T, Kurita S, Takeshita Y, et al. Metformin prevents and reverses inflammation in a non-diabetic mouse model of nonalcoholic steatohepatitis. *PloS One* (2012) 7(9):e43056. doi: 10.1371/journal.pone.0043056

20. Kahle M, Horsch M, Fridrich B, Seelig A, Schultheiss J, Leonhardt J, et al. Phenotypic comparison of common mouse strains developing high-fat diet-induced hepatosteatosis. *Mol Metab* (2013) 2(4):435–46. doi: 10.1016/j.molmet.2013.07.009

21. Dorn C, Engelmann JC, Saugspier M, Koch A, Hartmann A, Muller M, et al. Increased expression of c-jun in nonalcoholic fatty liver disease. *Lab Invest* (2014) 94(4):394–408. doi: 10.1038/labinvest.2014.3

22. Kobori M, Takahashi Y, Sakurai M, Ni Y, Chen G, Nagashimada M, et al. Hepatic transcriptome profiles of mice with diet-induced nonalcoholic steatohepatitis treated with astaxanthin and vitamin e. *Int J Mol Sci* (2017) 18 (3):593. doi: 10.3390/ijms18030593

23. Maradana MR, Yekollu SK, Zeng B, Ellis J, Clouston A, Miller G, et al. Immunomodulatory liposomes targeting liver macrophages arrest progression of nonalcoholic steatohepatitis. *Metabolism* (2018) 78:80–94. doi: 10.1016/j.metabol.2017.09.002

24. Xiong X, Wang Q, Wang S, Zhang J, Liu T, Guo L, et al. Mapping the molecular signatures of diet-induced Nash and its regulation by the hepatokine tsukushi. *Mol Metab* (2019) 20:128–37. doi: 10.1016/j.molmet.2018.12.004

25. Min-DeBartolo J, Schlerman F, Akare S, Wang J, McMahon J, Zhan Y, et al. Thrombospondin-I is a critical modulator in non-alcoholic steatohepatitis (Nash). *PloS One* (2019) 14(12):e0226854. doi: 10.1371/journal.pone.0226854

26. Heintz MM, McRee R, Kumar R, Baldwin WS. Gender differences in diet-induced steatotic disease in Cyp2b-null mice. *PloS One* (2020) 15(3):e0229896. doi: 10.1371/journal.pone.0229896

27. Lu Y, Shao M, Xiang H, Zheng P, Wu T, Ji G. Integrative transcriptomics and metabolomics explore the mechanism of kaempferol on improving nonalcoholic steatohepatitis. *Food Funct* (2020) 11(11):10058–69. doi: 10.1039/d0fo02123g

28. Bates J, Vijayakumar A, Ghoshal S, Marchand B, Yi S, Kornyeyev D, et al. Acetyl-coa carboxylase inhibition disrupts metabolic reprogramming during hepatic stellate cell activation. *J Hepatol* (2020) 73(4):896–905. doi: 10.1016/j.jhep.2020.04.037

29. Pyo JJ, Ahn S, Jin SH, An S, Lee E, Choi J, et al. Keratinocyte-derived il-36gamma plays a role in hydroquinone-induced chemical leukoderma through inhibition of melanogenesis in human epidermal melanocytes. *Arch Toxicol* (2019) 93(8):2307–20. doi: 10.1007/s00204-019-02506-6

30. Dali-Youcef N, Vix M, Costantino F, El-Saghire H, Lhermitte B, Callari C, et al. Interleukin-32 contributes to human nonalcoholic fatty liver disease and insulin resistance. *Hepatol Commun* (2019) 3(9):1205–20. doi: 10.1002/hep4.1396

31. Gluchowski NL, Gabriel KR, Chitraju C, Bronson RT, Mejhert N, Boland S, et al. Hepatocyte deletion of triglyceride-synthesis enzyme acyl coa: Diacylglycerol acyltransferase 2 reduces steatosis without increasing inflammation or fibrosis in mice. *Hepatology* (2019) 70(6):1972–85. doi: 10.1002/hep.30765

32. Teufel A, Itzel T, Erhart W, Brosch M, Wang XY, Kim YO, et al. Comparison of gene expression patterns between mouse models of nonalcoholic fatty liver disease and liver tissues from patients. *Gastroenterology* (2016) 151 (3):513–25 e0. doi: 10.1053/j.gastro.2016.05.051

33. Dai H, Wang L, Li L, Huang Z, Ye L. Metallothionein 1: A new spotlight on inflammatory diseases. *Front Immunol* (2021) 12:739918. doi: 10.3389/fimmu.2021.739918

34. Berna G, Romero-Gomez M. The role of nutrition in non-alcoholic fatty liver disease: Pathophysiology and management. *Liver Int* (2020) 40 Suppl 1:102–8. doi: 10.1111/liv.14360

35. Pickett-Blakely O, Young K, Carr RM. Micronutrients in nonalcoholic fatty liver disease pathogenesis. *Cell Mol Gastroenterol Hepatol* (2018) 6(4):451–62. doi: 10.1016/j.jcmgh.2018.07.004

36. Aigner E, Strasser M, Haufe H, Sonnweber T, Hohla F, Stadlmayr A, et al. A role for low hepatic copper concentrations in nonalcoholic fatty liver disease. *Am J Gastroenterol* (2010) 105(9):1978–85. doi: 10.1038/ajg.2010.170

37. Im YR, Hunter H, de Gracia Hahn D, Duret A, Cheah Q, Dong J, et al. A systematic review of animal models of nafld finds high-fat, high-fructose diets most closely resemble human nafld. *Hepatology* (2021) 74(4):1884–901. doi: 10.1002/hep.31897

38. Martinez-Chantar ML, Vazquez-Chantada M, Ariz U, Martinez N, Varela M, Luka Z, et al. Loss of the glycine n-methyltransferase gene leads to steatosis and hepatocellular carcinoma in mice. *Hepatology* (2008) 47(4):1191–9. doi: 10.1002/hep.22159

39. Charlton M, Viker K, Krishnan A, Sanderson S, Veldt B, Kaalsbeek AJ, et al. Differential expression of lumican and fatty acid binding protein-1: New insights into the histologic spectrum of nonalcoholic fatty liver disease. *Hepatology* (2009) 49(4):1375–84. doi: 10.1002/hep.22927

40. Coilly A, Desterke C, Guettier C, Samuel D, Chiappini F. Fabp4 and Mmp9 levels identified as predictive factors for poor prognosis in patients with nonalcoholic fatty liver using data mining approaches and gene expression analysis. *Sci Rep* (2019) 9(1):19785. doi: 10.1038/s41598-019-56235-y

41. Chalasani N, Abdelmalek MF, Garcia-Tsao G, Vuppalanchi R, Alkhouri N, Rinella M, et al. Effects of belapectin, an inhibitor of galectin-3, in patients with nonalcoholic steatohepatitis with cirrhosis and portal hypertension. *Gastroenterology* (2020) 158(5):1334–45.e5. doi: 10.1053/j.gastro.2019.11.296

42. Eriksen PL, Vilstrup H, Rigbolt K, Suppli MP, Sørensen M, Heebøll S, et al. Non-alcoholic fatty liver disease alters expression of genes governing hepatic nitrogen conversion. *Liver Int* (2019) 39(11):2094–101. doi: 10.1111/liv.14205

43. García-Cañaveras JC, Donato MT, Castell JV, Lahoz A. A comprehensive untargeted metabonomic analysis of human steatotic liver tissue by rp and hilic chromatography coupled to mass spectrometry reveals important metabolic alterations. *J Proteome Res* (2011) 10(10):4825–34. doi: 10.1021/pr200629p

44. Masoodi M, Gastaldelli A, Hyötyläinen T, Arretxe E, Alonso C, Gaggini M, et al. Metabolomics and lipidomics in nafld: Biomarkers and non-invasive diagnostic tests. *Nat Rev Gastroenterol Hepatol* (2021) 18(12):835–56. doi: 10.1038/s41575-021-00502-9

45. Ryaboshapkina M, Hammar M. Human hepatic gene expression signature of non-alcoholic fatty liver disease progression, a meta-analysis. *Sci Rep* (2017) 7 (1):12361. doi: 10.1038/s41598-017-10930-w

46. Jia X, Zhai T. Integrated analysis of multiple microarray studies to identify novel gene signatures in non-alcoholic fatty liver disease. *Front Endocrinol (Lausanne)* (2019) 10:599. doi: 10.3389/fendo.2019.00599

frontiers | Frontiers in Endocrinology

TYPE Original Research
PUBLISHED 18 November 2022
DOI 10.3389/fendo.2022.1039786

OPEN ACCESS

EDITED BY
Bertrand Duvillie,
Institut Curie, France

REVIEWED BY
Guozhu Ye,
Institute of Urban Environment (CAS), China
Jarlei Fiamoncini,
University of São Paulo, Brazil

*CORRESPONDENCE
Yanfeng Liu
u202918@163.com
Peiyuan Yin
yinpeiyuan@dmu.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Cancer Endocrinology, a section of the journal Frontiers in Endocrinology

RECEIVED 08 September 2022
ACCEPTED 03 November 2022
PUBLISHED 18 November 2022

CITATION
Pan C, Deng D, Wei T, Wu Z, Zhang B, Yuan Q, Liang G, Liu Y and Yin P (2022) Metabolomics study identified bile acids as potential biomarkers for gastric cancer: A case control study. *Front. Endocrinol.* 13:1039786. doi: 10.3389/fendo.2022.1039786

Metabolomics study identified bile acids as potential biomarkers for gastric cancer: A case control study

Chen Pan[1,2,3†], Dawei Deng[1,2,4†], Tianfu Wei[1], Zeming Wu[5], Biao Zhang[1], Qihang Yuan[1], Guogang Liang[1], Yanfeng Liu[1*] and Peiyuan Yin[2,6*]

[1]Department of General Surgery, First Affiliated Hospital of Dalian Medical University, Dalian, China, [2]Clinical Laboratory of Integrative Medicine, First Affiliated Hospital of Dalian Medical University, Dalian, China, [3]Department of General Surgery, The First Affiliated Hospital of University of Science and Technology of China (USTC), Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China, [4]Department of Hepato-Biliary-Pancreas, Affiliated Hospital of North Sichuan Medical College, Nanchong, China, [5]iPhenome Biotechnology (Yun Pu Kang) Inc., Dalian, China, [6]Institute of Integrative Medicine, Dalian Medical University, Dalian, China

Gastric cancer (GC) is a common lethal malignancy worldwide. Gastroscopy is an effective screening technique for decreasing mortality. However, there are still limited useful non-invasive markers for early detection of GC. Bile acids are important molecules for the modulation of energy metabolism. With an in-depth targeted method for accurate quantitation of 80 bile acids (BAs), we aimed to find potential biomarkers for the early screening of GC. A cohort with 280 participants was enrolled, including 113 GC, 22 benign gastric lesions (BGL) and 145 healthy controls. Potential markers were identified using a random forest machine algorithm in the discovery cohort (n=180), then validated in an internal validation cohort (n=78) and a group with 22 BGL. The results represented significant alterations in the circulating BA pool between GC and the controls. BAs also exhibited significant correlations with various clinical traits. Then, we developed a diagnostic panel that comprised six BAs or ratios for GC detection. The panel showed high accuracy for the diagnosis of GC with AUC of 1 (95%CI: 1.00-1.00) and 0.98 (95%CI: 0.93-1.00) in the discovery and validation cohort, respectively. This 6-BAs panel was also able to identify early GC with AUC of 1 (95%CI: 0.999-1.00) and 0.94 (95%CI: 0.83-1.00) in the discovery and validation cohort, respectively. Meanwhile, this panel achieved a good differential diagnosis between GC and BGL and the AUC was 0.873 (95% CI: 0.812-0.934). The alternations of serum bile acids are characteristic metabolic features of GC. Bile acids could be promising biomarkers for the early diagnosis of GC.

KEYWORDS

gastric cancer, bile acids, metabolomics, biomarkers, LC MS

Introduction

Gastric cancer (GC) is a highly aggressive and fatal malignancy with high mortality and accounts for the second leading cause of cancer-related deaths worldwide (1, 2). Pathological grading of GC plays an essential role in determining patient prognosis (3, 4). Screening and early diagnosis of GC are critical in the prevention and treatment of GC. However, existing non-invasive tumor markers, such as carcinoembryonic antigen (CEA), have low GC evaluation efficiency, especially early gastric cancer (EGC) (5, 6). Endoscopy is commonly utilized in the screening and diagnosis of GC in clinical practice and has dramatically improved the disease outcome. However, gastroscopy consumes tremendous medical resources, and its cost-benefit remains debatable. Besides, because of its invasiveness, gastroscopy causes great anxiety to the subjects (7, 8). Therefore, current screening strategies only cover high-risk individuals who are older than 40 years or those with a prior history of gastropathy (9). There is an urgent need for innovative biomarkers to screen high-risk populations who require gastroscopy.

GC development has been associated with both genetic and environmental factors. In addition, metabolites are the end products of a complex interplay between intrinsic metabolism, environmental exposure and genetic predisposition (10). Occurrence of metabolic reprogramming in GC coupled with variations in the metabolites facilitates understanding of tumor biology. Previous metabolomic data showed that energy metabolism, amino acid metabolism and lipid metabolism are related to GC progression (11–13). Besides, GC with peritoneal metastasis depends on unique metabolic features (14). Thus, metabolomics, a new omics technique, provides a powerful tool for GC understanding.

On the other hand, bile reflux has been shown to be an independent risk factor for precancerous gastric lesions and GC (15). Bile acids (BAs), an important component of bile, play a significant role in regulating the digestive system and homeostasis of intestinal flora. A previous study demonstrated that BAs interact with the gut flora to influence human health (16). The metabolism of BAs was significantly disrupted in patients with Alzheimer's disease (AD), which was associated with cognitive impairment (17). Increasing evidence has indicated that BAs were involved in the occurrence and development of gastrointestinal tumors. For instance, deoxycholic acid (DCA) was shown to induce the expression of hepatocyte inflammatory genes, whose long-term expression was strongly associated with hepatocellular carcinoma (HCC) (18). However, the relationship between BA metabolism disorders and GC remains unclear.

Previous studies have shown that metabolites in blood or urine have the potential to be GC biomarkers (14, 19). Since quantification of the metabolites is the ultimate goal of metabolomics (20), most of the studies had a relatively small sample size and lacked a validation cohort, thus low quantification power. Besides, most of the previous studies utilized non-targeted metabolomic methods which had poor data stability, repeatability and quantitative linear range, limiting the clinical transformation of the outcome. This study used an in-depth targeted method developed for accurate quantitation of the BAs to analyze 280 blood samples from GC patients, benign gastric lesions(BGL) and healthy controls. Our study aimed to dissect the pathophysiologic interaction between BA metabolism and GC to identify biomarkers for early diagnosis.

Abbreviations: AD, Alzheimer's disease; AUC, area under the curve; BA, bile acid; CEA, carcinoembryonic antigen; CI, credible intervals; DCA, deoxycholic acid; DCA-3G, deoxycholic acid 3-glucuronide; ECG, early gastric cancer; FDR, false discovery rate; GC, gastric cancer; HCA, hyocholic acid; HCC, hepatocellular carcinoma; HDCA, hyodeoxycholic acid; IHC, immunohistochemistry; LCA, lithocholic acid; NF-κB, nuclear factor κB; NorCA, norcholic acid; OPLS-DA, orthogonal projections to latent structures discriminant analysis; PKC, protein kinase C; QC, quality control; ROC, receiver operating characteristic; SD, standard deviation; TLCA, taurolithocholic acid; TLCA-3S, taurolithocholic acid 3-sulfate; XIC, extracted ion chromatogram.

Materials and methods

Participants and criteria

Serum samples were collected from the GC patients, BGL patients and healthy participants (Con) at the First Affiliated Hospital of Dalian Medical University from May 2020 to October 2021. The GC patients were pathologically diagnosed by biopsy and divided into sub-groups based on the AJCC staging system, 8th edition (21). The sub-groups included the degree of differentiation, TNM stage as well as early or advanced GC. Samples in the Con group were collected from healthy participants during physical examination, and there was no obvious abnormality as assessed by gastroscopy. The age and gender ratio of the Con group were matched with the GC group. We excluded patients in the GC who had other forms of cancer, liver or renal insufficiency, severe cardiopulmonary diseases, metabolic diseases, active bleeding, and other mental or physical diseases. All 22 cases of BGL were confirmed by pathological biopsy. It mainly included chronic atrophic gastritis, adenomatous polyps, gastric ulcer and low grade intraepithelial neoplasia. On the other hand, we excluded any patients who had a history of gastrointestinal diseases, such as acute or chronic gastritis, upper gastrointestinal ulcers, upper gastrointestinal perforation, gastroesophageal reflux disease or benign tumors in the Con group. All the participants signed informed consent forms, and the study was approved by the Ethics Committee of First Affiliated Hospital of Dalian Medical University.

Serum sample collection and pretreatment

The GC patient serum samples under fasting were collected on the first morning after admission. All the BGL and Con samples were collected simultaneously and under the same fasting conditions as the GC samples. The samples were then stored at -80°C and thawed before pretreatment. At the beginning of the pretreatment, we transferred 80 µL of serum to 1 mL 96-well plate, then added 320 µL of mixed BA isotope internal standard dissolved in acetonitrile: methanol (1:1, v:v), as shown in Table S4. After 3 minutes of vortexing, the solution was then centrifuged for 20 minutes. We transferred 260 µL of the supernatant to another 96-well plate and then dried the extraction by centrifugal vacuum concentration (Labconco Corporation, USA). The remaining supernatant in all samples was mixed and distributed at the same volume as those in quality control (QC) samples (22–24). Before the BA-targeted metabolomic analysis, the extraction was redissolved in 50% methanol in water.

Metabolomic analysis

After injection of 2.5 µL of redissolved BA extraction, a total of 63 BAs (Table S5) were target detected by a Shimadzu UPLC (Shimadzu, Kyoto, Japan), coupled with a Sciex 5500+ triple quadrupole (QQQ) mass spectrometer (AB Sciex, Singapore). The BAs were separated on a C18-PFP column (ACE, UK, 3 µm, 2.1 × 50 mm). Phase A was composed of 2 mM ammonium acetate in water, while phase B contained acetonitrile. The chromatographic gradient was configured as follows: in 0 minutes, 83% phase A and 17% phase B; in 10 minutes, 70% phase A and 30% phase B; in 13 minutes, 45% phase A and 55% phase B; and in 14 and 17 minutes, 5% phase A and 95% phase B. The last 5 minutes was used for column washing and equilibration. We used 0.4 mL min-1 as the flow rate. The BAs were ionized by a Turbo-V heated electrospray ionization source and then detected by scheduled multiple reaction monitoring modes.

Date processing

The targeted BA annotation was based on the BA standards (25, 26). We compared the primary and secondary mass spectrometry data of the targeted BAs with the standards, as previously described. We calculated quantitative data of each sample by combining the standard curve and the area under the curve (AUC) BA values. Finally, internal standards were used for calibration. The above analyses were conducted with Analyst and OS-MQ software (AB SCIEX, Singapore).

Statistical analysis

The BAs with missing values of more than 50% were excluded, and then we employed the K-nearest algorithm to impute the missing values (27). The molar concentration of the serum sample was calculated from the mass concentration and molecular weight. In addition, SIMCA-P software (Umetrics, Sweden) was used for orthogonal projections to latent structures discriminant analysis (OPLS-DA) (28). Benjamini-Hochberg false discovery rate (FDR) adjustment for Student's t-test was performed on the MetaboAnalyst website (29–31). In addition, Random Forest and glmnet package were executed in R software Version 4.0.5 (R Core Team, 2021) (32). The receiver operating characteristic (ROC) curves (33), column diagrams, scatter plots and heat-maps were drawn by GraphPad Prism 9.0 (GraphPad Software Inc., USA). Besides, a biological network was generated by Cytoscape 3.8.2 (Cytoscape Consortium, USA) (34, 35).

Results

Study design and characteristics of BA metabolism

As shown in Figure 1, totally 180 participants were enrolled in the discovery cohort, including 79 GC and 101 controls. Based on the AJCC staging system, 8th edition, there were 32 patients in stage I, 9 patients in stage II, 22 patients in stage III and 16 patients in stage IV (Figure 1). We defined cohort 1 as the discovery set and used machine learning to conduct a diagnosis model (Table S1). To verify the results, cohort 2 and cohort 3 were included. Cohort 2 was validation set, including 34GC and 44Con, to detect repeatability of the model (Table S2). Cohort 3 included 22 BGL for verifying the differential diagnosis effect of the diagnostic model (Table S3). There was no significant difference in age and gender ratio (male/female) among GC and Con, whether in discovery set or validation set. Other related clinical traits were shown in Table 1 by the form of mean ± standard deviation (SD). Besides, we presented an extracted ion chromatogram (XIC), which provided a visual representation of the analyzed targeted BAs in the GC and Con groups as shown in (Figure S1). Through univariate and multivariate analyses, we obtained differentially expressed BAs between the GC and Con groups. Notably, our analyses showed that there was no significant difference between the two groups in clinical features such as gender and age.

After screening out the BAs with excess missing values or unstable detection, we obtained a total of 49 BAs. We then added biologically significant ratios and total concentrations of sub-classes to then BAs. By calculating the BAs (ratios or sub-classes) with significant differences between the GC and Con groups, a heat-map of 34 features was drawn (Figure 2A). Overall, patients

10.3389/fendo.2022.1039786

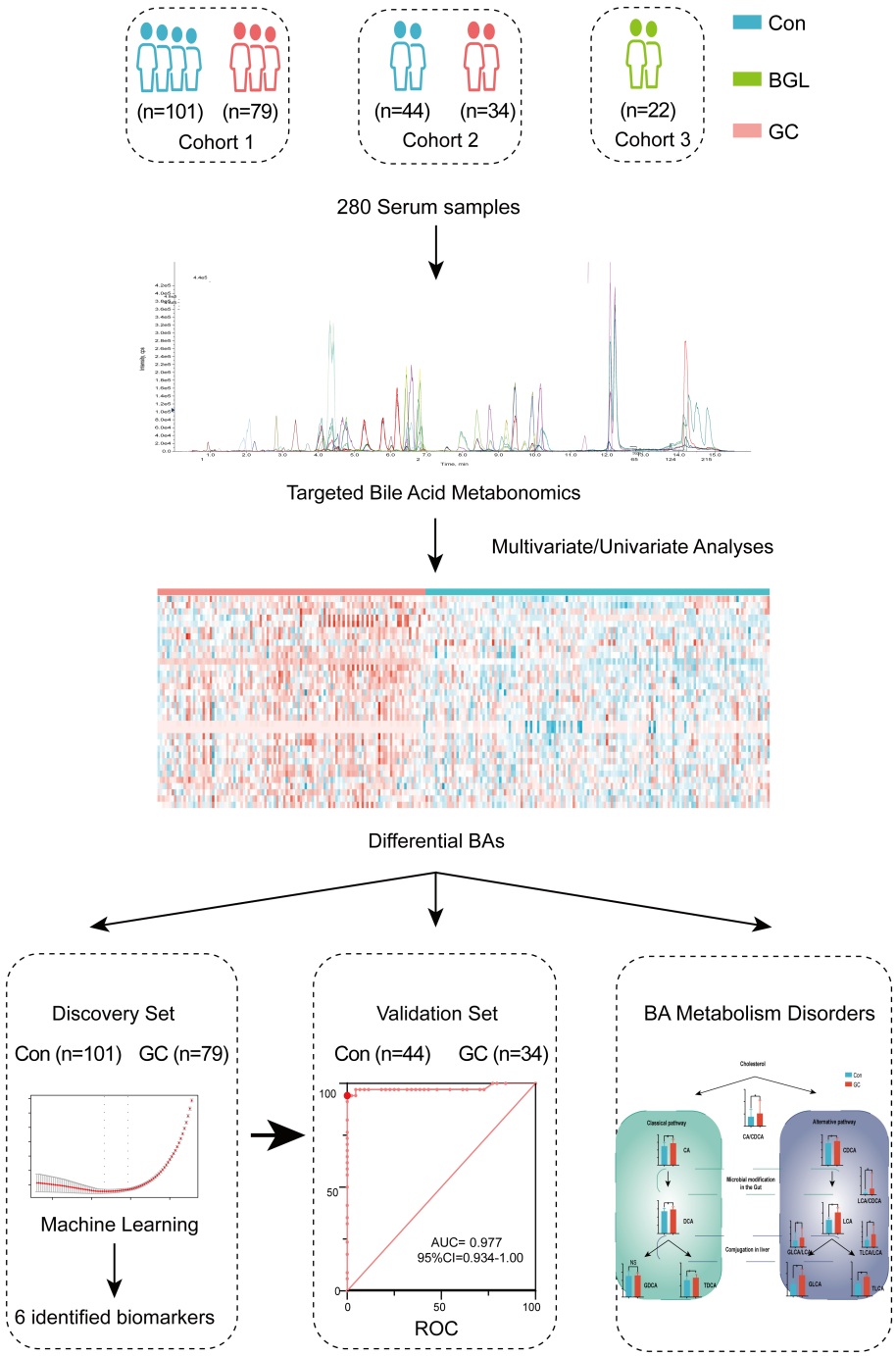


FIGURE 1
Global design and schematic representation of the study.

in the GC group had higher total BA concentration compared with the Con group. Interestingly, in the GC group, the levels of some BA sub-classes of interest, including conjugated BAs, unconjugated BAs, sulfate BAs, glucuronide BAs and HCAs showed different degrees of increase (Figure 2B).

Correlation analysis of the BAs and clinical traits

To analyze the correlation among the serum BAs and clinical traits in the patients with GC, we performed a Spearman

10.3389/fendo.2022.1039786

TABLE 1 Patient characteristics for the discovery and internal validation cohorts.

Characteristics	Discovery cohort (n = 180)			Validation cohort (n = 78)		
	GC (n = 79)	Con (n = 101)	p-value	GC (n = 34)	Con (n = 44)	p-value
Age (years)	63.19 ± 11.31	60.71 ± 11.92	0.21	63.65 ± 8.91	59.39 ± 12.67	0.99
Gender, male (%)	70.89	72.28	0.94	67.65	79.55	0.24
WBC (*10^9/L)	5.61 ± 1.5	6.00 ± 1.47	0.085	6.01 ± 1.69	6.11 ± 1.3	0.78
HGB (g/L)	124.14 ± 23.85	133.10 ± 41.50	0.091	121.29 ± 26.49	136.04 ± 40.97	0.074
Cre (μmol/L)	70.06 ± 17.22	70.97 ± 12.09	0.69	69.56 ± 23.79	73.52 ± 12.53	0.35
Urea (μmol/L)	12.02 ± 56.11	5.40 ± 1.14	0.25	5.65 ± 1.87	5.62 ± 1.56	0.94
Glucose (mmol/L)	5.49 ± 2.01	5.58 ± 1.41	0.72	5.24 ± 1.14	7.36 ± 13.11	0.35
ALT (U/L)	15.65 ± 10.4	23.24 ± 13.90	<0.001	14.91 ± 9.04	22.07 ± 12.97	0.0078
AST (U/L)	18.37 ± 7.63	21.94 ± 6.32	<0.001	17.12 ± 5.61	21.07 ± 6.06	0.0044
TBIL(μmol/L)	12.82 ± 5.58	16.02 ± 5.92	<0.001	11.33 ± 4.37	15.16 ± 6.34	0.004
DBIL(μmol/L)	3.6 ± 2.09	5.16 ± 5.03	0.012	3.48 ± 1.65	4.63 ± 2.07	0.011
TBA (μmol/L)	5.93 ± 5.83	—	—	4.52 ± 2.43	—	—
Cancer Stage						
I	32	—	—	15	—	—
II	9	—	—	1	—	—
III	22	—	—	12	—	—
IV	16	—	—	6	—	—
CEA (ng/ml)	8.19 ± 18.69	—	—	7.02 ± 22.75	—	—
CA199 (U/mL)	54.25 ± 170.26	—	—	154.19 ± 492.32	—	—

Data are presented as mean ± SD. WBC, white blood cell; HGB, hemoglobin; Cre, creatinine; ALT, alanine transaminase; AST, glutamic oxaloacetic transaminase; TBIL, total bilirubin; DBIL, direct bilirubin; TBA, total bile acid; CEA, carcinoembryonic antigen; CA199, carbohydrate antigen 199.

correlation analysis. We selected eight clinical traits related to GC, including Size, CEA or Stage, which represented the pathological features of GC. We separately classified the BAs strongly correlated with each clinical trait (Figure 3; Table S6). These data indicated that the BAs have commendable reactivity with clinical traits related to GC and could be characteristic biomarkers for clinical diagnosis of GC.

Identification of diagnostic markers for GC

Due to lack of effective screening tools and diagnostic markers for GC, we used targeted quantitated BAs to construct a diagnostic signature. LASSO regression analysis based on glmnet R package was performed to screen BA biomarkers that could be used for GC diagnosis. Six BAs (ratio), including HCA, TLCA, NorCA, DCA-3G, TLCA-3S and HDCA/LCA were obtained (Table 2).

The discovery set was used to construct the diagnostic panel. The OPLS-DA score plot showed an obvious separation between the GC group and the Con group (Figure 4A). To validate the classification, we performed a permutation test (Figure 4B). The results showed that the Y-axis intercept of R2 and Q2 was 0.212 and -0.428 (usually R2 and Q2 were less than 0.4 and 0, respectively), with a significant positive slope, which indicated that the data in the discovery set was not overfitted and reliable. Surprisingly, in the discovery set, the predictive ability of the GC diagnostic panel based on the RF model was perfect, and its sensitivity and specificity were 100%. Meanwhile, the AUC value of the model was 1 (95%CI: 1.00-1.00) (Figure 4C). Thereafter, we included another cohort as a validation set to evaluate the diagnostic model. As shown in Figure 4D, there were significant differences between the GC and the Con groups. Similarly, the verification data was not overfitted (Figure 4E). In the validation set, the predictive ability of the GC diagnostic panel was also satisfactory. Its sensitivity and specificity were 94.1% and 100%, respectively (Figure 4F). In addition, the AUC value of the verification set model was 0.98 (95%CI: 0.93-1.00). Besides, among all the GC and Con samples in the combined set, age (60 years old as the boundary) and gender had no significant effect on the predicted ability of the diagnostic panel (Figure S2). In addition, there was no correlation between the predicted ability of the diagnostic panel and whether GC was in an advanced stage, nor with AJCC tumor stage (Figure S3). These data showed that the diagnostic ability of the developed diagnostic panel was free from tumor load, which makes it an optimal diagnostic tool for the detection of GC.

The calculated model cut-off value in the discovery set was applied to the validation set as shown in Figure 4G. Participants

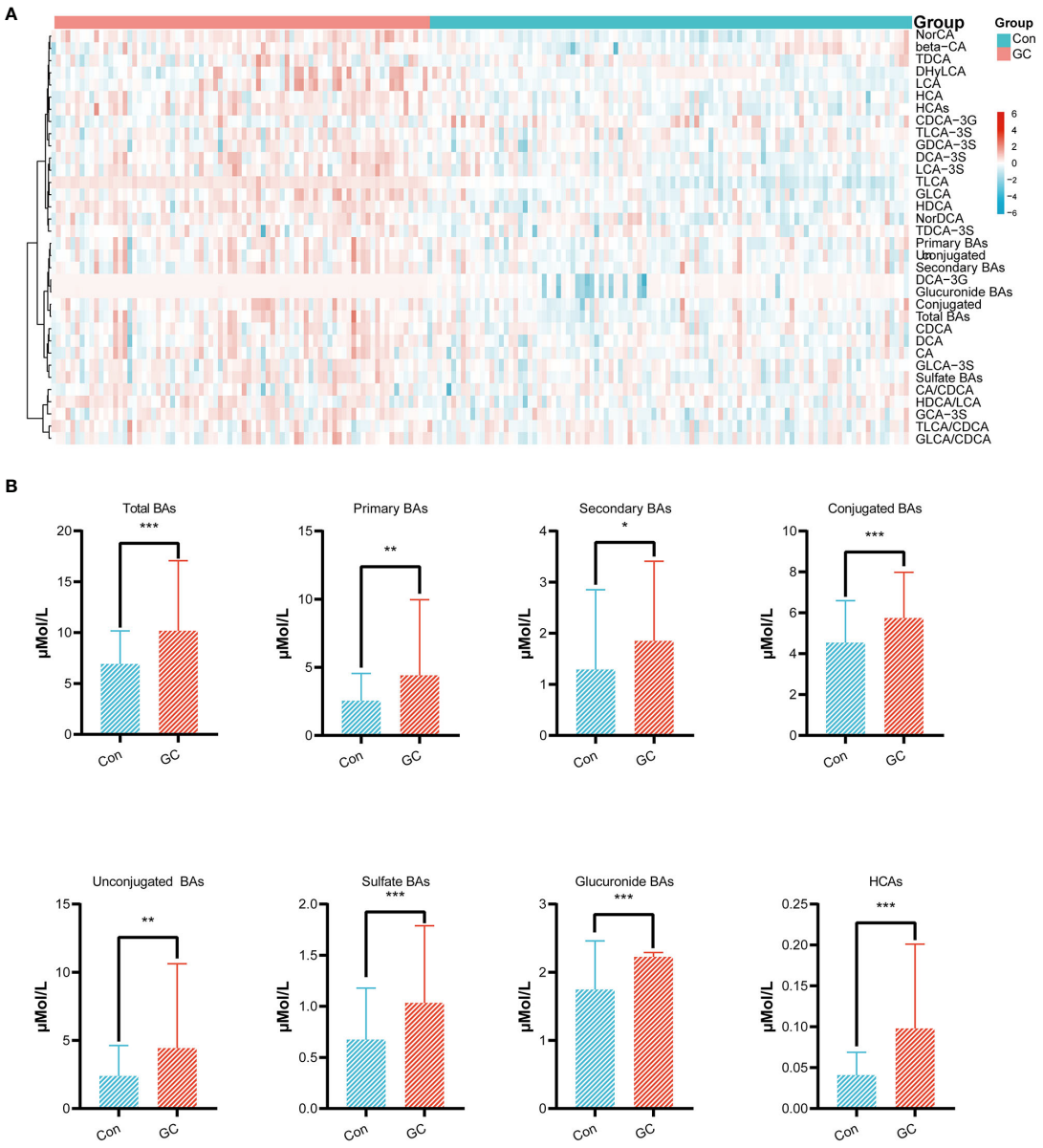


FIGURE 2
Overview of Bile acids. Differential expression of BAs, BA ratios and total concentration of BA sub-classes between the GC and Con groups were plotted as heat-map **(A)**. In this plot, red represents a higher concentration, while blue represents a lower concentration. Concentrations of eight kinds of BA sub-classes in the GC and Con group. For all figures, FDR-adjusted Q-value: *Q < 0.05; **Q < 0.01; ***Q < 0.001 **(B)**.

with a prediction probability of more than 0.64 were categorized as GC; otherwise, they were classified as Con group. To visually demonstrate the difference between the diagnostic panel and the CEA, a scatter plot was created to distinguish GC from Con (Figure 4H). It was not surprising to observe that CEA had superior sensitivity in distinguishing GC from Con, but lacked specificity. Together, the BA diagnostic panel showed promising sensitivity and specificity in the diagnosis of GC.

Identification of diagnostic markers for EGC

Early diagnosis is crucial for the prognosis of GC. Here, we constructed a diagnostic model for EGC. Our analysis showed that there are obvious metabolic differences between the EGC group and the Con group (Figure 5A). Likewise, the data of the discovery set was not overfitted (Figure 5B). The Y-axis intercept of R2 and

10.3389/fendo.2022.1039786

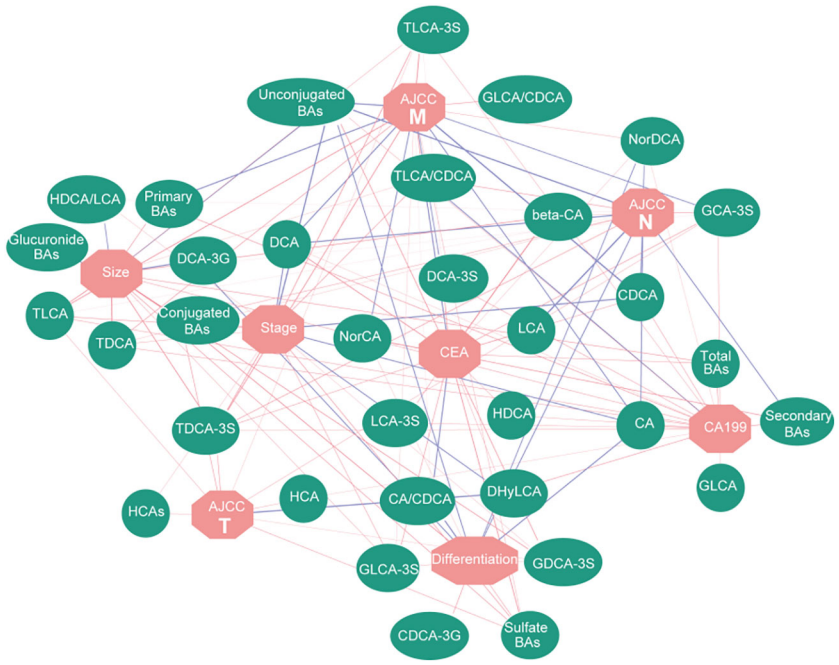


FIGURE 3
Correlation network based on Spearman correlation analysis. It shows the correlation between BAs and clinical traits. Red lines represent positive correlation, while blue lines represent negative correlation; the width of the line represents the correlation coefficient.

Q2 was 0.262 and -0.446, respectively, with a positive slope. Similarly, the ROC curve was also surprisingly perfect, and its sensitivity and specificity were 100% and 99%, respectively (Figure 5C). Besides, the AUC value of this model was 1 (95% CI: 0.999-1.00). In addition, the diagnostic model of the verification set yielded better results, and the OPLS-DA clearly separated the EGC group from the Con group (Figure 5D). Moreover, the data from the verification set was still not overfitted (Figures 5E, S4A). Unexpectedly, as shown in Figure 5F, the prediction rate of ROC curve remained high (Sensitivity=92.9%, Specificity=100%, 95%CI: 0.83-1.00, AUC=0.94). As previously mentioned, we used the same method to determine a cut-off value of 0.63, which distinguished the EGC from the Con (Figures 5G, H).

Identification of differential diagnostic markers for GC and BGL

The differential diagnosis between BGL and GC can help doctors to make a preliminary judgment, whether the disease is benign or malignant, without relying on pathological examination. Hence, we used 22 cases of BGL and 113 cases of GC to further verify the differential diagnostic efficacy of 6-BAs diagnostic panel. The results suggested that there were significant metabolic differences between GC and BGL group (Figure 6A). Meanwhile, this model did not perform a tendency of overfit (Figure 6B). The AUC value of ROC plot was 0.873 (95%CI: 0.812-0.934), which sensitivity and specificity were 63.7% and 100% respectively (Figure 6C). Likewise, The

TABLE 2 Characteristics of differential expression of the six markers for GC detection identified in this study.

Biomarker	P-value	Q-value	Log2 (Fold Change)	AUC of ROC
HCA	9.38E-05	0.00048	0.81	0.65 (95% CI:0.58-0.71)
TLCA	4.64E-67	3.57E-65	1.32	0.99 (95% CI:0.97-1)
NorCA	1.34E-13	5.16E-12	0.81	0.79(95% CI:0.73-0.84)
DCA-3G	7.17E-12	1.84E-10	0.38	0.70 (95% CI:0.64-0.77)
TLCA-3S	0.0016	0.0056	0.67	0.60 (95% CI:0.53-0.67)
HDCA/LCA	0.0061	0.016	0.91	0.62 (95% CI:0.55-0.69)

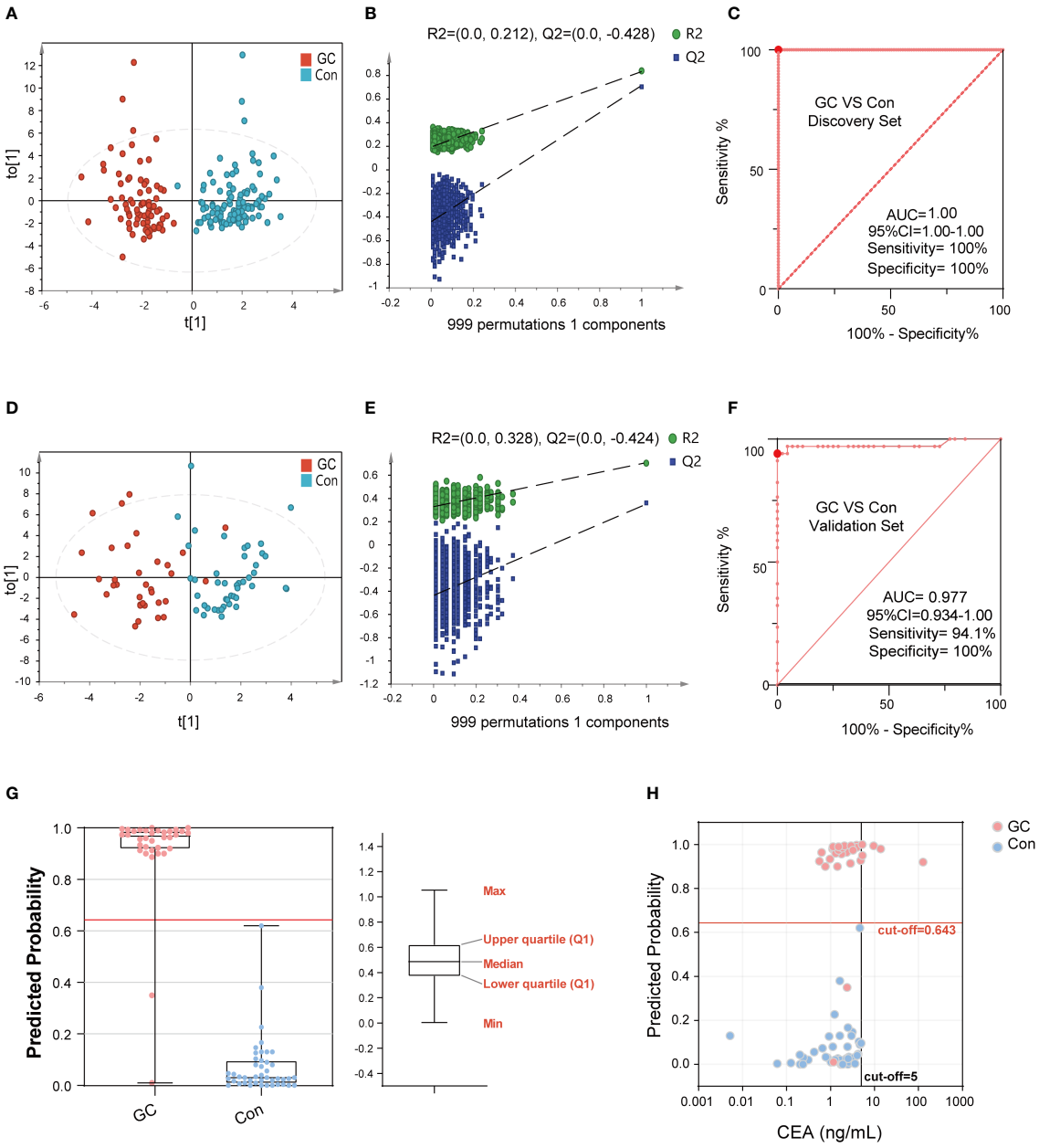


FIGURE 4
The performance of BA biomarkers in distinguishing GC and Con group. The OPLS-DA score plot of the GC and Con groups in discovery set **(A)**. The permutation test result of the above OPLS-DA, which could verify whether the classification is overfitted **(B)**. The ROC curve of 6-BAs diagnostic panel between the GC and Con group, which was constructed using the discovery set **(C)**. The OPLS-DA score plot of the GC and Con groups in the validation set **(D)**. The permutation test result of the above OPLS-DA. **(E)** The ROC curve of 6-BAs diagnostic panel between the GC and Con group, which was constructed in discovery set and test in validation set **(F)**. Predicted probability of the 6-BAs diagnostic panel identified in the discovery set and applied to the validation set **(G)**. Scatter plot for comparing the 6-BAs diagnostic panel and CEA. The CEA concentration of the Con is randomly selected from the upper limit of the normal value and zero **(H)**.

OPLS-DA score plot of EGC and BGL group also suggested that there were metabolic differences between BGL and the early stage of gastric cancer (Figure 6D). This model also has not been overfitted (Figures 6E, S4B). Finally, we found that for EGC, the AUC value of the ROC curve was 0.823 (95%CI: 0.725-0.921). Its sensitivity decreased to 55.6%. However, the specificity remains 100%. Obviously, unlike disease screening, differential diagnosis paid more attention to specificity. Therefore, 6-BAs diagnostic panel can also be used as a marker for differential diagnosis.

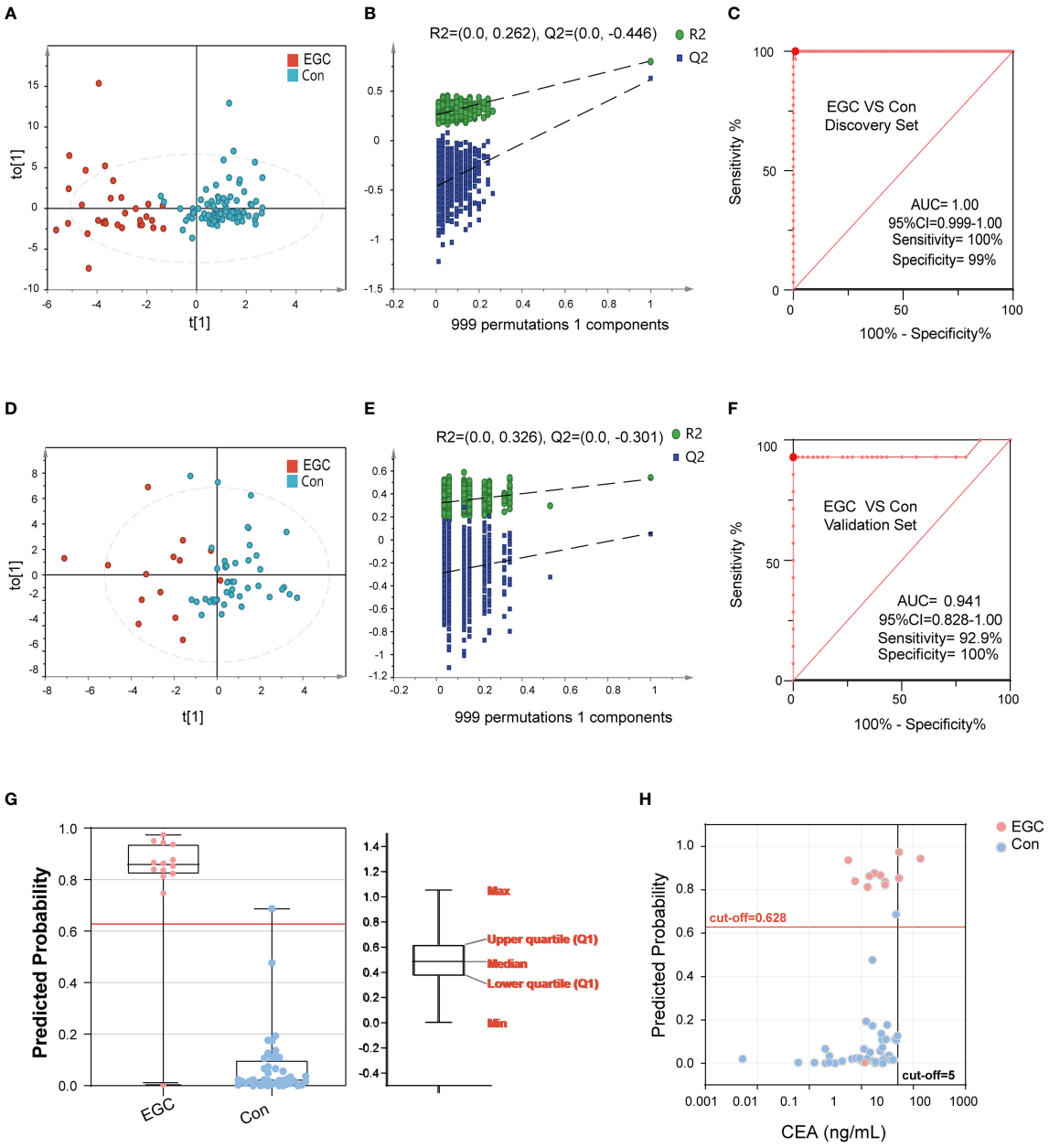


FIGURE 5
The performance of BA biomarkers in distinguishing early GC and Con group. **(A)**: OPLS-DA score plot of early GC and Con groups in the discovery set. **(B)**: permutation test result of the above OPLS-DA. **(C)**: ROC curve of 6-BAs diagnostic panel between early GC and Con group, which was constructed in the discovery set. **(D)**: OPLS-DA score plot of early GC and Con groups in the validation set. **(E)**: permutation test result of the above OPLS-DA. **(F)**: ROC curve of 6-BAs diagnostic panel between early GC and Con group, which was constructed in the discovery set and validation set. **(G)**: Predicted probability of the 6-BAs diagnostic panel identified in the discovery set and applied to the validation set. **(H)**: Scatter plot for comparing the 6-BAs diagnostic panel and CEA.

Discussion

In this study, we profiled serum BAs of 280 GC patients, BGL patients and healthy controls using targeted BA metabolomics. Our data showed that the total serum BA pool was significantly altered in the patients with GC. Moreover, these differential BAs were associated with clinical traits. Besides, we constructed a diagnostic panel using a machine learning algorithm consisting of six BA molecules. The constructed model demonstrated good diagnostic efficiency for BGL and EGC. The metabolism of BAs can be considered as a crucial indicator to the prevention and treatment of GC.

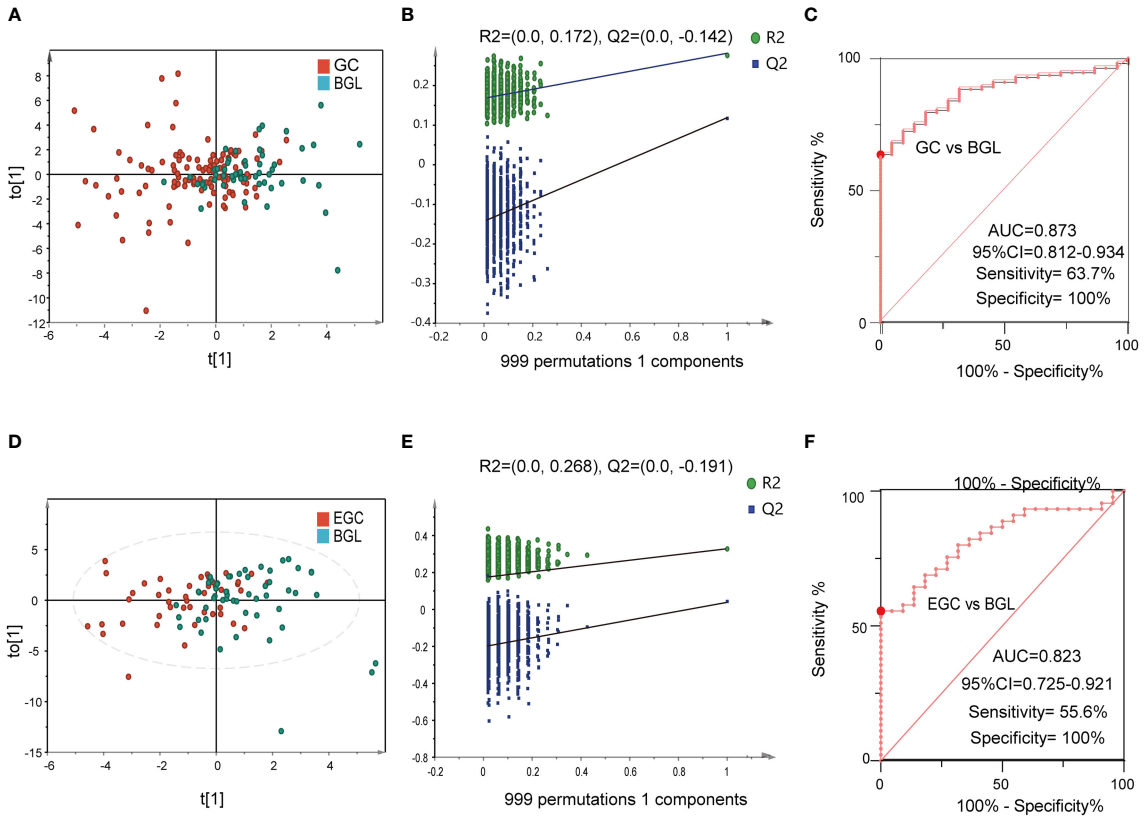


FIGURE 6
The performance of BA biomarkers in distinguishing GC and BGL group. **(A)**: OPLS-DA score plot of GC and BGL groups. **(B)**: permutation test result of the OPLS-DA result in **(A, C)**: ROC curve of 6-BAs diagnostic panel between GC and BGL group. **(D)**: OPLS-DA score plot of early GC (EGC) and BGL groups. **(E)**: permutation test result of the OPLS-DA result in **(D, F)**: ROC curve of 6-BAs diagnostic panel between EGC and BGL group.

Recent studies have shown that BAs are important signal molecules. In biological systems, BAs maintain glucose and lipid metabolism homeostasis and energy expenditure by acting on the receptors in peripheral tissues and organs, such as Farnesol X receptor FXR and TGR-5 (36–38). BAs are mainly metabolized through enterohepatic circulation and play a crucial role in regulating the functions of digestive tract and intestinal immunity (39). Besides, studies have shown that BAs are important regulatory molecules in tumors. In HCC, BAs were shown to directly incapacitate the plasma membrane, leading to the activation of protein kinase C (PKC), which activated the P38-MAPK pathway, resulting in increased activation of p53 and nuclear factor κB (NF-κB) which mediate cellular apoptosis and inflammation (40). In this study, there were significant metabolic changes in BAs in patients with GC. Increased synthesis of hepatic BAs or reabsorption of intestinal BAs led to an increased pool volume of the BAs. Improved BAs pool in GC may be beneficial in disease development. Besides, glucuronidated BAs and sulfated BAs were significantly increased in GC (Figure 7). The pathophysiological role of the BAs in GC remains unclear.

Pathological characteristics such as tumor stage and grade accurately reflects the degree of malignancy of cancers, and these characteristics have a strong guiding significance for clinic (41). However, these features can often only be accurately captured in an invasive fashion. Markers which are easily obtained under a noninvasive manner reflecting disease characteristics are urgently needed. Correlation analysis showed that BAs were closely related to T stage, N stage and grade, etc. Circulating BAs can be used as indicators reflecting the characteristics of GC.

The ratio of upstream and downstream molecules in the metabolic pathway can indirectly reveal the change of corresponding catalytic enzyme activity. By comparing the ratio of expression of adjacent metabolites, we demonstrated that there was enhancement of the metabolic activity of both the classical and alternative pathways in GC (42). Among these, the classical pathway has been more emphasized (CA/CDCA). However, there is a dramatic occurrence of metabolic disorders in the alternative pathway. Increased LCA/CDCA in GC indicated that intestinal flora enhances the catalytic activity of primary BAs CDCA, resulting in increased cytotoxic LCA. With the increased LCA uptake in the intestinal tract, there was

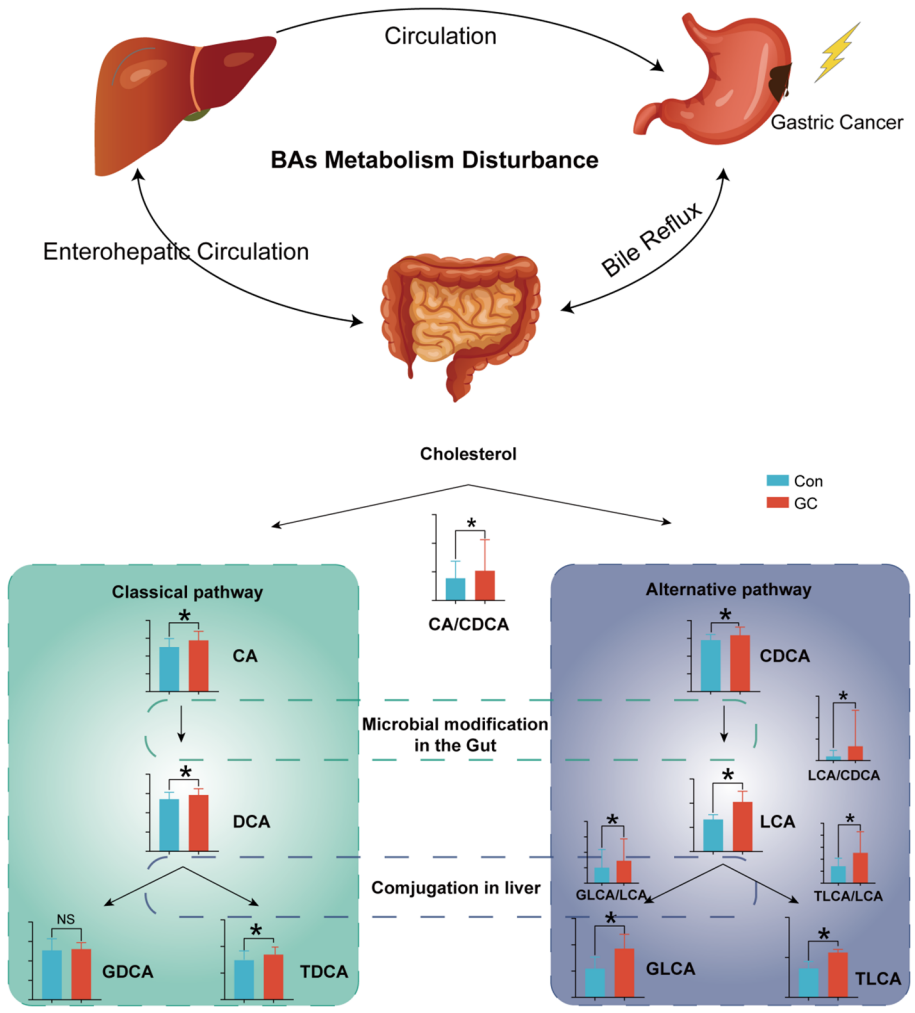


FIGURE 7
Overview of the BA metabolic dysregulation in GC. FDR-adjusted Q-value: *Q < 0.05.

increased activity of LCA-modifying enzymes in the liver (GLCA/LCA, TLCA/LCA). CDCA, LCA, TLCA are hydrophobic and cytotoxic, which lead to cell damage and inflammation (43, 44).

In this study, the diagnostic model constructed by the 6 molecules and ratios in GC had good diagnostic performance, even in EGC. The constructed panel had a high sensitivity and specificity for screening high-risk GC populations. Moreover, the diagnostic efficiency of the diagnostic model was independent of tumor load, which is an important biomarker for early cancer screening. Screening for early-stage cancer is the most critical step in the improvement of the current GC status, as these patients undergo less trauma, fewer complications and better prognosis. Among the 6 molecules, TLCA expression was the most differentially expressed metabolite, with an AUC of 0.99 (95% CI:0.97-1). TLCA was the strongest activator of TGR5 among all the BAs and promoted the occurrence of liver cancer by activating the TGR5 (45). Remodeling of energy metabolism is a common feature in tumor metabolic reprogramming. Hyocholic acid species (HCA and HDCA/LCA) play an important role in regulating insulin sensitivity, glucose homeostasis and energy expenditure (46).

The metabolism of BAs is mainly influenced by the catalysis of liver and intestinal microbiota. However, different diseases have been found to have specific features of BAs metabolism (42, 47). The diagnostic specificity can be enhanced by using multiple molecular model. Compared with a single molecule, our model is composed of six molecules and ratios which can reduce molecular noise and increase the accuracy of diagnosis. These 6 BAs could make up for the lack of non-invasive markers for GC screening and resolve the cost associated with endoscopy as a screening tool for GC.

We used an in-depth metabolomics approach which provided high-quality and information-rich BAs data in patients with GC. Differential BAs, including many low-abundant BAs, were annotated based on the BA standards. Due to limited detection methods, data on the low-abundance BAs remain scanty. Recent data has shown that some of these BAs are closely related to human health. Besides, BAs play a key role in regulating gastric mucosal homeostasis, which mediates stomach upset. Thus, the newly reported BAs could serve as screening markers. In addition, these molecules provide a new tool in understanding the pathophysiology of GC.

Although our study had an important outcome, there is a need to confirm the diagnostic value of those candidates in a multi-center large sample cohort. In addition, the molecular mechanism of BA metabolism disorder on the occurrence and development of GC needs further evaluation in cell assays.

Conclusion

Taken together, our data demonstrated that the BAs metabolism disorder is involved in GC development. Our diagnostic model using 6 BAs or ratio provided promising diagnostic efficiency for GC, which could perform early screening of high-risk populations and promote early diagnosis of GC.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of First Affiliated Hospital of Dalian Medical University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

CP: data curation, formal analysis, software, methodology, writing - original draft, and writing - review and editing. DD: data curation, formal analysis, investigation, methodology, project administration, writing - original draft, and writing - review and editing. TW: data curation, formal analysis, investigation, and methodology. ZW: methodology, formal analysis, resources, and validation. BZ: data curation and formal analysis. QY: data curation and formal analysis. GL: conceptualization, resources, validation, and supervision. YL: conceptualization, resources, project administration, validation, and supervision. PY: conceptualization, resources, validation, supervision, funding acquisition, project administration, writing - review and editing. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by The Key Research And Development Project Of Liaoning Province (No. 2018225054) and The National Natural Science Foundation of China (No. 81873156).

Conflict of interest

ZW and PY are co-founders of iPhenome Yun Pu Kang Biotechnology Inc. ZW is an employee of iPhenome Yun Pu Kang Biotechnology Inc.

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/fendo.2022.1039786/full#supplementary-material

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin* (2019) 69:7–34. doi: 10.3322/caac.21551

2. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. *CA Cancer J Clin* (2016) 66:115–32. doi: 10.3322/caac.21338

3. Son T, Hyung WJ. Laparoscopic gastric cancer surgery: Current evidence and future perspectives. *World J Gastroenterol* (2016) 22:727–35. doi: 10.3748/wjg.v22.i2.727

4. Zong L, Abe M, Seto Y, Ji J. The challenge of screening for early gastric cancer in China. *Lancet* (2016) 388:2606. doi: 10.1016/S0140-6736(16)32226-7

5. Feng F, Tian Y, Xu G, Liu Z, Liu S, Zheng G, et al. Diagnostic and prognostic value of CEA, CA19-9, AFP and CA125 for early gastric cancer. *BMC Cancer* (2017) 17:737. doi: 10.1186/s12885-017-3738-y

6. Hu P-J, Chen M-Y, Wu M-S, Lin Y-C, Shih P-H, Lai C-H, et al. Clinical evaluation of CA72-4 for screening gastric cancer in a healthy population: A multicenter retrospective study. *Cancers (Basel)* (2019) 11:E733. doi: 10.3390/cancers11050733

7. Ho SWT, Tan P. Dissection of gastric cancer heterogeneity for precision oncology. *Cancer Sci* (2019) 110:3405–14. doi: 10.1111/cas.14191

8. Chen J, Wu L, Sun Y, Luo C, Chen X, Wu L, et al. Diagnostic value and clinical significance of circulating miR-650 and CA211 in detecting of gastric carcinoma. *Oncol Lett* (2020) 20:254. doi: 10.3892/ol.2020.12117

9. Choi IJ. Endoscopic gastric cancer screening and surveillance in high-risk groups. *Clin Endosc* (2014) 47:497–503. doi: 10.5946/ce.2014.47.6.497

10. Peng B, Li H, Peng X-X. Functional metabolomics: from biomarker discovery to metabolome reprogramming. *Protein Cell* (2015) 6:628–37. doi: 10.1007/s13238-015-0185-x

11. Song H, Peng J-S, Dong-Sheng Y, Yang Z-L, Liu H-L, Zeng Y-K, et al. Serum metabolic profiling of human gastric cancer based on gas chromatography/mass spectrometry. *Braz J Med Biol Res* (2012) 45:78–85. doi: 10.1590/s0100-879x2011007500158

12. Huang S, Guo Y, Li Z-W, Shui G, Tian H, Li B-W, et al. Identification and validation of plasma metabolomic signatures in precancerous gastric lesions that progress to cancer. *JAMA Netw Open* (2021) 4:e2114186. doi: 10.1001/jamanetworkopen.2021.14186

13. Liang Q, Wang C, Li B. Metabolomic analysis using liquid Chromatography/Mass spectrometry for gastric cancer. *Appl Biochem Biotechnol* (2015) 176:2170–84. doi: 10.1007/s12010-015-1706-z

14. Kaji S, Irino T, Kusuhara M, Makuuchi R, Yamakawa Y, Tokunaga M, et al. Metabolomic profiling of gastric cancer tissues identified potential biomarkers for predicting peritoneal recurrence. *Gastric Cancer* (2020) 23:874–83. doi: 10.1007/s10120-020-01065-5

15. Zhang LY, Zhang J, Li D, Liu Y, Zhang DL, Liu CF, et al. Bile reflux is an independent risk factor for precancerous gastric lesions and gastric cancer: An observational cross-sectional study. *J Dig Dis* (2021) 22:282–90. doi: 10.1111/1751-2980.12986

16. Sato Y, Atarashi K, Plichta DR, Arai Y, Sasajima S, Kearney SM, et al. Novel bile acid biosynthetic pathways are enriched in the microbiome of centenarians. *Nature* (2021) 599:458–64. doi: 10.1038/s41586-021-03832-5

17. MahmoudianDehkordi S, Arnold M, Nho K, Ahmad S, Jia W, Xie G, et al. Altered bile acid profile associates with cognitive impairment in alzheimer's disease-an emerging role for gut microbiome. *Alzheimers Dement* (2019) 15:76–92. doi: 10.1016/j.jalz.2018.07.217

18. Yoshimoto S, Loo TM, Atarashi K, Kanda H, Sato S, Oyadomari S, et al. Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome. *Nature* (2013) 499:97–101. doi: 10.1038/nature12347

19. Kwon HN, Lee H, Park JW, Kim Y-H, Park S, Kim JJ. Screening for early gastric cancer using a noninvasive urine metabolomics approach. *Cancers (Basel)* (2020) 12:E2904. doi: 10.3390/cancers12102904

20. Yan M, Xu G. Current and future perspectives of functional metabolomics in disease studies-a review. *Anal Chim Acta* (2018) 1037:41–54. doi: 10.1016/j.aca.2018.04.006

21. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, et al. The eighth edition AJCC cancer staging manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA Cancer J Clin* (2017) 67:93–9. doi: 10.3322/caac.21388

22. Salem MA, Jüppner J, Bajdzienko K, Giavalisco P. Protocol: a fast, comprehensive and reproducible one-step extraction method for the rapid preparation of polar and semi-polar metabolites, lipids, proteins, starch and cell wall polymers from a single sample. *Plant Methods* (2016) 12:45. doi: 10.1186/s13007-016-0146-2

23. Alseekh S, Aharoni A, Brotman Y, Contrepois K, D'Auria J, Ewald J, et al. Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. *Nat Methods* (2021) 18:747–56. doi: 10.1038/s41592-021-01197-1

24. Zelena E, Dunn WB, Broadhurst D, Francis-McIntyre S, Carroll KM, Begley P, et al. Development of a robust and repeatable UPLC-MS method for the long-term metabolomic study of human serum. *Anal Chem* (2009) 81:1357–64. doi: 10.1021/ac8019366

25. Choucair I, Nemet I, Li L, Cole MA, Skye SM, Kirsop JD, et al. Quantification of bile acids: a mass spectrometry platform for studying gut microbe connection to metabolic diseases. *J Lipid Res* (2020) 61:159–77. doi: 10.1194/jlr.RA119000311

26. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. Proposed minimum reporting standards for chemical analysis chemical analysis working group (CAWG) metabolomics standards initiative (MSI). *Metabolomics* (2007) 3:211–21. doi: 10.1007/s11306-007-0082-2

27. Bartel J, Krumsiek J, Theis FJ. Statistical methods for the analysis of high-throughput metabolomics data. *Comput Struct Biotechnol J* (2013) 4:e201301009. doi: 10.5936/csbj.201301009

28. Saccenti E, Hoefsloot HCJ, Smilde AK, Westerhuis JA, Hendriks MMWB. Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics* (2014) 10:361–74. doi: 10.1007/s11306-013-0598-6

29. Vinaixa M, Samino S, Saez I, Duran J, Guinovart JJ, Yanes O. A guideline to univariate statistical analysis for LC/MS-based untargeted metabolomics-derived data. *Metabolites* (2012) 2:775–95. doi: 10.3390/metabo2040775

30. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc: Ser B (Methodological)* (1995) 57:289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

31. Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, et al. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res* (2018) 46:W486–94. doi: 10.1093/nar/gky310

32. Cutler A, Stevens JR. Random forests for microarrays. *Methods Enzymol* (2006) 411:422–32. doi: 10.1016/S0076-6879(06)11023-X

33. Lasko TA, Bhagwat JG, Zou KH, Ohno-Machado L. The use of receiver operating characteristic curves in biomedical informatics. *J BioMed Inform* (2005) 38:404–15. doi: 10.1016/j.jbi.2005.02.008

34. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* (2003) 13:2498–504. doi: 10.1101/gr.1239303

35. Basu S, Duren W, Evans CR, Burant CF, Michailidis G, Karnovsky A. Sparse network modeling and metscape-based visualization methods for the analysis of large-scale metabolomics data. *Bioinformatics* (2017) 33:1545–53. doi: 10.1093/bioinformatics/btx012

36. Lefebvre P, Cariou B, Lien F, Kuipers F, Staels B. Role of bile acids and bile acid receptors in metabolic regulation. *Physiol Rev* (2009) 89:147–91. doi: 10.1152/physrev.00010.2008

37. Chávez-Talavera O, Tailleux A, Lefebvre P, Staels B. Bile acid control of metabolism and inflammation in obesity, type 2 diabetes, dyslipidemia, and nonalcoholic fatty liver disease. *Gastroenterology* (2017) 152:1679–1694.e3. doi: 10.1053/j.gastro.2017.01.055

38. Watanabe M, Houten SM, Mataki C, Christoffolete MA, Kim BW, Sato H, et al. Bile acids induce energy expenditure by promoting intracellular thyroid hormone activation. *Nature* (2006) 439:484–9. doi: 10.1038/nature04330

39. Jia W, Xie G, Jia W. Bile acid-microbiota crosstalk in gastrointestinal inflammation and carcinogenesis. *Nat Rev Gastroenterol Hepatol* (2018) 15:111–28. doi: 10.1038/nrgastro.2017.119

40. Anwer MS. INTRACELLULAR SIGNALING BY BILE ACIDS. *J Biosci (Rajshari)* (2012) 20:1–23. doi: 10.3329/jbs.v20i0.17647

41. Atef E, El Nakeeb A, El Hanafy E, El Hemaly M, Hamdy E, El-Geidie A. Pancreatic cystic neoplasms: predictors of malignant behavior and management. *Saudi J Gastroenterol* (2013) 19:45–53. doi: 10.4103/1319-3767.105927

42. Nho K, Kueider-Paisley A, MahmoudianDehkordi S, Arnold M, Risacher SL, Louie G, et al. Altered bile acid profile in mild cognitive impairment and alzheimer's disease: Relationship to neuroimaging and CSF biomarkers. *Alzheimers Dement* (2019) 15:232–44. doi: 10.1016/j.jalz.2018.08.012

43. Ignacio Barrasa J, Olmo N, Pérez-Ramos P, Santiago-Gómez A, Lecona E, Turnay J, et al. Deoxycholic and chenodeoxycholic bile acids induce apoptosis *via*

oxidative stress in human colon adenocarcinoma cells. *Apoptosis* (2011) 16:1054–67. doi: 10.1007/s10495-011-0633-x

44. Perez M-J, Briz O. Bile-acid-induced cell injury and protection. *World J Gastroenterol* (2009) 15:1677–89. doi: 10.3748/wjg.15.1677

45. Kawamata Y, Fujii R, Hosoya M, Harada M, Yoshida H, Miwa M, et al. A G protein-coupled receptor responsive to bile acids. *J Biol Chem* (2003) 278:9435–40. doi: 10.1074/jbc.M209706200

46. Zheng X, Chen T, Jiang R, Zhao A, Wu Q, Kuang J, et al. Hyocholic acid species improve glucose homeostasis through a distinct TGR5 and FXR signaling mechanism. *Cell Metab* (2021) 33:791–803.e7. doi: 10.1016/j.cmet.2020.11.017

47. Deng D, Pan C, Wu Z, Sun Y, Liu C, Xiang H, et al. An integrated metabolomic study of osteoporosis: Discovery and quantification of hyocholic acids as candidate markers. *Front Pharmacol* (2021) 12:725341. doi: 10.3389/fphar.2021.725341

frontiers | Frontiers in Endocrinology

TYPE Original Research
PUBLISHED 29 November 2022
DOI 10.3389/fendo.2022.1035158

OPEN ACCESS

EDITED BY
Prasoon Agarwal,
National Bioinformatics Infrastructure Sweden, Sweden

REVIEWED BY
Shuai Ma,
Second Affiliated Hospital of Zhengzhou University, China
Shang-Qian Xie,
University of Idaho, United States
Yanhui Tan,
Guangxi Normal University, China
Xiujie Gao,
Guangdong Pharmaceutical University, China

*CORRESPONDENCE
Qingli Quan
bioquanqingli@163.com
Jia Tang
tony2081129@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Systems Endocrinology, a section of the journal Frontiers in Endocrinology

RECEIVED 02 September 2022
ACCEPTED 15 November 2022
PUBLISHED 29 November 2022

CITATION
Quan Q, Wu J, Yu M and Tang J (2022) Immune micro-environment and drug analysis of peritoneal endometriosis based on epithelial-mesenchymal transition classification. *Front. Endocrinol.* 13:1035158. doi: 10.3389/fendo.2022.1035158

Immune micro-environment and drug analysis of peritoneal endometriosis based on epithelial-mesenchymal transition classification

Qingli Quan[1*†], Jiabao Wu[1†], Meixing Yu[2] and Jia Tang[1*]

[1]NHC Key Laboratory of Male Reproduction and Genetics, Guangdong Provincial Reproductive Science Institute (Guangdong Provincial Fertility Hospital), Guangzhou, China, [2]Guangzhou Women and Children's Medical Center, Guangzhou Medical University, Guangzhou, China

Background: Epithelial-mesenchymal transition (EMT) is a complex event that drives polar epithelial cells transform from adherent cells to motile mesenchymal cells, in which are involved immune cells and stroma cells. EMT plays crucial roles in migration and invasion of endometriosis. The interaction of endometrial implants with the surrounding peritoneal micro-environment probably affects the development of peritoneal endometriosis. To date, very few studies have been carried out on peritoneal endometriosis sub-type classification and micro-environment analysis based on EMT. The purpose of this study is to investigate the potential application of EMT-based classification in precise diagnosis and treatment of peritoneal endometriosis.

Method: Based on EMT hallmark genes, 76 peritoneal endometriosis samples were classified into two clusters by consistent cluster classification. EMT scores, which calculated by Z score of 8 epithelial cell marker genes and 8 mesenchymal cell marker genes, were compared in two clusters. Then, immune scores and the abundances of corresponding immune cells, stroma scores and the abundances of corresponding stroma cells were analyzed by the "xCell" package. Futhermore, a diagnostic model was constructed based on 9 diagnostic markers which related to immune score and stroma score by Lasso-Logistic regression analysis. Finally, based on EMT classification, a total of 8 targeted drugs against two clusters were screened out by drug susceptibility analysis *via* "pRRophetic" package.

Results: Hallmark epithelial-mesenchymal transition was the mainly enriched pathway of differentially expressed genes between peritoneal endometriosis tissues and endometrium tissues. Compared with cluster 2, EMT score and the abundances of most infiltrating stroma cell were significantly higher, while the abundances of most infiltrating immune cells were dramatically less. The diagnostic model could accurately distinguish cluster 1 from cluster 2. Pathway analysis showed drug candidates targeting cluster 1 mainly act on the IGF-1 signaling pathway, and drug candidates targeting cluster 2 mainly block the EGFR signaling pathway.

Conclusion: In peritoneal endometriosis, EMT was probably promoted by stroma cell infiltration and inhibited by immune cell infiltration. Besides, our study highlighted the potential uses of the EMT classification in the precise diagnosis and treatment of peritoneal endometriosis.

KEYWORDS

epithelial-mesenchymal transition, peritoneal endometriosis, immune micro-environment, diagnostic model, drug susceptibility analysis

Introduction

Endometriosis is characterized by the presence of normal endometrium (like stroma and glands) abnormally invaded in body parts other than the uterine cavity, which shares many characteristics with malignant tumour (1) (2). Although ectopic endometrial tissue can be implanted in any parts of body, abdominal cavity is one of the most frequently locations that endometriotic tissue implanted into, leading to peritoneal endometriosis (1–4). Over the past decades, several systems have been proposed for endometriosis classification. The most widely accepted is American Society for Reproductive Medicine (rASRM) classification and the updated Enzian classification (Supplement to ASRM Classification) (5). However, the rASRM score has limitations in deep infiltrating endometriosis description and Enzian classification has not included peritoneal endometriosis classification (6), which is greatly limiting accurate diagnosis and treatment of peritoneal endometriosis.

Epithelial-mesenchymal transition (EMT) lead to the increased motility *via* rearrangements of cellular contact junctions, loss of cell adhesion, apicobasal polarity and epithelial cell morphology, thus promoting lesion metastasis (7, 8). In general, EMT of ectopic endometrial tissue is more active than that of eutopic endometrial tissue, which may be beneficial for migration and invasion of ectopic tissue (9). After endometrium attaches to peritoneum, endometrial epithelial cells also undergo EMT (10). Furthermore, the expressions of EMT induced transcription factors that may trigger EMT were significantly increased in deep endometriotic lesions than in eutopic endometrium (11, 12). These indicate EMT is a factor contributing to progression of endometriosis. Classification based on EMT hallmarks has been widely used in diseases sub-classify (13, 14), we supposed classification based on EMT also has a potential to be used on peritoneal endometriosis classification.

Immune micro-environment affects EMT (15, 16). Peritoneal endometriosis is markedly characterized by increased numbers of peritoneal macrophages and elevated concentrations of pro-inflammatory chemokines, which associated with endometriosis-related pain and infertility (17, 18). Macrophages induced EMT in pancreatic cancer cells (19). And inflammatory mediators in retrograde menstrual fluid probably contribute to ectopic endometrial EMT in the presence of peritoneal hypoxia (20). Besides, in superficial peritoneal endometriosis, the migration and infiltration of peritoneal endometriotic tissue were also associated with the formation and differentiation of stroma cells, such as myofibroblasts and smooth muscles (SM)-like cells (21). All these made us curious about the differences in immune cell infiltration and stroma cell infiltration of peritoneal endometriosis classified based on EMT classification.

Here, we classified peritoneal endometriosis into two clusters based on EMT hallmark genes by consistent cluster classification, which is suitable for diseases classification from the perspective of molecular (22, 23) Then, we compared the immune micro-environment and stroma cells infiltration of two clusters. What was more, based on EMT classification, we established a diagnostic model and screened potential drugs against different clusters. In conclusion, our study provided a potential strategy for peritoneal endometriosis diagnosis and treatment.

Methods and materials

Data collection

The RNA sequencing dataset of 76 peritoneal endometriosis tissues and 37 endometrium tissues was fetched from GSE141549. The clinical information all subjects was provided in Supplementary Table 1. Another RNA sequencing dataset that containing 11 peritoneal endometriosis tissues and 11 endometrium tissues was GSE5108. The single cell RNA-seq dataset (ScRNA-Seq) of 8 peritoneal endometriosis tissues was fetched from GSE179640. All the above datasets were downloaded from GEO DataSet. EMT hallmark genes were referred from the HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION gene set in Molecular

Signatures Database v7.5.1 (https://www.gsea-msigdb.org/gsea/msigdb/). The data of msigdb.v7.4.entrez.gmt was downloaded from Gene Set Enrichment Analysis website (https://www.gsea-msigdb.org/gsea/msigdb/).

Gene set enrichment analysis

In order to explore potential mechanisms of EMT in peritoneal endometriosis, we performed gene set enrichment analysis (GSEA) on GSE141549 and GSE5108. Firstly, logFc values of all genes between peritoneal endometriosis tissues and endometrium tissue genes were obtained by"limma" package. Then, GSEA based on msigdb.v7.4.entrez.gmt by "clusterProfiler" package were performed (24). At last, the results was visualized by gseaplot2 of the "enrichplot" package (25).

Consistent cluster analysis based on EMT

To classify peritoneal endometriosis, we performed consistent clustering analysis on GSE141549 based on the 200 EMT hallmark genes by using the "ConsensusClusterPlus" package (26). Samples were divided into two clusters according to the expression characteristics of EMT hallmark genes.

Single cell RNA-seq data analysis

ScRNA-Seq analysis and visualization for GSE179640 were performed with "Seurat" package (version 4.1.1) (27, 28). Briefly, we removed low-quality cells with feature RNA< 500 or > 6000 and mitochondrial reads > 20%. Then, the top 2000 highly variable genes were selected after the gene expression normalization. After gene expression integration, cells were clustered and two-dimensional visualization was performed using uniform manifold approximation and projection (UMAP). Clusters were annotated based on the average gene expression of the following major cell types: fibroblasts (COL1A1, COL3A1, COL1A2), macrophages/monocytes (CD68, MS4A4A, MS4A7, CD14), endothelial cells (PECAM1, VWF), epithelial cells (EPCAM), mesenchymal cells (VIM), CD8+ T cells (PTPRC, CD2, CD3G, CD8A), CD4+ T cells (PTPRC, CD2, CD3G, CD4), dendritic cells (DC) (IL3RA, CLEC4C), mast cells (KIT, TPSB2, TPSAB1), natural killer cells (NK) (NCAM1) and neutrophils (FCGR3A) (29–32).

EMT score calculation

To screen mesenchymal cell marker genes and epithelial marker cell genes for EMT score of peritoneal endometriosis, we firstly referenced 8 epithelial cell marker genes (CD24, CDH1, DSP, EPCAM, FOLR1, KRTI8, KRT19 and OCLN) and 14 mesenchymal cell marker genes (ACTA2, CD44, CDH2, FN1, ITGA5, MMP2, S100A4, SNAI2, TNC, TWIST1, VIM, WNT5A, ZEB1 and ZEB2) of the CellMarker website (http://xteam.xbio.top/CellMarker/). Then, we compared the expression of these genes in epithelial cells cluster and mesenchymal cells cluster (GSE179640). Finally, 8 epithelial genes and 8 mesenchymal genes were selected for EMT score. EMT score was the sum of Z scores of mesenchymal genes minus the sum of Z scores of epithelial genes (33).

Calculation of immune score, stroma score, abundances of immune cells and stroma cells

"xCell" provides a novel method to infer immune and stromal cell types, immune score and stroma score based on genetic characteristics (34). Here, we used the "xCell" package to analyze the relative abundance of immune cells and stroma cells, immune score and stroma score in peritoneal endometriosis samples.

Screening and functional enrichment analysis of differentially expressed genes

In order to figure out the functional differences of the differentially expressed genes (DEGs), the differential genes between cluster 1 and cluster 2 were screened by using of the "limma" package (adj. p. val< 0.05, |log FC| > 1) (35). Then, the DEGs were analyzed by Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) through the website (https://cn.string-db.org/). By setting FDR< 0.05, the significant terms were selected and visualized with the "ggplot2" package (36).

Weighted gene co-expression network analysis

Weighted correlation network analysis (WGCNA) can be used for finding clusters (modules) of highly correlated genes, for summarizing such clusters using the module eigengene or an intramodular hub gene (37). To identify immune score or stroma score associated modules or genes, the "WGCNA" package was used to construct the co-expression network analysis of the mRNA expression matrix of DEGs. samples were clustered according to pearson's correlation analysis and the outliers were removed. The soft thresholding parameter (β) was selected when the scale free topology model fit > 0.85. Afterward, the adjacency matrix was transformed into a

topological overlap matrix (TOM) and genes were assigned to different gene modules according to dissimilarity matrix (1-TOM). Similar dynamic modules were merged when coefficient of dissimilarity< 0.2. Pearson correlation analysis was performed to identify the module with the strongest association with immune score and stroma score. The module eigengenes related to immune score or stroma score were selected with gene significance (GS) > 0.55 and module membership (MM) > 0.85, respectively.

Lasso-logistics regression

We extracted expression matrix of immune score and stroma score related genes from GSE141549. Then, 76 samples in this expression matrix were randomly divided into training dataset and test dataset in a ratio of 1:1. In the training dataset, the Lasso-Logisitic regression analysis was performed based on the classification information of cluster 1 and cluster 2 using the "glmnet" package (38). The diagnostic markers were screen and a diagnostic model was built. Furthermore, the diagnostic model was validated in the test dataset. The ROC curves were plotted using the "ROCR" package and AUC value was calculated (39).

Drug susceptibility analysis

The "pRRophetic" package was used to analyze the half maximal inhibitory concentration (IC50) of 251 drugs (40). Then, the drug candidates for cluster 1 or cluster 2 were screened by setting the adj. p. val<0.05.

Statistics of data

All statistical analyses performed in our study were conducted in R studio (version 4.1.2). Comparisons of mRNA expression were analyzed by Wilcoxon test. All correlation analyses were performed by Pearson correlation analysis using the "corrplot" package (41). Differences were significant when $P < 0.05$.

Results

The classification based on the EMT hallmark genes

Results of GSEA on deferences gene expression between peritoneal endometriosis tissues and endometrium tissues of both GSE141549 and GSE5108 showed that hallmark epithelial-mesenchymal transition (EMT) was the mainly enriched pathway (Figures 1A, B). In order to analysis peritoneal endometriosis from the perspective of EMT, we performed consistent clustering analysis on GSE141549 that containing 76 patients with peritoneal endometriosis based on the EMT hallmark genes. Samples could be clearly divided into cluster 1 (n = 34) and cluster 2 (n = 42) (Figures 1C-E).

EMT score comparison between cluster 1 and cluster 2

In order to calculate the EMT score of peritoneal endometriosis tissues, we performed ScRNA-Seq analysis on GSE179640 for selecting marker genes of epithelial cells and mesenchymal cells. The entire cell population was categorised into 18 major cell clusters. All cell clusters were identified as 11 cell types, consist of fbroblasts cells, macrophages/monocytes, endothelial cells, epithelial cells, other T cells, mesenchymal cells, CD8+ T cells, DC, mast cells, NK/neutrophils and unknown based on expression of markers (Figures 2A, B). We compared the expression of 8 epithelial cell marker genes and 14 mesenchymal cell marker genes in both epithelial cells and mesenchymal cells. Results showed the expression of 8 epithelial cell marker genes (CD24, CDH1, DSP, EPCAM, FOLR1, KRTI8, KRT19 and OCLN) were significantly higher in epithelial cells compared these with mesenchymal cell. And 8 mesenchymal cell marker genes (ACAT2, CD44, FN1, S1004A, TNC, VIM, ZEB1 and ZEB2) were just the oppose (Figure 2C). Hence, we selected these 16 genes as the marker genes for EMT score. Then, EMT score of peritoneal endometriosis (GSE141549) based on the Z score of these marker genes were calculated. Results showed that EMT score of cluster 1 was significantly higher than that of cluster 2 ($p< 0.0001$) (Figure 2D). Results indicated EMT appears much more robust in cluster 1 than that in cluster 2.

Screening and functional enrichment analysis of the differential gene between cluster 1 and cluster 2.

In order to explore the differences between cluster 1 and cluster 2 comprehensively, we analyzed the DEGs between cluster 1 and cluster 2. Results showed there were 95 up-regulated genes and 57 down-regulated genes in cluster 1 compared with cluster 2 (Figures 3A, B). Pathway enrichment indicated that the mainly enriched BP were Regulation of midbrain dopaminergic neuron differentiation and Negative regulation of smooth muscle cell matrix adhesion, the mainly enriched MF were Chemokine activity and CCR chemokine receptor binding, and the mainly CC were mainly Z disc, Stress fiber and Dystrophin-associated glycoprotein complex (Figure 3C). The mainly KEGG-enriched pathways were Cytokine-cytokine receptor interaction, Chemokine signaling, Toll-like receptor signaling pathway and NFkB signaling

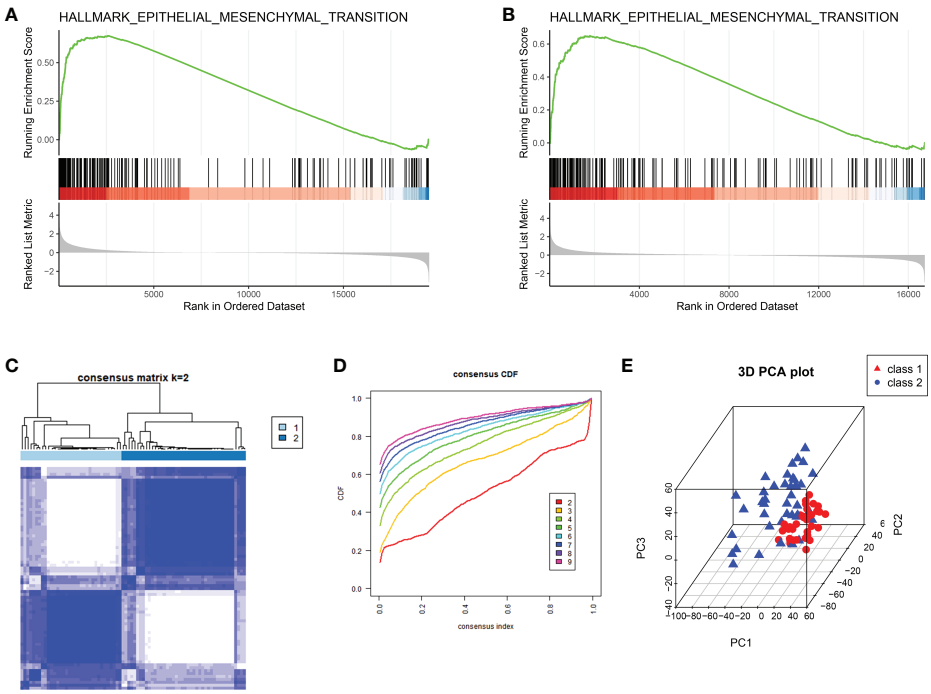


FIGURE 1
The classification of peritoneal endometriosis based on EMT. **(A, B)**. GSEA analysis of the whole transcriptome of GSE141549 and GSE5108, respectively. **(C)**. EMT modification patterns identified with K-means clustering. **(D)**. The cumulative distribution function (CDF) curve of the clustering. **(E)**. PCA plot of cluster1 and cluster2.

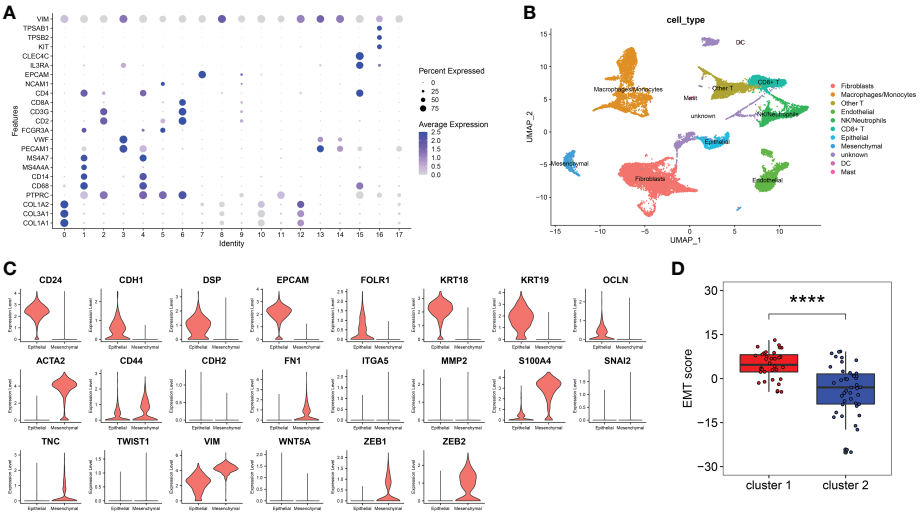


FIGURE 2
EMT score marker genes selection and EMT score calculation. **(A)**. Marker genes expression of 18 clusters were shown on bubble diagram. **(B)**. UMAP plots of 11 types of cells. Cells were colored for types. **(C)**. The expression level of EMT score marker genes in CellMarker website. The upper 8 genes were epithelial cell marker genes and the other 14 genes were mesenchymal cell marker genes. **(D)**. EMT score of cluster 1 and cluster 2. ($p < 0.0001$) (****$p < 0.0001$).

pathway (Figure 3D). Results showed DEGs between cluster 1 and cluster 2 mainly in volved in chemokines signaling pathways, including inflammatory chemokines pathways (Toll-like receptors pathway and NF-kappa B pathway).

Screening of genes related to immune score and stroma score

Given that the functional differences between cluster 1 and cluster 2 were mainly enriched in chemotaxis and inflammatory responses, we further analyzed the immune micro-environment. The immune score of cluster 1 was significantly lower than that of cluster 2 ($p<0.01$), while the stroma score was dramatically higher than that of cluster 2 ($p<0.0001$) (Figure 4A). Furthermore, we selected immune score related gene and stroma score by WGCNA. Four modules were identified when the Diss Thres was set as 0.2 after merging dynamic modules, as shown in the clustering dendrograms (Figure 4B). The brown module and turquoise module were associated with Immune score and stroma score respectively (Figure 4C). Finally, 7 Immune score-related genes were set selected by setting GS>0.55 and MM>0.85 (Figure 4D). Results showed that the expression of all 7 Immune score-related genes in cluster 1 were significantly lower than those in cluster 2 ($p<0.0001$) (Figure 4E). All these 7 genes were significantly positively correlated with Immune score ($p<0.0001$) (Figure 4F). Similarly, 14 stroma score-related genes were selected and the expression of these 14 genes in cluster 1 were remarkably higher than those in cluster 2 ($p<0.05$) (Figures 4G, H). All 14-stroma score-related genes were significantly positively correlated with Immune score ($p<0.05$) (Figure 4I). In conclusion, the immune cells infiltration of cluster 1 was significantly higher than that of cluster 2, while the infiltration of stroma cells was remarkably lower in cluster 2. We speculated that, in peritoneal endometriosis lesions, high infiltration of immune cells inhibited the progression of EMT, while high infiltration of stroma cell contributes to EMT.

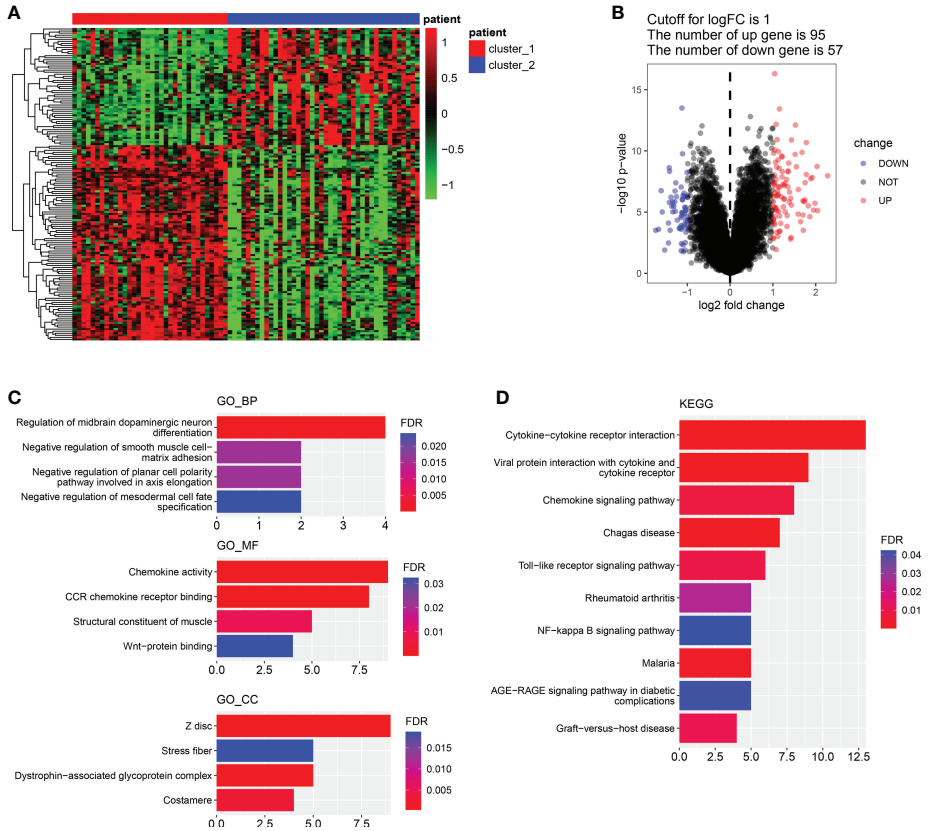


FIGURE 3
Differential genes screening and functional enrichment analysis. **(A, B)**. Heatmap of DEGs between cluster 1 and cluster 2. **(B)**. Volcano plot of DEGs between cluster 1 and cluster 2. **(C)**. GO enrichment analysis of DEGs about BP, MF and CC. **(D)**. KEGG-enriched analysis. (BP, biological process. MF, molecular function).

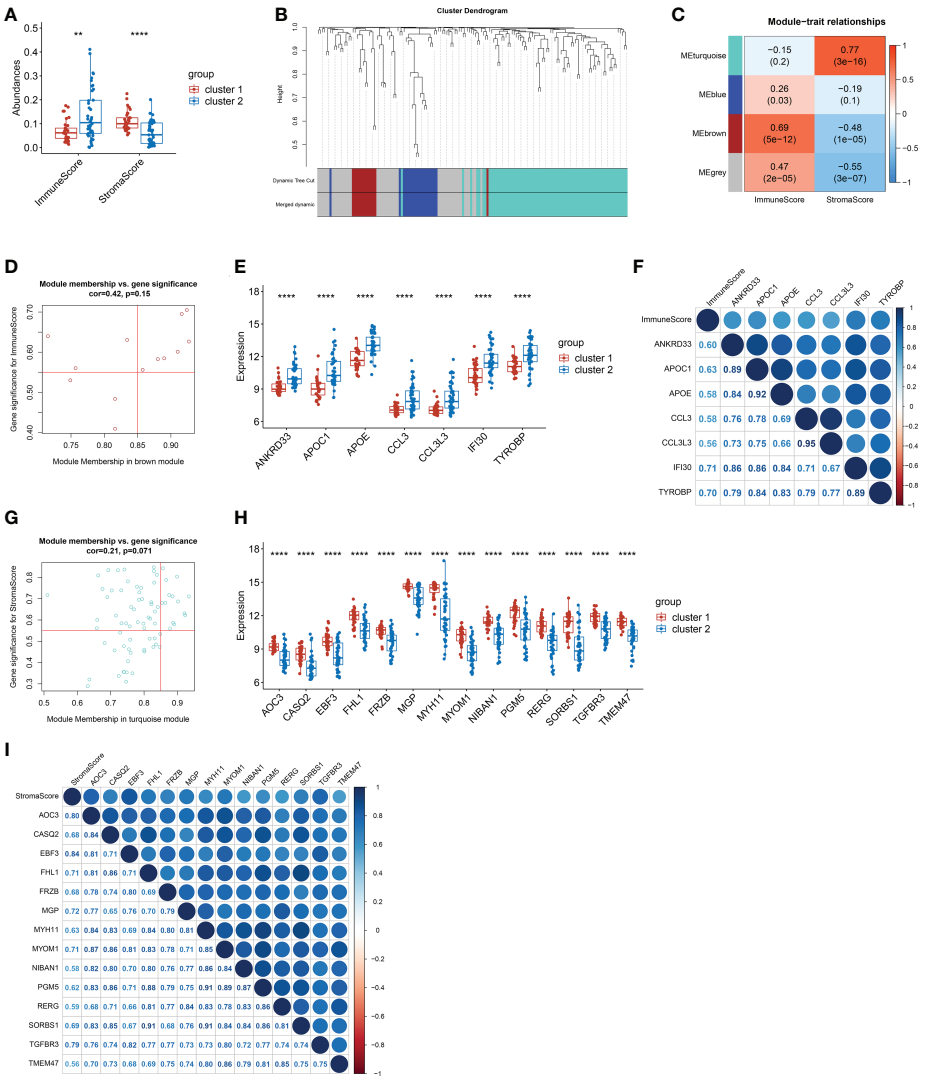


FIGURE 4
The screening of genes that related to Immune score and stroma score. **(A)**. The immune score and stroma score of cluster and cluster 2. **(B)**. Cluster dendrogram of the co-expression network modules. **(C)**. Correlations between the modules and immune scores, and correlations between the modules and immune scores (p-values were shown). **(D)**. Scatter plot analysis of the brown module. 7 Immune score-related genes were screened out in the upper-right area where GS > 0.55 and MM > 0.85. **(E)**. Comparison of the 7 immune score-related genes between cluster 1 and cluster 2. **(F)**. Correlation analysis between the 7 immune score-related genes. **(G)**. Scatter plot analysis of the turquoise module. 14 stroma score-related genes were screened out in the upper-right area where GS > 0.55 and MM > 0.85. **(H)**. Comparison of the 14 stroma score-related genes between cluster 1 and cluster 2. **(I)**. Correlation analysis between the 14 stroma score-related genes. (**p< 0.01; ****p< 0.0001; GS, gene significance. MM, module membership.).

The abundances of immune cells and stroma cells

Given the significant differences in immune score and stroma score between the two clusters, we further analyzed the abundances of immune cells and stroma cells. Results showed that the abundances of 12 kinds of immune cells, namely DC cells, iDC cells, Monocytes, Macrophages, M1 Macrophages, M2 Macrophages, Basophils, Th1 cells, Th2 cells, CD4+ Tem cells, B cells and memory B cells,were significantly lower in cluster 1 than that of cluster 2 (p< 0.05) (Figure 5A). Correlation analysis showed the abundances of above 12 kinds of immune cells were almost remarkably positively correlated with the expression of all 7 immune score-related genes (p< 0.05) (Figure 5B). This indicated that immune cells in cluster 2 were more active than those in cluster 1. And immune cells were positively regulated by

immune score-related genes. Additionally, the abundance of Epithelial cells, Keratinocytes and Osteoblasts in cluster 1 were significantly lower than those in cluster 2 ($p<0.05$) (Figure 5C) and had significantly negative correlations with the stroma-related genes (Figure 5D). While the abundances of Fibroblasts, ly Endothelial cells, Myocytes, Chondrocytes and Skeletal muscle cells were significantly higher in cluster 1 than that of cluster 2 ($p<0.05$) (Figure 5C) had significantly positive correlations with the whole stroma-related genes (Figure 5D). In addition, abundances of Adipocytes and Smooth muscle cells were also higher in cluster 1. The epithelial cell abundance of cluster 1 was lower, which consistent with the EMT score (Figure 2D). It was suggested that the increased abundance of Fibroblasts, ly Endothelial cells, Skeletal muscle cells and Smooth muscle cells probably contribute to EMT in peritoneal endometriosis.

Construction of the diagnostic model

To construct a diagnostic model, diagnostic markers were screened from immune score- related genes and stroma score-related genes by lasso-logistic regression analysis in the training dataset. The minimum binomial deviance was obtained when log(λ) was -5.773583, and 9 genes were selected as diagnostic markers (Figure 6A). The coefficients of TMEM47 and FRZB were larger than the other 7 genes (Figure 6B) (Supplementary Table 2). A diagnostic model was constructed with the following formula:

$$cd \quad score = \sum_{i=1}^{n}(Coefficient_i \times Expression_i)$$

The ROC analysis showed that the AUC of the training dataset was 0.955 when the cut-off value of the cd-score was -36.070 (Figure 6C). Sample was classified as cluster 1 when the cd-score was less than or equal to the cut-off value, otherwise sample was classified as cluster 2. According to the cut-off of the training dataset, the AUC of the test dataset was 0.862 (Figure 6D). Additionally, cd score was significantly negatively correlated with EMT score in training dataset, test dataset and entire dataset (Figure 6E-G). Therefore, the diagnostic model constructed from these 9 genes and their coefficients had high specificity and sensitivity.

Candidate drug screening

Based on the clusters classified by EMT hallmark genes, drug susceptibility was analyzed. In the training dataset, the IC50 of BMS-754807 and Lisitinib in cluster 1 was significantly lower than that in cluster 2 ($p<0.05$), while the IC50 of Methotrexate, Gefitinib, Veliparib, GW 4441756, CCT007093 and Temozolomide in cluster 1 were remarkably higher in cluster 2 ($p<0.0001$) (Figure 7A). The drug susceptibility trends of all candidate drugs in the test dataset were consistent with that in the training dataset (Figure 7B). Then, we classified the dataset into cluster 1 and cluster 2 by the diagnostic model we established. Except for GW 441756, the susceptibility trends of all candidate drugs in the test dataset predicted by the above

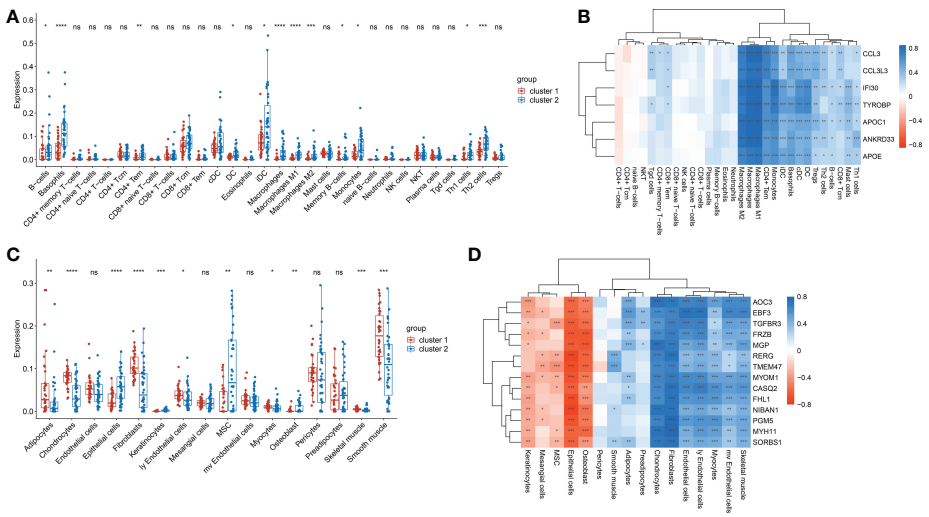


FIGURE 5
The abundances of immune cells and stroma cells in cluste1 and cluster2. **(A)**. Comparison of the immune cell abundances in cluster 1 and cluster2. **(B)**. Correlation of the immune cell abundances and the expression of immune-related genes. **(C)**. The comparison of the stroma cells abundances in cluster 1 and cluster2. **(D)**. Correlation of the stroma cells abundances and the expression of stroma score-related genes. (*p<0.05; **p<0.01; ***p<0.001; ****p<0.0001; ns, no significance).

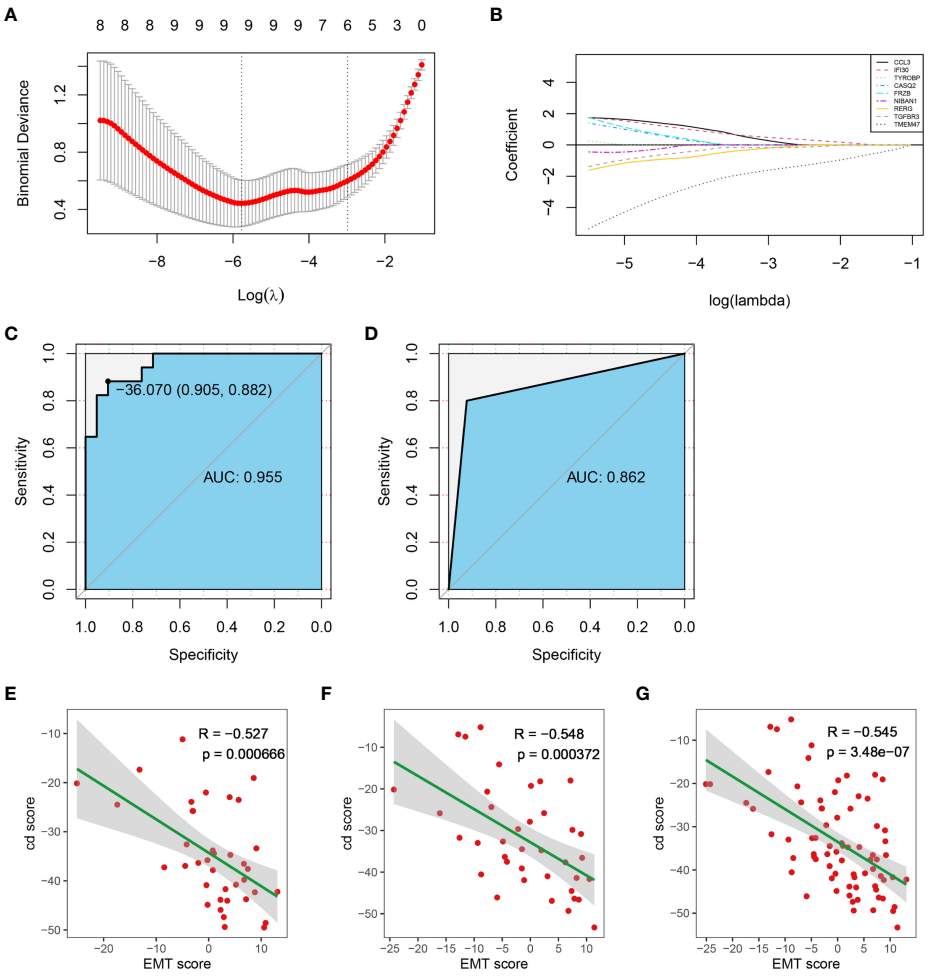


FIGURE 6
Construction of the diagnostic model. **(A)** Diagnostic markers screening. **(B)** The coefficients of all diagnostic markers. **(C)** ROC of the diagnostic model with the 14 diagnostic markers in the training dataset. **(D)** Validation of the diagnostic model in the test dataset. **(E–G)**. The correlation between cd score and EMT score in test dataset, training dataset and entire dataset, respectively. (AUC, Area Under Curve).

diagnostic model were also consistent with the training dataset. (Figure 7C). Results showed BMS-754807 and Lisitinib were more sensitive for cluster 1, while Methotrexate, Gefitinib, Veliparib, CCT007093 and Temozolomide were more sensitive for cluster 2. It was suggested that the diagnostic classification models we established can be used for drug screening.

Discussion

Over decades, endometriosis classified traditionally based on lesion appearance, pelvic adhesions, or/and anatomic location of disease (42), but none of the current classification systems classify peritoneal endometriosis from molecular perspective. Here, we classified peritoneal endometriosis into two cluster based on EMT hallmark genes and found EMT scores of cluster 1 was significantly higher than cluster 2. What was more, we also found EMT in peritoneal endometriosis was related with both immune cell infiltration and stroma cell infiltration. In addition, based on immune score-related genes and stroma score-related genes, we established a diagnostic model and screened candidate drugs. Our study provided new ideas for classification, diagnosis and treatment of peritoneal endometriosis.

EMT is involved in the process of endometriosis. The migration and invasion abilities of endometrial stromal cells enhanced by facilitated EMT, and conversely inhibited EMT-related proteins reduced the volume and weight of endometriotic lesions in mice model (43–45). In pathological and physiological

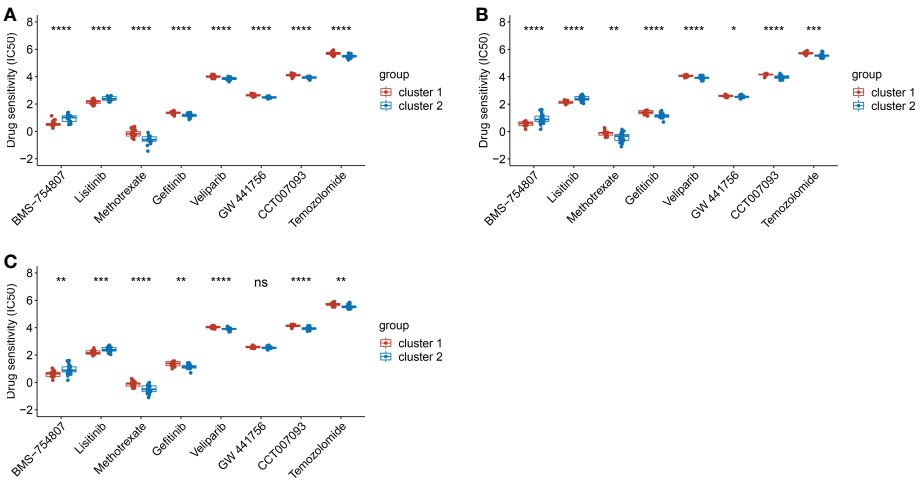


FIGURE 7
The comparison of drug sensitivity between cluster 1 and cluster 2. **(A)**. The comparison of drug sensitivity in training dataset. **(B)**. The comparison of drug sensitivity in test dataset. **(C)**. The comparison of drug sensitivity in the predicted cluster 1 and cluster 2 in test dataset. (*p< 0.05; **p< 0.01; ***p< 0.001; ****p< 0.0001).

EMT, both stroma cells and immune cells are involved (46–48). Researches concerning stroma cell involve in EMT are not rare. Adipocytes promote EMT progression by reducing epithelial cell characteristics or inducing EMT-related phenotypes and thus promote tumor invasiveness (49, 50). Ly endothelial cells mediate the preferential migration of cells that undergoing EMT to lymphatic vessels by secreted pro-inflammatory cytokines (51). Chemokines promote pulmonary fibrosis by promoting EMT (52). EMT induced tissue fibrosis, which probably stimulate the production of fibroblasts in (53). Here we found not only the stroma score but also the abundances of most infiltrating stroma cells were significantly higher in cluster 1 than these in cluster 2, including fibroblasts, adipocytes, ly endothelial cells, chondrocytes, skeletal muscle cells and smooth muscle cells. We proposed that the infiltration of stroma cells probably contribute to EMT in peritoneal endometriosis. Besides, T and B cells, DC cells and tumor-associated macrophages that present in the tumor micro-environment induce EMT (54). Macrophages may induce pathological EMT of epithelial cells in a denomyosis (55). EMT is strongly associated with a highly immunosuppressive environment (15). We found the immune score was significantly lower in cluster 1 in than that in cluster 2, while the abundances of all infiltrating immune cells were significantly higher in cluster 2 than that in cluster 1, particularly macrophages, DC cells, CD4 +T cells and B cells. Here, we proposed immune cell infiltration possibly inhibited the EMT of peritoneal endometriosis, especially macrophages, DC cells, CD4+T cells and B cells. Therefore, EMT classification is meaningful for peritoneal endometriosis accurate diagnosis and treatment.

Additionally, stroma score- and immune score-related genes possibably participate in stromal cells and immune cells infiltration. Aoc3 is an endothelial adhesion molecule that contributes to the extravasation of neutrophils, macrophages, and lymphocytes to sites of inflammation (56). CASQ2 is a calcium binding protein that stores calcium for muscle function (57). FRZB is involved in the regulation of chondrocytes development (58). MGP is a vitamin K-dependent protein, which is synthesized in bone and many other mesenchymal cells, which is also highly expressed by vascular smooth muscle cells (VSMCs) and chondrocytes (59). CCL3 and CCL3L3 are chemokines that produced by macrophage and monocyte respectively (60, 61). Ifi30 is an IFN-γ-inducible protein that is involved in MHC class II-restricted antigen processing and MHC class I-restricted cross-presentation pathways of adaptive immunity (62). Therefore, it was suggested that these genes regulate stroma cells and immune cellsinfiltration in peritoneal endometrisis.

To date, drugs treatment for endometriosis are mainly based on hormone regulation and inflammation inhibition, rarely concerning EMT. Here, based on EMT classification, we selected 2 candidate drugs for cluster 1 and 6 candidate drugs for cluster 2. As for cluster 2 drugs, Methotrexate blocks tumor cell proliferation mainly through the inhibition of dihydrofolate reductase (DHFR), which is also an immunosuppression (63). Gefitinib is a small molecule inhibitor of epidermal growth factor receptor (EGFR) tyrosine kinase (64). Veliparib is an inhibitor of PARP1 and PARP2 (65). GW 4441756 is a selective TrkA (NTRK1) inhibitor. CCT007093 is an inhibitor of protein phosphatase 1D (PPM1D Wip1) (66). Temozolomide reduces

the proliferative activity of tumor cells (67). Pathway enrichment analysis found that drugs for cluster 2 mainly acted on the EGFR signaling pathway (Supplementary Figure 1). And restraining EGFR pathway can inhibit EMT progression (68, 69). Among drugs for cluster 1, BMS-754807 is a potent small molecule inhibitor of IGF-1R/IR family kinases. Lisitinib is a dual inhibitor of IGF-1 and insulin receptor (IR) (70). IGF-1 is expressed in ectopic endometrial stroma cells (71). In addition, IGF-1 concentration in peritoneal fluid of patients with endometriosis are significantly higher than that of normal controls (72, 73). On the other hand, the peritoneal mesothelial cells with insufficient IGF-1R expression had lower migration ability and higher adhesion ability (74). In addition, inhibitors of IGF-1R hinder the growth of ectopic lesions and reverses the pain behavior in mice model (71, 73). It was indicated that inhibition of insulin-like growth factor pathway was crucial for the treatment for cluster 1. Of course, drugs we screened needed to be further validated.

In conclusion, we classified peritoneal endometriosis based on EMT. Then, we constructed diagnostic models based on the screened genes and performed drug screening. This will provide a new strategy for the precise diagnosis and medicine of peritoneal endometriosis.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE141549.

Ethics statement

Ethical review and approval was not required for the animal study because Our study is based on sequencing data downloaded from the GEO database.

Author contributions

JT and JW collected the research data and checked the data analysis. MY directed data analysis. QQ analyzed the data and wrote the draft. All authors contributed to the article and approved the submitted version.

Funding

We acknowledge the PhD workstation of Guangdong Provincial Reproductive Science Institute (Guangdong Provincial Fertility Hospital) for funding support (NO.: BS202201).

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/fendo.2022.1035158/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Pathway enrichment of targets of cluster 2 drugs.

SUPPLEMENTARY TABLE 1
Clinical information of the participants.

SUPPLEMENTARY TABLE 2
Gene coef.

SUPPLEMENTARY DATA SHEET 1
R script.

References

1. Vercellini P, Viganò P, Somigliana E, Fedele L. Endometriosis: pathogenesis and treatment. *Nat Rev Endocrinol* (2014) 10(5):261–75. doi: 10.1038/nrendo.2013.255

2. Lac V, Verhoef L, Aguirre-Hernandez R, Nazeran TM, Tessier-Cloutier B, Praetorius T, et al. Iatrogenic endometriosis harbors somatic cancer-driver mutations. *Hum Reprod* (2019) 34(1):69–78. doi: 10.1093/humrep/dey332

3. Ramírez-Pavez TN, Martínez-Esparza M, Ruiz-Alcaraz AJ, García-Peñarrubia P, Machado-Linde F.García-Peñarrubia P. The role of peritoneal macrophages in endometriosis. *Int J Mol Sci* (2021) 22(19):1–19. doi: 10.3390/ijms221910792

4. Cuevas M, Flores I, Thompson KJ, Ramos-Ortolaza DL, Torres-Reveron A, Appleyard CB. Stress exacerbates endometriosis manifestations and inflammatory

parameters in an animal model. *Reprod Sci* (2012) 19(8):851–62. doi: 10.1177/1933719112438443

5. Pfeifer SM, Attaran M, Goldstein J, Lindheim SR, Petrozza JC, Rackow BW, et al. ASRM müllerian anomalies classification. *Fertil Steril* (2021) 116(5):1238–52. doi: 10.1016/j.fertnstert.2021.09.025

6. Keckstein J, Hudelist G. Classification of deep endometriosis (DE) including bowel endometriosis: From r-ASRM to #Enzian-classification. *Best Pract Res Cl Ob* (2021) 71:27–37. doi: 10.1016/j.bpobgyn.2020.11.004

7. Wrighton KH. Cell migration: EMT promotes contact inhibition of locomotion. *Nat Rev Mol Cell Bio* (2015) 16(9):518. doi: 10.1038/nrm4045

8. Dongre A, Weinberg RA. New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. *Nat Rev Mol Cell Bio* (2019) 20(2):69–84. doi: 10.1038/s41580-018-0080-4

9. Liu X, Zhang Q, Guo SW. Histological and immunohistochemical characterization of the similarity and difference between ovarian endometriomas and deep infiltrating endometriosis. *Reprod Sci* (2018) 25(3):329–40. doi: 10.1177/1933719117718275

10. Matsuzaki S, Darcha C, Pouly JL, Canis M. Effects of matrix stiffness on epithelial to mesenchymal transition-like processes of endometrial epithelial cells: Implications for the pathogenesis of endometriosis. *Sci Rep* (2017) 7:44616. doi: 10.1038/srep44616

11. Furuya M, Masuda H, Hara K, Uchida H, Sato K, Sato S, et al. ZEB1 expression is a potential indicator of invasive endometriosis. *Acta Obstet Gyn Scan* (2017) 96(9):1128–35. doi: 10.1111/aogs.13179

12. Caramel J, Papadogeorgakis E, Hill L, Browne GJ, Richard G, Wierinckx A, et al. A switch in the expression of embryonic EMT-inducers drives the development of malignant melanoma. *Cancer Cell* (2013) 24(4):466–80. doi: 10.1016/j.ccr.2013.08.018

13. Lüönd F, Sugiyama N, Bill R, Bornes L, Hager C, Tang F, et al. Distinct contributions of partial and full EMT to breast cancer malignancy. *Dev Cell* (2021) 56(23):3203–21.e11. doi: 10.1016/j.devcel.2021.11.006

14. Bakir B, Chiarella AM, Pitarresi JR, Rustgi AK. EMT MET plasticity and tumor metastasis. *Trends Cell Biol* (2020) 30(10):764–76. doi: 10.1016/j.tcb.2020.07.003

15. Cai J, Cui Y, Yang J, Wang S. Epithelial-mesenchymal transition: When tumor cells meet myeloid-derived suppressor cells. *Bba-Rev Cancer* (2021) 1876 (1):188564. doi: 10.1016/j.bbcan.2021.188564

16. Peyre L, Meyer M, Hofman P, Roux J. TRAIL receptor-induced features of epithelial-to-mesenchymal transition increase tumour phenotypic heterogeneity: potential cell survival mechanisms. *Brit J Cancer* (2021) 124(1):91–101. doi: 10.1038/s41416-020-01177-w

17. Lousse JC, Van Langendonckt A, Defrere S, Ramos RG, Colette S, Donnez J. Peritoneal endometriosis is an inflammatory disease. *Front Biosci (Elite Ed)* (2012) 4(1):23–40. doi: 10.2741/e358

18. Gardella B, Dominoni M, Gritti A, Arrigo A, Antonucci S, Carletti GV, et al. Endometriosis pain and epithelial neutrophil activating peptide-78 levels. *Sci Rep* (2022) 12(1):3227. doi: 10.1038/s41598-022-07349-3

19. Kuwada K, Kagawa S, Yoshida R, Sakamoto S, Ito A, Watanabe M, et al. The epithelial-to-mesenchymal transition induced by tumor-associated macrophages confers chemoresistance in peritoneally disseminated pancreatic cancer. *J Exp Clin Cancer Res* (2018) 37(1):307. doi: 10.1186/s13046-018-0981-2

20. Kusama K, Fukushima Y, Yoshida K, Sakakibara H, Tsubata N, Yoshie M, et al. Endometrial epithelial-mesenchymal transition (EMT) by menstruation-related inflammatory factors during hypoxia. *Mol Hum Reprod* (2021) 27(6):1–11. doi: 10.1093/molehr/gaab036

21. Ibrahim MG, Sillem M, Plendl J, Taube ET, Schüring A, Götte M, et al. Arrangement of myofibroblastic and smooth muscle-like cells in superficial peritoneal endometriosis and a possible role of transforming growth factor beta 1 (TGFβ1) in myofibroblastic metaplasia. *Arch Gynecol Obstet* (2019) 299(2):489–99. doi: 10.1007/s00404-018-4995-y

22. Jiang Z, Luo Y, Zhang L, Li H, Pan C, Yang H, et al. A novel risk score model of lactate metabolism for predicting over survival and immune signature in lung adenocarcinoma. *Cancers (Basel)* (2022) 14(15):3727. doi: 10.3390/cancers14153727

23. Qiu C, Shi W, Wu H, Zou S, Li J, Wang D, et al. Identification of molecular subtypes and a prognostic signature based on inflammation-related genes in colon adenocarcinoma. *Front Immunol* (2021) 12:769685. doi: 10.3389/fimmu.2021.769685

24. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an r package for comparing biological themes among gene clusters. *OMICS* (2012) 16(5):284–7. doi: 10.1089/omi.2011.0118

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

26. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* (2010) 26 (12):1572–3. doi: 10.1093/bioinformatics/btq170

27. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell* (2021) 184(13):3573–3587.e29. doi: 10.1016/j.cell.2021.04.048

28. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive integration of single-cell data. *Cell* (2019) 177(7):1888–1902.e21. doi: 10.1016/j.cell.2019.05.031

29. Tan Y, Flynn WF, Sivajothi S, Luo D, Bozal SB, Davé M, et al. Single-cell analysis of endometriosis reveals a coordinated transcriptional programme driving immunotolerance and angiogenesis across eutopic and ectopic tissues. *Nat Cell Biol* (2022) 24(8):1306–18. doi: 10.1038/s41556-022-00961-5

30. Ma J, Zhang L, Zhan H, Mo Y, Ren Z, Shao A, et al. Single-cell transcriptomic analysis of endometriosis provides insights into fibroblast fates and immune cell heterogeneity. *Cell Biosci* (2021) 11(1):125. doi: 10.1186/s13578-021-00637-x

31. Shih AJ, Adelson RP, Vashistha H, Khalili H, Nayyar A, Puran R, et al. Single-cell analysis of menstrual endometrial tissues defines phenotypes associated with endometriosis. *BMC Med* (2022) 20(1):315. doi: 10.1186/s12916-022-02500-3

32. Konrad L, Dietze R, Riaz MA, Scheiner-Bobis G, Behnke J, Horné F, et al. Epithelial-mesenchymal transition in endometriosis-when does it happen? *J Clin Med* (2020) 9(6):1915. doi: 10.3390/jcm9061915

33. Thompson JC, Hwang WT, Davis C, Deshpande C, Jeffries S, Rajpurohit Y, et al. Gene signatures of tumor inflammation and epithelial-to-mesenchymal transition (EMT) predict responses to immune checkpoint blockade in lung cancer with high accuracy. *Lung Cancer* (2020) 139:1–8. doi: 10.1016/j.lungcan.2019.10.012

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

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

36. Ito K, Murphy D. Application of ggplot2 to pharmacometric graphics. *CPT Pharmacometrics Syst Pharmacol* (2013) 2(10):e79. doi: 10.1038/psp.2013.56

37. Langfelder P, Horvath S. WGCNA: an r package for weighted correlation network analysis. *BMC Bioinf* (2008) 9:559. doi: 10.1186/1471-2105-9-559

38. Trevor H, Robert T, Ryan T. Best subset forward stepwise or lasso? glmnet: Analysis and recommendations based on extensive comparisons. *Stat Sc* (2020) 35 (4):579–92. doi: 10.1214/19-STS733

39. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in r. *Bioinformatics* (2005) 21(20):3940–1. doi: 10.1093/bioinformatics/bti623

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

41. Michael F. Corrgrams: Exploratory displays for correlation matrices. *Am Statistician* (2002) 256:316–24. doi: 10.1198/000313002533

42. Johnson NP, Hummelshoj L, Adamson GD, Keckstein J, Taylor HS, Abrao MS, et al. World endometriosis society consensus on the classification of endometriosis. *Hum Reprod* (2017) 32(2):315–24. doi: 10.1093/humrep/dew293

43. Xiong W, Zhang L, Liu H, Li N, Du Y, He H, et al. E2 -mediated EMT by activation of β-catenin/Snail signalling during the development of ovarian endometriosis. *J Cell Mol Med* (2019) 23(12):8035–45. doi: 10.1111/jcmm.14668

44. Zheng QM, Chen XY, Bao QF, Yu J, Chen LH. ILK enhances migration and invasion abilities of human endometrial stromal cells by facilitating the epithelial-mesenchymal transition. *Gynecol Endocrinol* (2018) 34(12):1091–6. doi: 10.1080/09513590.2018.1498477

45. Hsu YW, Chen HY, Chiang YF, Chang LC, Lin PH, Hsia SM, et al. The effects of isoliquiritigenin on endometriosis *in vivo* and *in vitro* study. *Phytomedicine* (2020) 77:153214. doi: 10.1016/j.phymed.2020.153214

46. Ricciardi M, Zanotto M, Malpeli G, Bassi G, Perbellini O, Chilosi M, et al. Epithelial-to-mesenchymal transition (EMT) induced by inflammatory priming elicits mesenchymal stromal cell-like immune-modulatory properties in cancer cells. *Brit J Cancer* (2015) 112(6):1067–75. doi: 10.1038/bjc.2015.29

47. Guo S, Deng CX. Effect of stromal cells in tumor microenvironment on metastasis initiation. *Int J Biol Sci* (2018) 14(14):2083–93. doi: 10.7150/ijbs.25720

48. Han L, Wang S, Wei C, Fang Y, Huang S, Yin T, et al. Tumour microenvironment: a non-negligible driver for epithelial-mesenchymal transition in colorectal cancer. *Expert Rev Mol Med* (2021) 23:e16. doi: 10.1017/erm.2021.13

49. Kushiro K, Chu RA, Verma A, Núñez NP. Adipocytes promote B16BL6 melanoma cell invasion and the epithelial-to-Mesenchymal transition. *Cancer Microenviron* (2012) 5(1):73–82. doi: 10.1007/s12307-011-0087-2

50. Lee Y, Jung WH, Koo JS. Adipocytes can induce epithelial-mesenchymal transition in breast cancer cells. *Breast Cancer Res Tr* (2015) 153(2):323–35. doi: 10.1007/s10549-015-3550-9

51. Pang MF, Georgoudaki AM, Lambut L, Johansson J, Tabor V, Hagikura K, et al. TGF-β1-induced EMT promotes targeted migration of breast cancer cells through the lymphatic system by the activation of CCR7/CCL21-mediated chemotaxis. *Oncogene* (2016) 35(6):748–60. doi: 10.1038/onc.2015.133

52. Ma Z, Ma C, Zhang Q, Bai Y, Mu K, Liu X, et al. Role of CXCL16 in BLM-induced epithelial-mesenchymal transition in human A549 cells. *Respir Res* (2021) 22(1):42. doi: 10.1186/s12931-021-01646-7

53. Zeisberg M, Duffield JS. Resolved: EMT produces fibroblasts in the kidney. *J Am Soc Nephrol* (2010) 21(8):1247–53. doi: 10.1681/ASN.2010060616

54. Chattopadhyay I, Ambati R, Gundamaraju R. Exploring the crosstalk between inflammation and epithelial-mesenchymal transition in cancer. *Mediat Inflamm* (2021) 2021:9918379. doi: 10.1155/2021/9918379

55. An M, Li D, Yuan M, Li Q, Zhang L, Wang G, et al. Different macrophages equally induce EMT in endometria of adenomyosis and normal. *Reproduction* (2017) 154(1):79–92. doi: 10.1530/REP-17-0174

56. Dunkel J, Aguilar-Pimentel JA, Ollert M, Fuchs H, Gailus-Durner V, de Angelis MH, et al. Endothelial amine oxidase AOC3 transiently contributes to adaptive immune responses in the airways. *Eur J Immunol* (2014) 44(11):3232–9. doi: 10.1002/eji.201444563

57. Rossi D, Gamberucci A, Pierantozzi E, Amato C, Migliore L, Sorrentino V. Calsequestrin, a key protein in striated muscle health and disease. *J Muscle Res Cell M* (2021) 42(2):267–79. doi: 10.1007/s10974-020-09583-6

58. Yagi H, Takahata Y, Murakami T, Nakaminami Y, Hagino H, Yamamoto S, et al. Transcriptional regulation of FRZB in chondrocytes by osterix and Msx2. *J Bone Miner Metab* (2022) 40(5):723–34. doi: 10.1007/s00774-022-01345-3

59. Bjørklund G, Svanberg E, Dadar M, Card DJ, Chirumbolo S, Harrington DJ, et al. The role of matrix gla protein (MGP) in vascular calcification. *Curr Med Chem* (2020) 27(10):1647–60. doi: 10.2174/0929867325666180716104159

60. Sheng D, Ma W, Zhang R, Zhou L, Deng Q, Tu J, et al. Ccl3 enhances docetaxel chemosensitivity in breast cancer by triggering proinflammatory macrophage polarization. *J Immunother Cancer* (2022) 10(5):e003793. doi: 10.1136/jitc-2021-003793

61. Lara S, Akula S, Fu Z, Olsson AK, Kleinau S, Hellman L, et al. The human monocyte-a circulating sensor of infection and a potent and rapid inducer of inflammation. *Int J Mol Sci* (2022) 23(7):3890. doi: 10.3390/ijms23073890

62. Guo J, Zhou M, Liu X, Pan Y, Yang R, Zhao Z, et al. Porcine IFI30 inhibits PRRSV proliferation and host cell apoptosis *in vitro*. *Gene* (2018) 649:93–8. doi: 10.1016/j.gene.2018.01.065

63. Zhao S, Wang ZP, Lin Z, Wei G, He Y, Wen X, et al. Drug repurposing by siderophore conjugation: Synthesis and biological evaluation of siderophore-methotrexate conjugates as antibiotics. *Angew Chem Int Edit* (2022) 61(36): e202204139. doi: 10.1002/anie.202204139

64. Maemondo M, Inoue A, Kobayashi K, Sugawara S, Oizumi S, Isobe H, et al. Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. *New Engl J Med* (2010) 362(25):2380–8. doi: 10.1056/NEJMoa0909530

65. Murai J, Huang SY, Das BB, Renaud A, Zhang Y, Doroshow JH, et al. Trapping of PARP1 and PARP2 by clinical PARP inhibitors. *Cancer Res* (2012) 72 (21):5588–99. doi: 10.1158/0008-5472.CAN-12-2753

66. Zhang L, Liu L, He Z, Li G, Liu J, Song Z, et al. Inhibition of wild-type p53-induced phosphatase 1 promotes liver regeneration in mice by direct activation of mammalian target of rapamycin. *Hepatology* (2015) 61(6):2030–41. doi: 10.1002/hep.27755

67. Willis JA, Overman MJ. Inducing hypermutability to promote anti-PD-1 therapy response. *Cancer Discovery* (2022) 12(7):1612–14. doi: 10.1158/2159-8290.CD-22-0492

68. Yuan W, Wei F, Ouyang H, Ren X, Hang J, Mo X, et al. CMTM3 suppresses chordoma progress through EGFR/STAT3 regulated EMT and TP53 signaling pathway. *Cancer Cell Int* (2021) 21(1):510. doi: 10.1186/s12935-021-02159-5

69. Sheng W, Chen C, Dong M, Wang G, Zhou J, Song H, et al. Calreticulin promotes EGF-induced EMT in pancreatic cancer cells *via* Integrin/EGFR-ERK/MAPK signaling pathway. *Cell Death Dis* (2017) 8(10):e3147. doi: 10.1038/cddis.2017.547

70. Wang Q, Zhang Y, Zhu J, Zheng H, Chen S, Chen L, et al. IGF-1R inhibition induces MEK phosphorylation to promote survival in colon carcinomas. *Signal Transduct Tar* (2020) 5(1):153. doi: 10.1038/s41392-020-0113-2

71. Zhou Y, Zeng C, Li X, Wu PL, Yin L, Yu XL, et al. IGF-I stimulates ERβ and aromatase expression *via* IGF1R/PI3K/AKT-mediated transcriptional activation in endometriosis. *J Mol Med* (2016) 94(8):887–97. doi: 10.1007/s00109-016-1396-1

72. Heidari S, Kolahdouz-Mohammadi R, Khodaverdi S, Tajik N, Delbandi AA. Expression levels of MCP-1 HGF and IGF-1 in endometriotic patients compared with non-endometriotic controls. *BMC Womens Health* (2021) 21(1):422. doi: 10.1186/s12905-021-01560-6

73. Forster R, Sarginson A, Velichkova A, Hogg C, Dorning A, Horne AW, et al. Macrophage-derived insulin-like growth factor-1 is a key neurotrophic and nerve-sensitizing factor in pain associated with endometriosis. *FASEB J* (2019) 33 (10):11210–22. doi: 10.1096/fj.201900797R

74. Xia Y, Wan C, Zhang Q, Wang H, Feng Y, Jiang C, et al. Role of IGF-1R in epithelial-mesenchymal trans differentiation of human peritoneal mesothelial cells. *Clin Exp Nephrol* (2022) 26(7):630–9. doi: 10.1007/s10157-022-02209-w

frontiers | Frontiers in Endocrinology

TYPE Original Research
PUBLISHED 02 December 2022
DOI 10.3389/fendo.2022.1015557

OPEN ACCESS

EDITED BY
Katsumi Iizuka,
Fujita Health University, Japan

REVIEWED BY
Sergio Polakof,
l'alimentation et l'environnement (INRAE), France
Takaaki Murakami,
Kyoto University, Japan

*CORRESPONDENCE
Oluf Pedersen
oluf@sund.ku.dk

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

SPECIALTY SECTION
This article was submitted to Systems Endocrinology, a section of the journal Frontiers in Endocrinology

RECEIVED 09 August 2022
ACCEPTED 11 November 2022
PUBLISHED 02 December 2022

CITATION
Clos-Garcia M, Ahluwalia TS, Winther SA, Henriksen P, Ali M, Fan Y, Stankevic E, Lyu L, Vogt JK, Hansen T, Legido-Quigley C, Rossing P and Pedersen O (2022) Multiomics signatures of type 1 diabetes with and without albuminuria. *Front. Endocrinol.* 13:1015557. doi: 10.3389/fendo.2022.1015557

Multiomics signatures of type 1 diabetes with and without albuminuria

Marc Clos-Garcia[1,2†], Tarunveer S. Ahluwalia[3,4†], Signe A. Winther[3†], Peter Henriksen[3], Mina Ali[3], Yong Fan[1], Evelina Stankevic[1], Liwei Lyu[1], Josef K. Vogt[1,5], Torben Hansen[1], Cristina Legido-Quigley[3], Peter Rossing[3,6] and Oluf Pedersen[1,7*]

[1]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, [2]LEITAT Technological Center, Terrassa, Spain, [3]Complications Research, Steno Diabetes Center Copenhagen, Herlev, Denmark, [4]The Bioinformatics Center, Department of Biology, University of Copenhagen, Copenhagen, Denmark, [5]Clinical Microbiomics, Copenhagen, Denmark, [6]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark, [7]Center for Clinical Metabolic Research, Gentofte University Hospital, Copenhagen, Denmark

Aims/hypothesis: To identify novel pathophysiological signatures of longstanding type 1 diabetes (T1D) with and without albuminuria we investigated the gut microbiome and blood metabolome in individuals with T1D and healthy controls (HC). We also mapped the functional underpinnings of the microbiome in relation to its metabolic role.

Methods: One hundred and sixty-one individuals with T1D and 50 HC were recruited at the Steno Diabetes Center Copenhagen, Denmark. T1D cases were stratified based on levels of albuminuria into normoalbuminuria, moderate and severely increased albuminuria. Shotgun sequencing of bacterial and viral microbiome in stool samples and circulating metabolites and lipids profiling using mass spectroscopy in plasma of all participants were performed. Functional mapping of microbiome into Gut Metabolic Modules (GMMs) was done using EggNog and KEGG databases. Multiomics integration was performed using MOFA tool.

Results: Measures of the gut bacterial beta diversity differed significantly between T1D and HC, either with moderately or severely increased albuminuria. Taxonomic analyses of the bacterial microbiota identified 51 species that differed in absolute abundance between T1D and HC (17 higher, 34 lower). Stratified on levels of albuminuria, 10 species were differentially abundant for the moderately increased albuminuria group, 63 for the severely increased albuminuria group while 25 were common and differentially abundant both for moderately and severely increased albuminuria groups, when compared to HC. Functional characterization of the bacteriome identified 23 differentially enriched GMMs between T1D and HC, mostly involved in sugar and amino acid metabolism. No differences in relation to albuminuria stratification was observed. Twenty-five phages were differentially abundant between T1D and HC groups.

Six of these varied with albuminuria status. Plasma metabolomics indicated differences in the steroidogenesis and sugar metabolism and circulating sphingolipids in T1D individuals. We identified association between sphingolipid levels and Bacteroides sp. abundances. MOFA revealed reduced interactions between gut microbiome and plasma metabolome profiles albeit polar metabolite, lipids and bacteriome compositions contributed to the variance in albuminuria levels among T1D individuals.

Conclusions: Individuals with T1D and progressive kidney disease stratified on levels of albuminuria show distinct signatures in their gut microbiome and blood metabolome.

KEYWORDS

multiomics, type 1 diabetes, albuminuria, metabolomics, microbiome, lipidomics, phageome

1 Introduction

Chronic kidney disease (CKD) is a major health burden with a prevalence of about 15% in the United States (1) with a record global rise of 41.5% mortality rates among CKD reported during the past 3 decades (2). Elevated albuminuria is strongly associated with end stage renal disease, cardiovascular disease, and death among CKD (3). Diabetes is the leading cause of end stage kidney disease, and about one third of individuals with type 1 or type 2 diabetes develop CKD, also referred to as diabetic nephropathy or diabetic kidney disease (2). Diabetic nephropathy progression in type 1 diabetes can be clinically characterized by stages of increasing albuminuria (a) moderately increased albuminuria (previously called microalbuminuria) (urinary albumin 30 to 300 mg/g creatinine), (b) severely increased albuminuria (macroalbuminuria or proteinuria) (>300 mg/g), (c) loss of renal function (glomerular filtration rate), and (d) finally need for kidney replacement therapy.

The intestinal microbiome constantly interacts with its host, constituting a dynamic balance and synergy, and thereby playing a role in maintaining and complementing metabolic and physiological functions (4). Studies in animal models of T1D support the hypothesis that an altered gut microbiome may lead to a "leaky" intestinal mucosal barrier, an imbalance in innate and adaptive immune systems and eventually triggering various chronic non-communicable diseases (5, 6). A low diversity of the gut microbiome is associated with dysmetabolism and (7) a state of dysbiosis is hypothesized to worsen the metabolic status of individuals with T1D (8, 9). Furthermore, a pathophysiological role of an imbalanced gut microbiota in diabetic nephropathy has been suggested (6).

A recent study has proposed a mechanism by which the gut microbiota impacts host's insulin resistance and albuminuria development by upregulating G protein-coupled receptor 43 (GPR43) (10). Furthermore, the impact of the gut microbiota upon host's metabolic status is not limited to direct interaction between organisms, but also through specific bacterial metabolites. In this context, it is of interest that a role of bacteria-derived phenyl sulfate has been reported to induce albuminuria in experimental models of diabetes (11). As for the plasma metabolome, few studies including our previous work (12) have reported associations between albuminuria and sphingomyelins, phosphatidylcholines (13) and unsaturated fatty acids and phospholipids (13–16).

In this framework, we previously identified differences in the gut bacteriome of T1D individuals stratified by albuminuria levels using a 16S rRNA gene marker approach (12). However, due to the modest taxa resolution provided by this method we failed to gain deeper insights into the bacteriome and phageome at species level and bacteriome functional potentials. The present study included the same study participants (T1D and HC) (12) but with high-resolution whole microbiome sequencing combined with untargeted plasma lipidomics and polar metabolite profiling to further investigate the single and multi-omics profiles, and their functional relationship. Thus, the objectives of the current study were: (i) to analyze both the taxonomical composition and functional potential of the metagenomic communities in T1D stratified by albuminuria levels, and in HC; (ii) to characterize the untargeted plasma metabolome of T1D, stratified by albuminuria levels; and (iii) to associate metabolome with metagenomic features, to investigate pathophysiological multiomic signatures of longstanding T1D with and without albuminuria.

2 Results

2.1 Characteristics of study participants

The study comprised 161 T1D individuals (50 with normoalbuminuria, 50 with moderately increased albuminuria and 61 with severely increased albuminuria and 50 healthy controls (HC). A study overview is given in Figure 1.

The study participants were aged 60 ± 11 years (mean ± SD), 42% being women. The mean diabetes duration was 42 ± 15 years with an eGFR of 75 ± 25 ml min^{-1} $(1.73\ m)^{-2}$ among individuals with type 1 diabetes (T1D). Detailed clinical characteristics of the study groups have been reported (12) and are again presented in Supplementary Table 1.

T1D individuals with elevated albuminuria were treated with anti-hypertensive and proton pump inhibitor drugs more frequently than others (Supplementary Table 1). HbA1c, and fasting plasma hs-CRP levels were expectedly higher and hemoglobin, and fasting plasma concentrations of total cholesterol and LDL cholesterol were lower in the T1D individuals upon stratification for increasing albuminuria and when compared to HC (Supplementary Table 1).

Serum creatinine was higher and corresponding eGFR levels lower in moderate and severely increased albuminuria groups compared to T1D with normoalbuminuria. The Bristol stool scale score and estimated bowel movement frequency were comparable in each T1D albuminuria group. The daily dietary macronutrient intake differed significantly between the albuminuria groups as described previously (12).

2.2 Community, taxa, and functional modules of the gut bacterial microbiota

In total 9,229 genes were mapped for the gut bacteriome using KEGG Orthology (KO) annotation. The gene richness distribution (Supplementary Figure 1) and the alpha (or intragroup) diversity (Supplementary Figure 2) were not distinctive of the four study groups when using the rarefied data (QMP count; Supplementary Figure 3; Supplementary Table 2), except when using Shannon index ($p = 0.02$). However, pairwise comparison showed lower diversity in moderately and severely increased albuminuria groups compared to controls when using Shannon (p_{micro}= 0.015 and

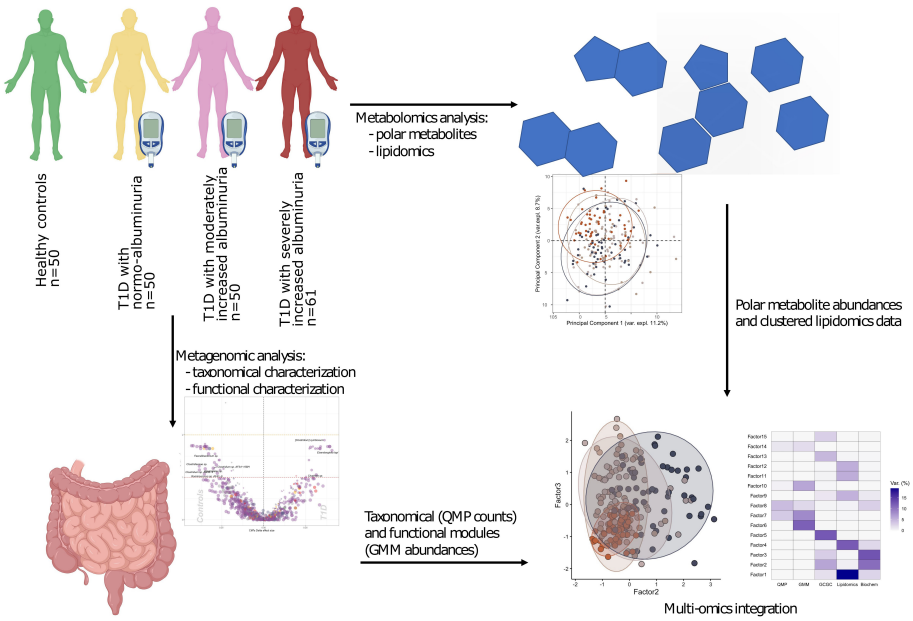


FIGURE 1
Study design and analytical overview. The study cohort comprised 50 healthy and 161 individuals with T1D, who stratified on albuminuria levels: normo-albuminuria (n=50, albuminuria levels =<3.39 mg/mmol), moderately increased albuminuria (n=50, albuminuria levels =3.39–33.79 mg/mmol) and severely increased albuminuria (n=61, albuminuria levels =≥33.90 mg/mmol). For each study participant 1) plasma samples were collected to perform non-targeted metabolomics analysis, including both polar metabolites and lipids, and 2) faecal samples collected for metagenomics analysis. For the metagenomic samples, differences between the clinical groups at both taxonomical and functional level were assessed. Lipidomics data was clustered into highly correlated lipid clusters in order to reduce dimensionality. Finally, different omics data types were integrated with the Multi Omics Factor Analysis (MOFA+) tool and metabolite origin screened to identify bacterial- and host-related metabolites through Least Shrinkage Selector Operator (LASSO) regression models. QMP, Quantitative Microbial Profiles; GMM, Gut Metabolic Modules.

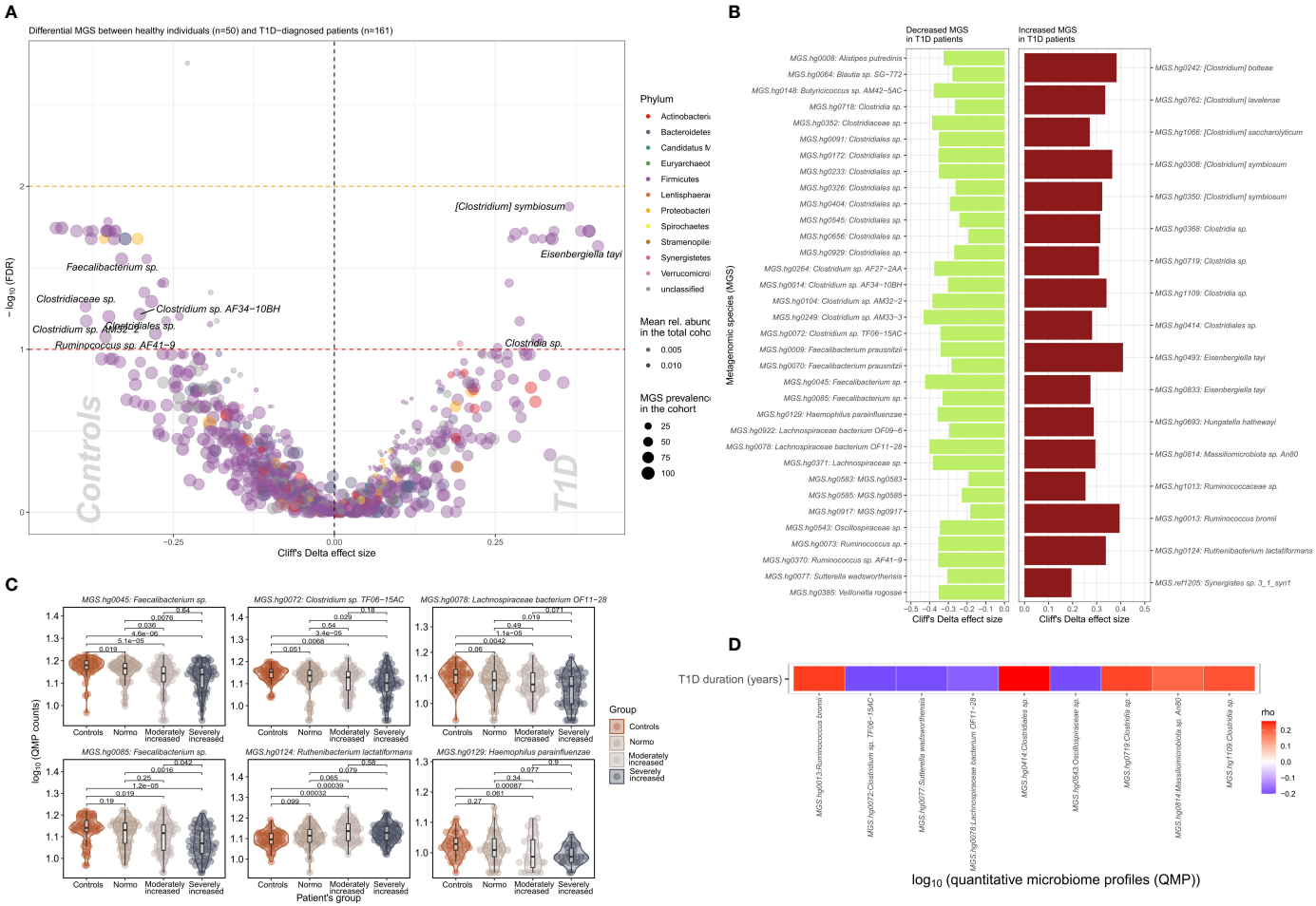


FIGURE 2
Differential bacterial abundance at metagenomic species (MGS) level, between T1D and healthy individuals. **(A)** Volcano plot showing difference in absolute abundance of metagenomic species (MGS) between T1D individuals (n=161) and the healthy controls (n=50). X-axis indicates Cliff's Delta effect size; Y-axis represents FDR-corrected (negative log) p−values. MGS that associated positively with T1D have been depicted towards increasing direction of effects (right). MGS circle color depicts the corresponding annotated phylum. Circle size corresponds to the number of individuals in the cohort where the specific MGS was found. Transparency of the circle corresponds to the average relative abundance in which each MGS is found within participants. **(B)** Significant contrasts in MGS abundance between T1D and healthy controls depicted with the bar length corresponding to Cliff's Delta effect size: green for higher and red for lower MGS abundances within T1D individuals. **(C)** Differential absolute MGS abundance between T1D subgroups stratified on levels of albuminuria. Individual distribution of the log10 transformed QMP counts (absolute abundance) is depicted for each MGS in violin and dot plots. Global distribution of the MGS counts for each T1D subgroup is depicted with a boxplot, indicating median value of the distribution with a horizontal line, first and third quartile with the limits of the white rectangle and the upper and lower limits of the distribution with vertical bars. Significance for pairwise comparison between different study groups is indicated with p-value. **(D)** Significant correlations between T1D duration in years and absolute abundance of MGS (QMP counts) are depicted as a heat map. Positive correlations are shown in red and inverse correlations in blue.

10.3389/fendo.2022.1015557

p_{macro} = 0.004), or Simpson (p_{micro}= 0.08 and p_{macro}= 0.02) indices (Supplementary Figure 2).

The community bacteriome dispersion varied between T1D and HC (p_{adonis}=0.001). These differences were mainly driven by T1D individuals with moderately and severely increased albuminuria groups ($p_{permanova}$ = 0.006), respectively. Furthermore, the T1D individuals were heterogeneously dispersed on the PCoA plot (Jensen-Shannon divergence index) compared to controls (Supplementary Figure 4). These observations were consistent with the 16s rRNA gene marker results previously reported by Winther et al. (12).

Shotgun sequencing allowed us to annotate the bacterial taxonomy to species level (Metagenomic species or MGS). Taxonomical annotations are provided in Supplementary Table 3.

We identified 51 bacterial metagenomic species (MGS, hereafter called species) that were differentially abundant between T1D and HC (Figures 2A, B; Supplementary Table 4) where the absolute abundance of 17 species were higher while 34 were lower in T1D individuals. In general lower absolute abundance of Short-Chain Fatty Acids (SCFA) producers such as *Veilonella rogosae* (17), *Faecalibacterium* sp., *Butyricicoccus* spp. *Clostridiales* sp. and *Lachnospiraceae bacterium* was observed in T1D compared to HC. *Clostridium* spp., including *C. saccharolyticum*, known for its saccharolytic activity in addition to *Eisenbergiella tayi, Hungatella hathewayi* and *Ruthenibacterium lactatiformans* were more abundant in T1D (Figures 2C, D). Some of the bacterial taxa (MGS) differences observed between overall T1D and HC individuals were also observed between T1D individuals with moderate or severe albuminuria and HC individuals (Figure 2C; Supplementary Figures 5, 6). Supplementary Figure 7 provides an overview of MGS absolute abundance within the T1D albuminuria subgroups and MGS specific and common to moderate and severely increased albuminuria groups, compared to HC.

Further shotgun sequencing of the metagenome facilitated mapping of functional metabolic potential and anaerobic fermentation capacity of the metagenome in form of Gut Metabolic Modules (GMMs) curation that represent a cellular enzymatic process defined by input and output metabolites. The GMMs computed using the Omixer Reference Pathway Mapper and KEGG Orthology (RPM) (18) differed in abundance (Supplementary Table 5) in T1D group compared to HC using univariate analyses (Supplementary Table 6; Supplementary Figure 8). The T1D bacteriome was enriched for modules of sugar degradation (most abundantly for ribose, fucose and trehalose) and for modules of amino acid metabolism, particularly for non-polar amino acids (alanine, glycine, isoleucine, methionine, and tryptophan), followed by acidic (lysine, cysteine, and histidine) and polar (threonine and glutamine) amino acids (P_{FDR}<0.10; Supplementary Figure 8; Supplementary Table 6).

2.3 Gut phageome and type 1 diabetes

We identified 502 highly abundant phages in our study participants of which 25 were differentially enriched between T1D and HC (Figure 3A). Interestingly, the relative abundance of six of these differentially enriched phages (uvig_37554, uvig_280596, uvig_296393, uvig_436746, uvig_514207, uvig_557689) changed with increasing level of albuminuria (Figure 3B).

Furthermore, CLR-normalized phage abundances were associated inversely with multiple clinical factors (adjusted for age, sex, and diet) including T1D status (25 phages), diabetes duration (20 phages), eGFR (21 phages) and plasma creatinine (7 phages) levels (Supplementary Figure 9). Hematocrit was found to be related to 17 phages overlapping with hemoglobin and partly glycosylated hemoglobin. All these associations were partly overlapping with T1D-associated phages (Supplementary Figure 9). The distribution of the samples derived from the bacteriome and the phageome analyses were similar, as assessed by Procrustes analysis (correlation = 0.67; Figure 3C).

2.4 Plasma metabolome and lipidome

2.4.1 Plasma polar metabolites

We examined the differential abundance of 398 (143 known and 255 unannotated) plasma polar metabolites (Supplementary Table 7) between i) T1D versus HC and, ii) albuminuria subgroups within T1D, using univariate and multivariate approaches.

To identify a subset of T1D - linked metabolites, we used the partial least squares-discriminant analysis (PLS-DA) approach (multivariate) splitting the dataset into 70% training and 30% validation sample subsets. Albeit we achieved a good separation (R^2: 91.8%) using five components, the reproducibility of the model was limited (Q^2: 25%; Supplementary Figure 10). Next, we selected polar metabolites (n=132) with a VIP (Variable Importance in Projection score) ≥ 1 in the PLS-DA analysis and we generated a PCA plot that effectively differentiated between T1D and HC groups, whereas the score did not provide any differentiation between albuminuria groups. (Figure 4A).

We identified 58 polar metabolites that were differentially abundant between T1D and HC (P_{FDR}<10%; Figures 4B, C; Supplementary Table 8) using the univariate approach. In T1D individuals, the plasma concentration of 1,5-anhydrosorbitol was significantly lower followed by cholesterol and butylated-hydroxytoluene while several sugar derived metabolites like lyxofuranose and beta-D-tagatopyranose were higher (Figure 4B; Supplementary Table 8). Within the T1D group a univariate comparison between moderately and severely increased albuminuria groups revealed significantly (P_{FDR}<10%) lower levels of ribitol, benzeneacetic acid, decanoic acid and 3-phenylpropanoic acid while higher levels

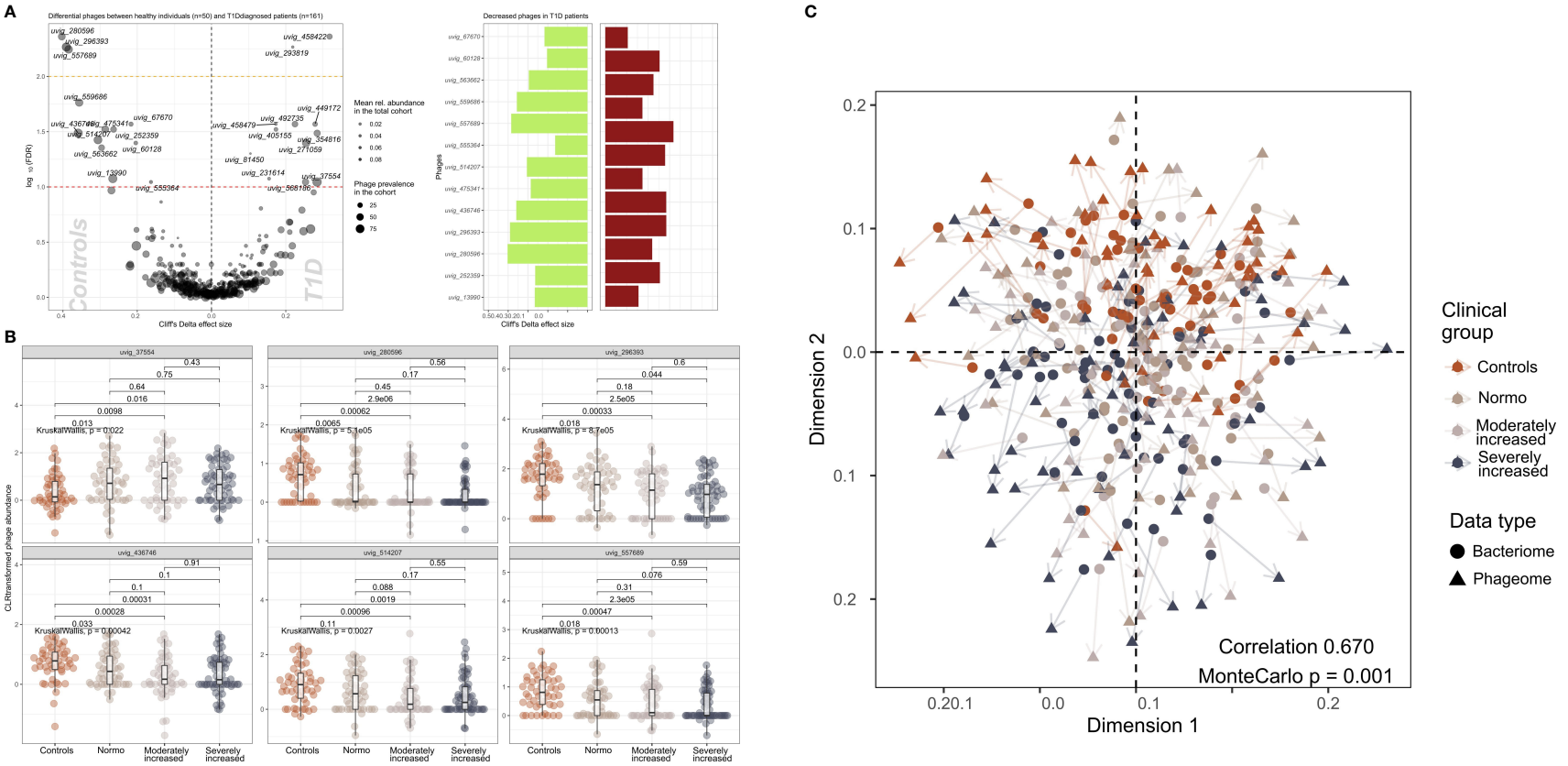


FIGURE 3
Contrasted relative abundance of bacteriophages in T1D and healthy control individuals. **(A)** In the left panel, a volcano plot compares the relative phage abundances between T1D and healthy individuals. The X-axis represents Cliff's Delta effect size while the Y-axis represents the association threshold for the comparison, (negative log) FDR-adjusted p-value. Significantly contrasted phages (above the red dotted line) are annotated. Dot size is relative to the global prevalence of phage in the present study sample, while their transparency is relative to their mean abundance in the cohort. In the right panel, significantly contrasted phages are shown as bar plots corresponding to Cliff's Delta effect sizes. Phages with significantly lower abundance in T1D are indicated as green bars while red bars represent phages with significantly higher abundance in T1D. **(B)** Distribution of abundance of selected phages when comparing healthy controls with the three groups of T1D stratified by albuminuria. For each phage, the CLR-transformed abundance distribution is represented in differently colored dots and boxplots showing the global group distribution. Multi-group comparison computed with Kruskal-Wallis test is included in each plot, as well as the pairwise comparisons between all study groups, performed with Wilcoxon test. **(C)** Comparison between the global composition of the phageome and the bacteriome, performed with Procrustes test. Principle coordinates analysis (PCoA) shows the disposition of individuals in the corresponding scores plot, for both bacteriome (circles) and phageome (triangles) based distance matrix. An arrow has been drawn connecting the same individuals. Individuals have been colored depending on their clinical group. Correlation and significance are indicated in the bottom right corner of the plot.

10.3389/fendo.2022.1015557

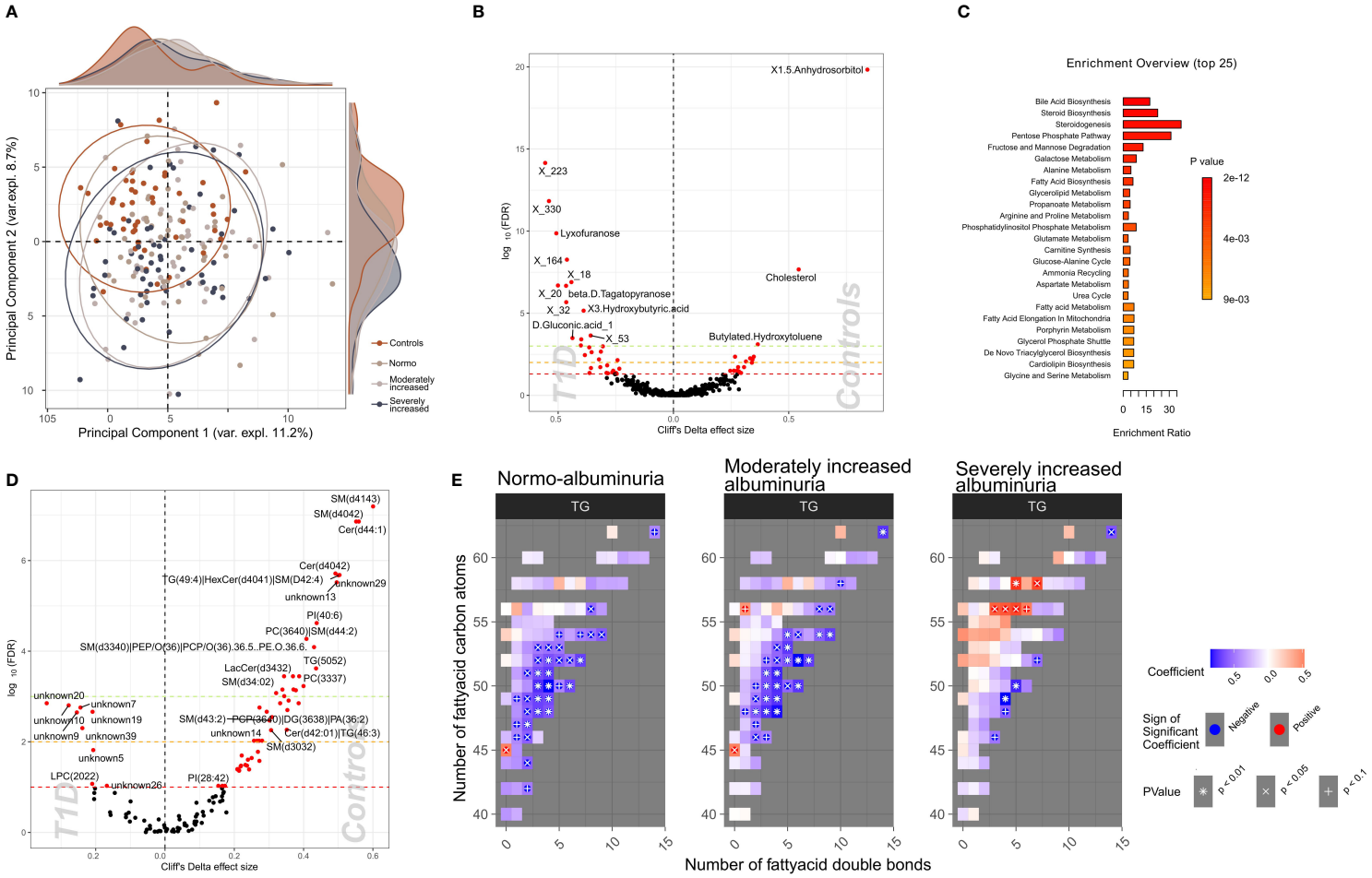


FIGURE 4
Profiles of plasma polar metabolites and lipids in T1D and healthy control individuals. **(A)** Scores plot resulting from the partial least square discriminant analysis (PLS-DA) demonstrating polar metabolites that contribute highly (Variable influence on project (VIP) > 1) to a T1D signature. Colors depict clinical study groups and ellipses demarcate the spread of individuals within each group. Explained variance for first two principal components (PC1 and PC2) is indicated on the respective X- and Y- axis. Global distribution of the participants from each clinical group is shown as a density plot in the secondary Y-axis. **(B)** Univariate analysis (volcano plot) depicting differential circulating polar metabolite abundances between healthy and T1D individuals. Significantly contrasted (FDR < 10%) polar metabolites are depicted in red, with increasing effects in direction of controls (right). **(C)** Functional metabolic pathway identification comparing T1D and healthy controls and depicting enriched metabolic pathways: annotations of top polar metabolites using the human metabolome database (HMDB). Bars have been ordered and colored by the enrichment p-value score. **(D)** Univariate analysis (volcano plot) depicting differential abundances of serum lipid cluster between healthy and T1D individuals. Significantly contrasted (FDR < 10%) lipid clusters are colored in red with increasing effects in direction of controls (right). **(E)** Triglyceride (TG) species associated with different T1D albuminuria groupings (compared to healthy controls) based on the number of carbon atoms (Y-axis) and the number of double bonds (X.axis) present in the chemical structure. Blue-to-red gradient heatmap suggests directionality of the association (negative to positive). Significant associations are indicated in each cell with the following symbols: + p<0.1, X p<0.05, *p<0.01.

10.3389/fendo.2022.1015557

of 2,3-dihydroxybutanoic acid in the severely increased albuminuria group (Supplementary Figure 11).

2.4.2 Functional enrichment of plasma polar metabolites and metabolic pathways

The functional enrichment analyses of plasma polar metabolites (T1D vs. HC) led to identification of several enriched human metabolic pathways. The differentially enriched pathways mainly comprised of steroidogenesis and steroid biosynthesis, bile acid biosynthesis, pentose phosphate pathway and several sugar metabolism pathways, such as fructose and mannose degradation and galactose metabolism (Figure 4C; Supplementary Table 9).

2.4.3 Plasma lipidomics

7,470 plasma lipids (476 known and 6,994 unannotated, Supplementary Table 10) were clustered into 122 strongly correlated lipid clusters ranging between 3 to 1,054 lipids per cluster (Supplementary Tables 11, 12).

We identified 60 lipid clusters (P_{FDR}<10%) differentially abundant between T1D and HC (Supplementary Table 13). The T1D lipidome was enriched in a set of lysophosphocholines (LPCs) and unknown lipids containing 20 to 22 carbon atoms (Figure 4D). Inversely, lipid clusters containing long chain ceramides (40-44 carbon atoms) and sphingomyelins (30-41 carbon atoms) were highly abundant in HC compared to T1D individuals.

Further, we analyzed the distribution of the annotated lipidome stratified by albuminuria status within the T1D and compared these to HC. Triglyceride (TG) lipid species with large number of carbon atoms (>55 carbon atoms) were positively associated with severely increased albuminuria, while comparatively shorter TGs (40-55 carbon atoms) were inversely associated to both normo-albuminuria and moderately increased albuminuria groups (Figure 4E).

2.5 Multi-omics factor analysis based on albuminuria levels

Multi-Omics Factor Analysis 2 (MOFA2) (19) tool allowed us to integrate all data on gut microbiome, plasma metabolites (metabolomics and lipidomics), and clinical biochemistry and, by a process of factorization, identify which data type was the most contributing to the individuals' T1D phenotype (stratified on albuminuria status).

From the factorization results, we observed that the lipidomics dataset explained a major part of the factors composition (~50% variance), followed by the polar metabolites, bioclinical variables, and the functional bacterial profiling (GMM) of the gut microbiome (~30% of factor's variance) (Figure 5A). Finally, analysis of the taxonomical composition of the bacteriome (QMP) explained about 15% of the factors' composition. Furthermore, we observed a rather limited interaction between the plasma metabolome and the gut microbiome data, as no factors were found that combined effectively these two data types (Figure 5B). Instead, the combination of biochemistry analysis and polar metabolites (factors 2 and 3) differentiated well between T1D and HC individuals. Thus, by splitting all the data combined into 15 factors we observed a trend by which T1D status (and partly albuminuria levels) influenced the position of the individuals along the generated principal components (Figure 5C).

Details on clinical and metabolomic components resulting in Factors 2 and 3 have been presented in (Supplementary Figure 12).

2.5.1 Relationships between the gut bacteriome and differential blood metabolome

Since metabolome composition explained the factors composition better than the other data types, we assessed the relationships of the gut bacteriome and the blood metabolome (Figure 6). Overall, 30% of the metabolites were associated with the bacteriome taxonomical profiling, while 40% to 50% of it was related to the bacteriome *via* functional profiling (Supplementary Figure 13)

All the differentially abundant polar metabolites in the T1D individuals were associated to the bacterial abundances of *Faecalibacterium prausnitzii, Clostridium spps.*, Lachnospiraceae spp. and *Eisenbergiella tayi*.

The lipidome composition was mostly associated with the bacteriome functional profiling (GMMs). In addition, the lipidome composition was associated to only a small subset of specific bacterial and/or archaeal species, such as *Akkermansia muciniphila* and *Methanobrevibacter smithii.*

3 Discussion

In the present study, applying deep metagenomic sequencing-based functional annotation and multi-omics factorization we identified multiple additional molecular signatures for T1D in the gut microbiome and plasma metabolome and lipidome both individually and when combined, compared to our previous findings (12). Moreover, we provide a gut phageome profile for T1D. While most significant differences in the gut microbial abundance and circulating metabolites and lipids were observed between T1D and HC, both moderately and severely increased albuminuria groups also evidenced significantly enriched bacteriome and plasma metabolite levels when compared to HC. In functional bacteriome analyses, we identified sugar, amino acid, and short chain fatty acid (SCFA) metabolizing species differentially enriched in T1D compared to HC, while no significant differences in the functional nature of the bacteriome was observed upon albuminuria stratification. The latter may suggest that nephropathy development is a continuous

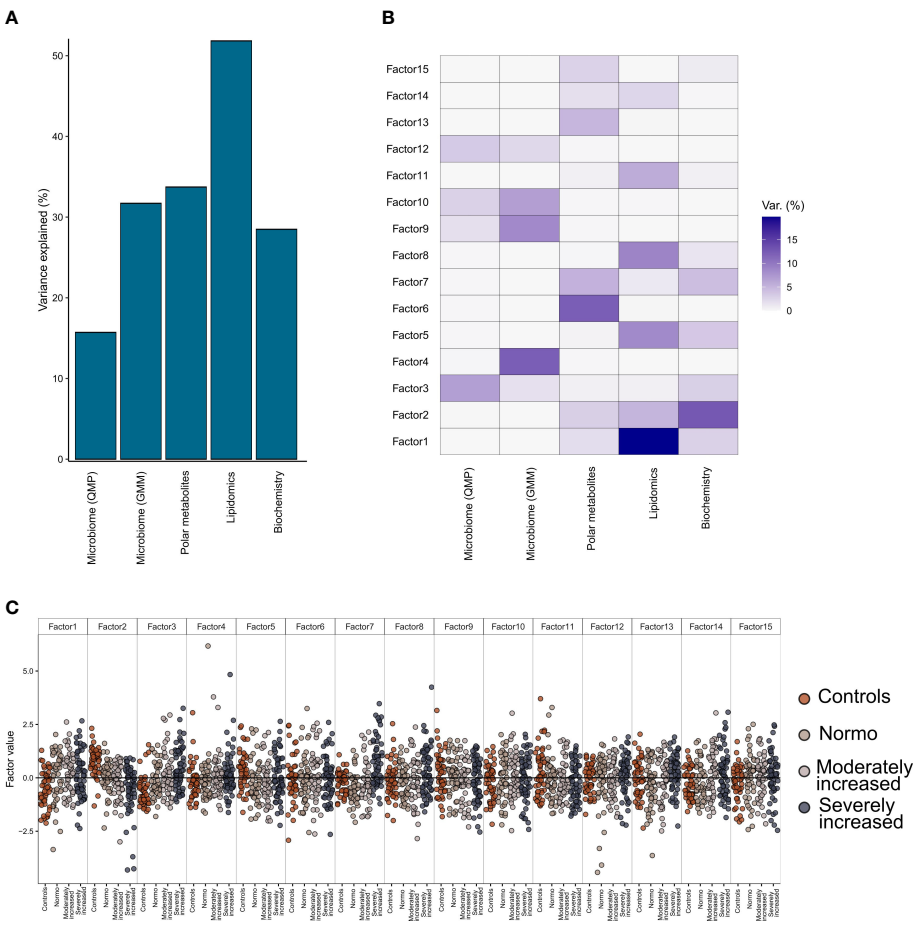


FIGURE 5
Multi-Omics Factor Analysis (MOFA). Results from multi-omics data integration after combining multiple data types: metagenomics data (taxonomical and functional), plasma metabolomics data (polar metabolites and clustered lipids) and biochemistry data using the MOFA+ tool. **(A)** Global explained variance for each of the data types (included in the integration step) has been represented as a bar chart. **(B)** Composition of 15 individual factors generated with the proportion of explained variance by each of the data types, displayed as a white-to-blue gradient in increasing order. **(C)** Distribution of the eigenvalues obtained for each of the factors for the combined dataset. Within each of those factors the individual, eigenvalues are represented as dot plots. Different colors in the dot plot depict different clinical study groups: orange for healthy controls, dark gray for T1D with normo-albuminuria, light gray for T1D with moderately increased albuminuria and black for T1D with severely increased albuminuria. Each of the study groups are also labelled in the bottom horizontal axis.

process, modifying both the microbiome and the metabolome in a constant manner. Through multi-omics we report that circulating lipids explained most of the phenotypic variance for T1D (stratified for albuminuria) followed by polar metabolites, clinical factors, and functional gut bacterial profiling (GMM). Albeit a limited interaction between circulating metabolome and gut microbiome was observed, a combination of circulating polar metabolites and clinical risk factors could best differentiate between T1D and HC individuals.

Deep shotgun sequencing of the microbiome allowed us to annotate the bacterial taxonomy to species or strain level (Metagenomic species or MGS) facilitating mapping of functional metabolic potential and anaerobic fermentation capacity of the metagenome in form of specie-function relationship or Gut Metabolic Modules (GMMs) (18). The absolute number of differentially abundant MGS were relatively higher in the group of severely increased albuminuria, followed by moderate and reduced in normo-albuminuria groups within the T1D cases (Supplementary Figure 7). Overall, a higher number of MGS seemed to occur with a lower absolute abundance in the T1D group compared to HC (Figure 2B). This is not surprising as most immune disorders have been generally associated with a loss of gut microbial diversity, specifically *Akkermansia* and *Faecalibacterium*, both potentially contributing to host immune tolerance (20). We also report significantly lower absolute abundance of *Faecalibacterium* in the T1D group compared to HC.

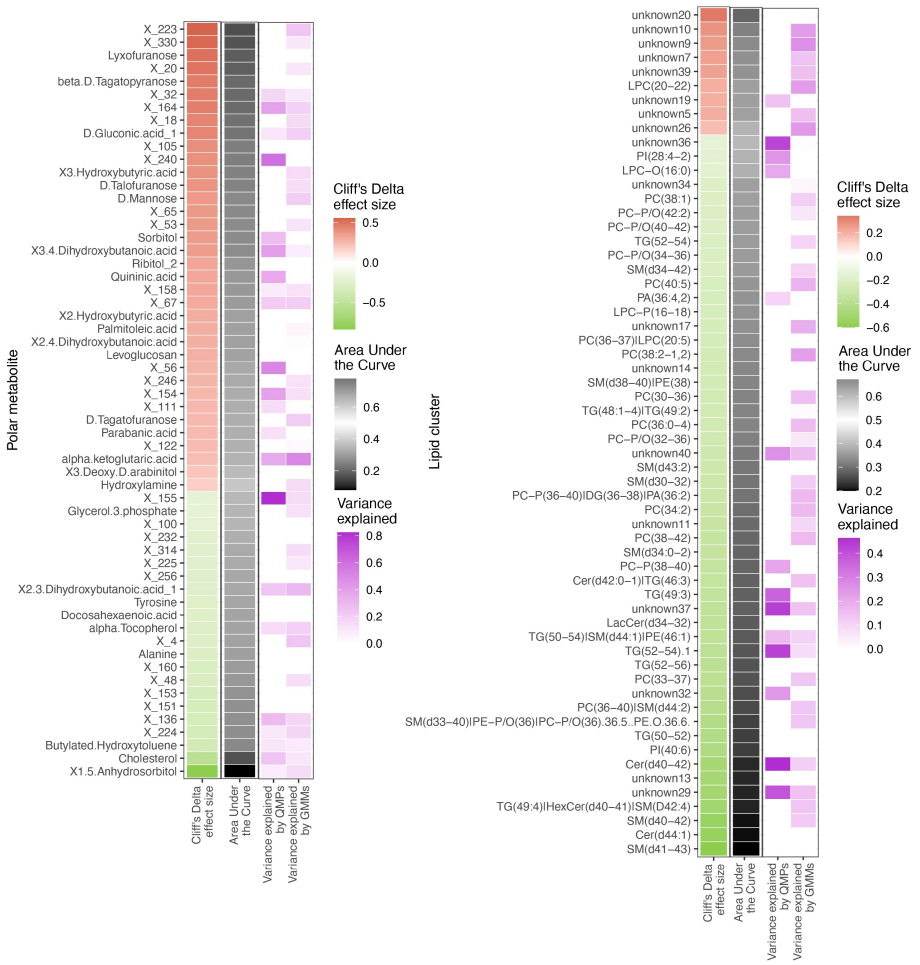


FIGURE 6
Differentially abundant circulating metabolites and microbiome origin and functionality assessment. Summary of the circulating metabolites and lipids data integrated with gut microbiome origin and functionality in the T1D vs healthy controls comparison. The two heatmaps display significantly contrasted (FDR < 10%) polar metabolites (left) and lipid clusters (right) between T1D and healthy individuals. Column 1 (red green of each of the two heatmaps shows the Cliff's Delta effect size for the clinical group comparison, ordered from higher to lower abundance in T1D individuals and colored from red (higher) to green (lower) gradient depending on its relative abundance in T1D individuals. Column 2 (black-white) shows the usefulness of the metabolite for discrimination between T1D and healthy individuals based on Area Under the Curve (AUC) analyses. The black-to-white-to-black gradient for AUC depicts assessment ability with 50% being uninformative AUC, 0% being the limit for the identification of healthy individuals and 100% being the limit for identification of T1D individuals. Column 3 (purple white) depicts relationship between the metabolites and bacterial QMP counts (absolute MGS abundance). Cells are colored if there is any significant association with metabolite levels (based on LASSO modelling) and the white-to-purple color gradient depicts the explained variance by the microbiome. Column 4 (purple white) depicts the relationship between the circulating metabolites and functional abundance of GMMs. Cells are colored if there is any significant association with metabolite levels (based on LASSO modelling) and the white-to-purple color gradient depicts the explained variance by the microbiome metabolic potential (or functionality).

Alteration of the gut microbiota in T1D resulted in lower SCFA production capabilities and increased saccharolytic activity. SCFAs are carboxylic acids with aliphatic tails of 1-6 Carbon atoms, acetate (C2), propionate (C3) and butyrate (C4), being most abundant produced by anerobic fermentation of polysaccharides or dietary fibers. SCFAs are mainly produced by *Bacteroidetes* (C2 and C3) and *Firmicutes* sp. (C4) further promoting beneficial bacteria survival (21, 22). Dietary fiber can upregulate carbohydrate metabolism enzymes, increasing SCFAs that enhance intestinal epithelial barrier function (particularly C4) reducing metabolic toxins and expression of inflammatory molecules (23). The C4 SCFA is also deemed important for hypoxia inducible factor 1 (HIF1) stability in maintaining epithelial barrier through C4-based oxygen balance and maintaining low oxygen concentrations in gut (21). A recent study also reported that lower SCFAs were associated with metabolic syndrome (22) and autoimmune disorders like inflammatory bowel disease (21). The absolute abundance of

the majority of *Clostridiales* sp. belonging to phylum Firmicutes was significantly lower in the T1D in the current study. A polysaccharide treatment induced increase in abundance of SCFA producing *Clostridiales* sp. has been reported to lower blood glucose levels, improve glucose tolerance and restore lipid balance in a rat model of type 2 diabetes (24) further supporting our findings and the relationship between lower bacterial abundance, correspondingly lower bacteriome function and T1D. The absolute abundance of *R. lactatiformans*, a lactate-producing bacterium (25), was also higher in T1D partially aligning with the elevated blood lactate levels recently reported in T1D (26). In addition, the absolute abundance of *Clostridium Spp.*, which is known for its sugar degradation capabilities was higher in the T1D group. Interestingly, abundance of *H. hathewayi*, which is reported to be positively associated with circulating taurine levels (27) was enriched in T1D. Given the fact that taurine levels reduce hyperglycemia (28, 29), abundance of *H. hathewayi* among T1D may explain the body's compensatory mechanism to counter hyperglycemia.

Functional classification of the T1D associating plasma metabolites suggested pathways enriched for steroidogenesis, bile acid biosynthesis and sugar metabolism. Additionally, the T1D plasma lipidome profile associated with alterations in the circulating sphingolipid (SL) levels, especially ceramides, which were partially produced by the gut microbiota, particularly Bacteroides sp. (Supplementary Table S14). Sphingolipids are known for their bioactive role as secondary messengers especially in metabolic disorders. Recently, it has been demonstrated that gut bacterial sphingolipids may pass the intestinal epithelium barrier, modifying the host's sphingolipid metabolism (30). Particularly, bacterial sphingolipids inhibit the processing of the host's own sphingolipids, including ceramides, the lipid type most altered in T1D. In the current study, lipids, especially long chain sphingolipids (ceramides and sphingomyelins) were multifold lower in T1D compared to HC, irrespective of the albuminuria stratifications. Recent studies supporting our findings observed associations between host circulating long chain ceramides and reduced kidney function (13) and diabetic kidney disease (14, 15) in T1D. We recently reported long chain sphingomyelins, to be inversely associated with albuminuria (especially severely increased albuminuria vs. normoalbuminuria) in 669 individuals with T1D (13) which were also observed in the DCCT/EDIC trial (15) but remained inconclusive in the current study potentially due to limited sample size. Animal studies have confirmed the role of sphingolipids-derived ceramides in insulin resistance of liver (31) while their pharmacological inhibition improved glucose homeostasis (32). However, the mechanism by which sphingolipids may influence albuminuria development is unclear. Bacteria-derived sphingolipids have been reported to act as ligands for G protein-coupled receptors (33), including those found in the intestinal epithelium (34, 35). These results are compatible with the albuminuria-inducing gut microbiota action hypothesis mediated through GPR43 upregulation (10).

The observed metabolomics profile was in accordance with our previous study looking into the pathophysiology of diabetic kidney disease (DKD) (36). In the present study we identified several associations between plasma metabolites and albuminuria in T1D. The main association was with polyols where we showed that plasma concentrations of sorbitol and ribitol both were higher in direct proportion to higher levels of albuminuria. The same metabolites, pointing to a validation in an independent cohort, applied to hydroxy butyrate 3,4-dihydroxybutanoic acids which were associated with moderate and severely increased albuminuria and have now replicated in the present study. Most interestingly sorbitol 3,4-dihydroxybutanoic and quininic acid were among the three identified metabolites showing strong associations to the abundance of bacterial species (Figure 2). However, the outcome of the functional enrichment results of circulating metabolites in T1D group can mainly be characterized as alterations related to two metabolic functions: cholesterol biosynthesis, and glucose metabolism.

The altered cholesterol levels resulted in a specific enrichment of steroid hormones biosynthesis and related metabolic pathways. Alterations of sexual hormone levels, with link to reduced fertility and increased risk for cardiovascular disease have been previously reported in T1D (20), which might be related to our results and the specific steroid metabolic alterations. For the glucose metabolism, the pentose phosphate metabolic pathway was highly enriched in our dataset. Interestingly, a protective role against chronic diabetes complications, including diabetic nephropathy, has been previously reported for this metabolic pathway (21, 22, 37).

Further factorization of multi-omics data in the current study demonstrated that the circulating lipidome could explain most of the (50%) T1D phenotypic variance followed by polar circulating metabolites, functional bacterial profiles (GMM) and finally 15% by taxonomical bacterial composition (QMP) (Figure 5). However, no distinction based on level of albuminuria could be made. We identified two distinct interactions that differentiated the T1D and HC groups best. These factors (factor 2 and 3, Figure 5) comprised a set of polar metabolites and bio-clinical markers. Interestingly the most significant drivers within these clinical and metabolite features were diabetes duration and a combination of sugar derivatives (Supplementary Figure 12). However, only a limited interaction between plasma metabolome and gut microbiome data was evidenced. While investigating metabolome origins we found polar metabolites that were associated with bacterial abundance reflected in relevant GMMs enriched for amino acids metabolism. Similarly, the lipidome was associated with bacterial function through relationships between the lipidome composition and abundance of GMMs involved in sugar degradation.

On this note, we need to consider the role of medication in both the gut microbiome (38, 39) and the blood metabolome

(40), especially considering the amount of time T1D individuals have received medication (30 to 45 years since diabetes diagnosis). We observed a strong influence of statins on abundance of *Clostridia* and *Ruminococcaceae spps.*, while the abundance of remaining gut bacteria seemed less influenced by the prescribed drug regimes.

Our study doesn't come without limitations, the most obvoius fact being this is a cross-sectional study, which limits the evaluation of the gut microbiome's contribution to albuminuria development. The findings reported here are based on bioinformatics analysis of combined omics technologies, which is a recent research field lacking standardized protocols, hampering reproducibility. The same holds for the metabolomics data analysis, which could only annotate a modest fraction of the identified metabolites. Moreover, identification of phages in the current study was limited to annotations available in the Gut phage database (19). Finally, the findings reported here are based on one specific cohort. Considering that the microbiome and the metabolome are also affected by a combination of environmental and biological factors, a validation on a different cohort with different environmental conditions would provide with more robust results.

Gut microbiome composition is known to be modifiable by several environmental factors, such as diet (41) or exercise (42) and/or through fecal microbiota transplant (FMT). Since this study demonstrates a relationship between the dysbiotic gut microbiota and an altered plasma metabolome composition in T1D cases with albuminuria, and if our findings are replicated in independent studies, it might serve as a basis for future microbiota-based interventions in T1D with albuminuria. Such interventions might include diet modifications or prescription of second-generation probiotics.

4 Materials and methods

4.1 Study design

A cross-sectional study conducted during 2016-2017 recruited 161 type 1 diabetes (T1D) individuals followed at the Steno Diabetes Center Copenhagen (SDCC) outpatient clinic and 50 non-diabetic age and gender matched healthy control individuals (12). All participants were >18 years of age and type 1 diabetes was diagnosed according to the WHO-criteria. Exclusion in the current study participation involved presence of at least one of the following conditions, (a) non-diabetic kidney disease; (b) renal failure (estimated glomerular filtration rate or eGFR <15 ml min^{-1}[1.73 $m]^{-2}$), dialysis or kidney transplantation; (c) change in renin–angiotensin–aldosterone system (RAAS)-blocking treatment during the month prior to study inclusion; (d) treatment with systemic antibiotics in the 3 months prior to recruitment; and (e) treatment with systemic immunosuppressive agents. Individuals with T1D were stratified into three different albuminuria groups based on the highest urine albumin/creatinine ratio (UACR) level measured on the study visit or documented previously in two out of three consecutive urine samples within past 1 year (as 24 h urine albumin content in samples [UAER] or UACR). Basis albuminuria groupings, 50 individuals had normoalbuminuria (<3.39 mg/mmol corresponding to <30 mg/24 h or mg/g), 50 had moderately increased albuminuria (3.39–33.79 mg/mmol corresponding to 30–299 mg/24 h or mg/g) and 60 had severely increased albuminuria (≥33.90 mg/mmol corresponding to ≥300 mg/24 h or mg/g). There was no recorded history elevated albuminuria for participants classified as having normoalbuminuria. For the severely increased albuminuria group, at least 30 individuals were selected based on concurrent eGFR <60 ml min^{-1}[1.73 $m]^{-2}$. The study design has been described in Supplementary Section (Figure 1, Supp Text). The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the Danish Capital Region (protocol H-15018107). All participants gave written informed consent and provided with self-collected fecal sample for posterior metagenomics analysis (Figure 1).

4.2 Metagenomics

4.2.1 Sequencing

Sequencing and metagenomic species (MGS) generation was performed as previously described (43). Quality control of raw FASTQ files was performed using KneadData (v. 0.6.1) to remove low-quality bases and reads derived from the host genome as follows: Using Trimmomatic (v. 0.36), the reads were quality trimmed by removing Nextera adapters, leading and trailing bases with a Phred score below 20, and trailing bases in which the Phred score over a window of size 4 drops below 20. Trimmed reads shorter than 100 bases were discarded. Reads that mapped to the human reference genome GRCh38 (with Bowtie2 v. 0.2.3.2 using default settings) were discarded. Read pairs in which both reads passed filtering were retained; these were classified as high-quality non-host (HQNH) reads.

4.2.2 Metagenomic species generation

As reference gene catalogue, we used the Clinical Microbiomics Human Gut 22M gene catalogue (22 459 186 genes), which was created from >5000 deep-sequenced human gut specimens. For MGS abundance profiling, we used the Clinical Microbiomics HGMGS v.2.3 set of 1273 MGS, which has highly coherent abundance and base composition in a set of 1776 reference human gut samples (44).

HQNH reads were mapped to the gene catalogue using Burrows-Wheeler Alignment (BWA) men (v. 0.7.16a) with options to increase accuracy (-r 1 -D 0.3). PCR/optical

duplicates were removed using samtools (v. 1.6). Each individual read was considered mapped if the following criteria were met: an alignment of ≥ 100 bases, ≥ 95% identity in this alignment, and a mapping quality (MAPQ) ≥ 20. However, if a read failed to align to the gene sequence with > 10 bases at either end, it was considered unmapped. Reads meeting the alignment length and identity criteria but not the MAPQ threshold were considered multimapped. Reads failing the alignment length or identity criteria were considered unmapped. Read pairs were classified into one of three possible categories as follows:

1) Read pairs in which both individual reads were considered unmapped.
2) Read pairs in which both individual reads were multimapped, or were mapped to genes in different MGSs, or one was multimapped and the other was unmapped, were considered multimapped.
3) Read pairs in which both individual reads mapped to the same gene; or in which one read mapped to a gene and the other was unmapped, multimapped, or mapped to another gene in the same MGS (see below); were considered mapped. A gene counts table was created with the number of mapped read pairs (for each gene), unmapped read pairs, and multimapped read pairs.

For each MGS, the "core" genes were defined as the 100 genes specific for the MGS and with the highest correlation to the mean and lowest absolute deviation from the mean. A MGS counts table was created based on the total gene counts for the 100 core genes of each MGS. However, MGS was considered detected only if read pairs were mapped to at least three of its 100 core genes; counts for MGSs that did not satisfy this criterion were set to zero. The MGS counts table was normalized according to effective gene length (accounting for read length) and then normalized to sum to 100%, resulting in relative abundance estimates of each MGS. Down-sampled (rarefied) MGS abundance profiles were calculated by random sampling, without replacement, from each sample in the MGS counts table. Values with fewer than three counts after down-sampling were set to zero, and the counts table was normalized according to effective gene length and then normalized to sum 100%.

4.2.3 Computation of quantitative microbial profiles

Faecal samples were subjected to bacterial cell counting with flow cytometry. Aliquots of 0.08-0.15 g defrosted faeces were diluted 2,118 times in staining buffer (1 mM EDTA (*Sigma-Aldrich*), 0.01% Tween20 (*Sigma-Aldrich*), pH 7.2 DPBS (*Lonza BioWhittaker*), 1% BSA (*Sigma-Aldrich*)). In order to remove debris from the faecal solutions, samples were filtered using a sterile syringe filter (pore size 5 µm (*pluriSelect*)). Next, 170 µL of the bacterial cell suspension was stained with 20 µL DAPI (1mM in H_2O, *Sigma-Aldrich*). The flow cytometry analysis of the bacterial cells present in the suspension was performed using a BD Fortessa LSRII flow cytometer (*BD Biosciences*). Measurements were performed at a pre-set flow rate of 0.5 µL/sec. Fluorescence events were monitored using the 440/40 nm, 575/26 nm, and 695/40 nm optical detectors, respectively. Forward and sideways-scattered light was also collected. The BD FACSDiva™ Software was used to gate and separate the bacterial fluorescence events from the faecal sample background. A threshold value of 900 was applied on the area of forward scattered channel (FSC) and a threshold value of 200 was applied on the area of sideways scattered (SSC) channel. Other flow settings are listed in Supplementary Table 16.

Density plots of blue fluorescence (440/40 nm) versus FSC allowed for distinction between the stained microbial cells and instrument noise or sample background. Density plots of red fluorescence (695/40 nm) versus FSC allowed for distinction between the counting beads and other particles in the testing solution, including bacteria, instrument noise or sample background. The exact same gates and gating strategy were applied for all samples in the form of a fixed template to allow direct comparison between measured samples.

Bacterial cell counts were later used for quantitative microbial profiling (QMP), as described elsewhere (45). Briefly, data was rarefied to equal sampling depth and cell counts used to compute the total abundance of each MGS.

4.2.4 Computation of gut metabolic modules

Emapper software (v. 1.0.3, HMM mode) was used to compare each gene in the gene catalogue to the EggNOG (v. 4.5) orthologous groups database (http://eggnogdb.embl.de/), resulting in annotations for 65% of genes. These genes were then mapped from EggNOG to the Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology database (http://www.genome.jp/kegg/kegg1.html) using MOCAT2 lookup tables (http://mocat.embl.de/). The annotation of GMMs was performed in R applying Omixer-RPM (http://www.raeslab.org/software/gmms.html). The GMM counts are referring to GMM QMPs based on MGS QMP counts. The GMM abundance table was then transformed using the central log-ratio (CLR) to ensure normality and assess its compositionality nature.

4.2.5 Computation of the phageome from sequenced faecal DNA

Bulk sequence reads reads derived from sequencing of faecal DNA were aligned against Gut Phageome Database (GPD) (46) with BWA mem. Obtained phages were then quality-filtered, retaining only the reads aligning to, at least, 75% of the phage genome length. Phageome counts dataset was then rarefied to the minimal reads and those phages not found in, at least, 10% of the samples (n=21) were removed. This resulted in a total of 502 phages to be included in the final dataset. Phageome counts were

then transformed with CLR approach to assess for compositionality of the data.

4.2.6 Computation of differential microbiome features

For the taxonomical analyses, the absolute MGS and/or genus clustered counts were used, after transformation with faecal cell counts. We used linear mixed models, adjusting for age, sex, race, BMI, and dietary data to compute the significance. Multiple testing correction as computed using the False Discovery Rate (FDR) approach. Effect sizes for the differences observe were computed using the Cliff's Delta test. For the GMM differential analyses, we used the same approach as for the MGSs but using the abundances computations.

4.3 Analyses of the untargeted plasma polar metabolites and lipids

4.3.1 Polar metabolites

The plasma samples were stored at −80 °C until analysis. The polar metabolites were analyzed using two-dimensional gas chromatography combined with time-of-flight mass spectrometry (GC×GC-TOFMS, a LECO Pegasus 4D equipped with a consumable-free thermal modulator from LECO Corp). The method has previously been described in detail (36, 43, 47). Specifically, 400 µl methanol and 10 µl internal standard mixture (Heptadecanoic acid-d33, Valine-d8, Glutamic acid-d5 and succinic acid-d4) were added to 30 µl of plasma samples. The samples were vortex mixed and centrifuged for 5 min at 10,000 rpm and half of the supernatant was evaporated to dryness. This was followed by two-step derivatization using methoximation and trimethylsilylation by first adding 25 µl methoxamine (45°C, 60 min) and then 25 µl N-trimethylsilyl-N-methyl trifluoroacetamide (45°C, 60 min). Finally, a retention index standard mixture (n-alkanes) and an injection standard (4,4′ -dibromooctafluorobiphenyl), both in 50 µl hexane, were added to the mixture. The calibration consisted of six points for each quantified metabolite.

The columns were as follows: a phenyl methyl deactivated retention gap (1.5 m × 0.53 mm i.d.) was connected to 10 m × 0.18 mm Rtx-5MS (phase thickness 0.18 µm) and to 1.5 m × 0.1 mm BPX-50 (phase thickness 0.1 µm). Helium was used as the carrier gas at a constant pressure mode (40 psig). A 4-s separation time was used in the second dimension. The temperature program was as follows for the first-dimension column: 50°C (2 min), at 7° C/min to 240°C and at 25°C/min to 300°C (3 min).

The second-dimension column temperature was 15°C higher than the corresponding first-dimension column throughout the program.

ChromaTOF 4.72 vendor software (LECO) was used for within-sample data processing, and the Guineu software (43) was used for alignment, normalization, and peak matching across samples. The normalization was performed by correction with internal standards and specific target metabolites were additionally quantified using external calibration curves. Compounds were identified by comparison to in-house and NIST14 (48) library entries.

Of the 398 total polar metabolites, 143 of them were annotated while the remaining 255 were unannotated. Normalized and log-transformed data was firstly adjusted for individual eGFR (accessible at eGFR Calculator | National Kidney Foundation). The obtained data was then used for the subsequent analyses. We used a combined univariate-multivariate approach to identify relevant features for the discrimination of the T1D individuals from the healthy controls as well as the different T1D subgroups stratified on level of albuminuria. To do so, we used Principal Component Analysis (PCA), Partial Least Squares Discriminant Analysis (PLS-DA) and linear mixed models, with sex, age, BMI, and dietary data as fixed effects. Cliff's Delta effect size was used to determine the difference in metabolite abundance between the two compared groups. All obtained p-values were adjusted for multiple testing with the false discovery (FDR) approach, considering 10% FDR threshold as significant.

We then used the Human Metabolome Database (HMDB) for annotating polar metabolites enabling a functional enrichment analysis with MetaboAnalyst tool (49). To do so, we uploaded a dataset consisting the abundances of normalized, centered, and scaled HMDB-annotated polar metabolites. Then, we selected the metabolic pathway associated metabolite sets to compute the differentially enriched metabolic pathways in study participants.

4.3.2 Plasma lipidomics

The plasma samples were stored at −80 °C until analysis. The Folch procedure (50) was used for sample preparation with minor modifications based on previously published methods at Steno Diabetes Center Copenhagen (13, 51, 52). Briefly, plasma samples were randomized and lipids were extracted from 10 µL plasma using chloroform:methanol (2:1 v/v) following addition of nine different internal standards (stable isotope labelled and non-physiological lipid species). Samples were analyzed in random order in positive electrospray ionization mode using ultra-high-performance liquid chromatography-quadrupole time-of-flight mass spectrometry (UHPLC-Q-TOF-MS) from Agilent Technologies (Santa Clara, CA, USA). The lipidomics data were pre-processed with MZmine2 (53) in which lipids were semi-quantified by normalizing the peak areas to internal standards and corrected for batch effect. Missing values were imputed with the k-nearest neighbour algorithm and all values were log-2-transformed to achieve normal-distributed data.

For the analysis of lipidomics, we used the same approach as outlined above for the polar metabolites. Of the 7,470 total lipids, 476 of them were annotated, while the remaining 6,994 were unannotated. Following adjustment for individual eGFR value, we used weighted gene co-expression network analysis (WGCNA) to cluster all the strongly correlated lipids, in order to reduce the dimensionality of the data. From the original 7,470

lipids we obtained 122 clusters, ranging from 3 lipids to 1,054 lipids. Lipid clusters were annotated considering their lipid content. If all the lipids in the cluster were unannotated, we named the corresponding cluster as "unknown" followed by a number. From here on in lipids cluster analyses, we followed the same approach as described for the polar metabolites.

Individual annotated lipids were used to study the associations between the clinical groups and the lipid specie types with *LipidomeR* R-package, accessible at https://lipidomer.org/.

4.3.3 Metabolite origin assessment

To evaluate whether metabolome features could be related to bacterial metabolism or to the host's metabolism and/or lifestyle factors, we used Least Absolute Shrinkage and Selection Operator (LASSO) modeling. This approach allowed us to identify whether the abundance of metabolome features (polar metabolites, lipids and/or lipid clusters) was better predicted by bioclinical data, QMP or GMM data and/or lifestyle.

4.4 Drug deconfounding of all data sets in the present study

All the datasets used in this project were scrutinized for any drug-associated features. To do so, we used the R package *metadeconfoundR* (available at https://github.com/TillBirkner/metadeconfoundR) (54), with its default parameters.

4.5 Data integration

4.5.1 Multi-omics factor analysis

The normalized datasets were used to investigate interactions and potential signatures involving gut microbiome and plasma metabolome profiles in T1D and its subgroups stratified on albuminuria. Biochemistry data was also combined after normalization (log transformation). Finally, we combined QMP taxonomical counts, GMM CLR-transformed dataset, polar metabolites, lipidomics clusters and biochemistry data for the integration step. We used multi-omics factor analysis (MOFA+) (24) with default parameters, to study the interactions between different layers of data and the potential for identifying combinations of features useful for predicting T1D with varying levels of albuminuria.

4.6 Statistical analysis

All statistical analyses were performed in R software (https://cran.r-project.org/), running version 4.1.0. Significance tests results were corrected for multiple testing with False Discovery Rate (FDR) approach and significance set at 10% FDR threshold. Data visualization was done with *ggplot2* R package. Specific details on the methodology regarding specific data types has been reported in the corresponding section of the methods.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee of the Danish Capital Region (protocol H-15018107). The patients/participants provided their written informed consent to participate in this study.

Author contributions

Conceptualization: OP, PR, TSA, and SW. Methodology: MC-G, TSA, PH, JV YF, LL, ES, MA, TH and CL-Q. Investigation: MC-G, TSA, SW, YF, JV, LL, ES, PH, and CL-Q. Visualization: MC-G. Funding acquisition: PR. Project administration: PR, OP. Supervision: OP, PR. Writing – original draft: MC-G, TSA. Writing – review & editing: All authors. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Novo Nordisk Foundation (grant number NNF14OC0013659) PROTON Personalising treatment of diabetic kidney disease; internal funding was provided by Steno Diabetes Center Copenhagen (SDCC), Herlev, Denmark. TSA was supported by the Novo Nordisk Foundation (Grant # NNF18OC0052457) and SDCC.

Acknowledgments

We thank all the study participants and the SDCC laboratory staff for their contributions.

Conflict of interest

PR reports personal fees from Bayer during the conduct of the study. He has received research support and personal fees from AstraZeneca and Novo Nordisk, and personal fees from Astellas Pharma, Boehringer Ingelheim, Eli Lilly, Gilead Sciences, Mundipharma, Sanofi, and Vifor Pharma. All fees are given to Steno Diabetes Center Copenhagen.

Author MC-G was employed by company LEITAT Technological Center. Author JKV was employed by Clinical Microbiomics.

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fendo.2022.1015557/full#supplementary-material

References

1. Saran R, Robinson B, Abbott KC, Agodoa LYC, Bragg-Gresham J, Balkrishnan R, et al. US Renal data system 2018 annual data report: Epidemiology of kidney disease in the united states. *Am J Kidney Dis* (2019) 73: A7–8. doi: 10.1053/j.ajkd.2019.01.001

2. Bikbov B, Purcell CA, Levey AS, Smith M, Abdoli A, Abebe M, et al. Global, regional, and national burden of chronic kidney disease, 1990–2017: a systematic analysis for the global burden of disease study 2017. *Lancet* (2020) 395:709–33. doi: 10.1016/S0140-6736(20)30045-3

3. Carrero JJ, Grams ME, Sang Y, Ärnlöv J, Gasparini A, Matsushita K, et al. Albuminuria changes are associated with subsequent risk of end-stage renal disease and mortality. *Kidney Int* (2017) 91:244–51. doi: 10.1016/j.kint.2016.09.037

4. Bourlioux P, Koletzko B, Guarner F, Braesco Véronique. The intestine and its microflora are partners for the protection ofthe host: report on the danone symposium "The IntelligentIntestine," held in Paris, June 14, 2002. *Am J Clin Nutr* (2003) 78:675–83. doi: 10.1093/ajcn/78.4.675

5. Vaarala O, Atkinson MA, Neu J. The "perfect storm" for type 1 diabetes: The complex interplay between intestinal microbiota, gut permeability, and mucosal immunity. *Diabetes* (2008) 57:2555–62. doi: 10.2337/db08-0331

6. Anders HJ, Andersen K, Stecher B. The intestinal microbiota, a leaky gut, and abnormal immunity in kidney disease. *Kidney Int* (2013) 83:1010–6. doi: 10.1038/ki.2012.440

7. le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, et al. Richness of human gut microbiome correlates with metabolic markers. *Nature* (2013) 500:541–6. doi: 10.1038/nature12506

8. Zheng P, Li Z, Zhou Z. Gut microbiome in type 1 diabetes: A comprehensive review. *Diabetes Metab Res Rev* (2018) 34:1–9. doi: 10.1002/dmrr.3043

9. Arneth B, Arneth R, Shams M. Metabolomics of type 1 and type 2 diabetes. *Int J Mol Sci* (2019) 20:1–14. doi: 10.3390/ijms20102467

10. Lu J, Chen PP, Zhang JX, Li XQ, Wang GH, Yuan BY, et al. GPR43 deficiency protects against podocyte insulin resistance in diabetic nephropathy through the restoration of AMPKα activity. *Theranostics* (2021) 11:4728–42. doi: 10.7150/thno.56598

11. Kikuchi K, Saigusa D, Kanemitsu Y, Matsumoto Y, Thanai P, Suzuki N, et al. Gut microbiome-derived phenyl sulfate contributes to albuminuria in diabetic kidney disease. *Nat Commun* (2019) 10:1835. doi: 10.1038/s41467-019-09735-4

12. Winther SA, Henriksen P, Vogt JK, Hansen TH, Ahonen L, Suvitaival T, et al. Gut microbiota profile and selected plasma metabolites in type 1 diabetes without and with stratification by albuminuria. *Diabetologia* (2020) 63:2713–24. doi: 10.1007/s00125-020-05260-y

13. Tofte N, Suvitaival T, Ahonen L, Winther SA, Theilade S, Frimodt-Møller M, et al. Lipidomic analysis reveals sphingomyelin and phosphatidylcholine species associated with renal impairment and all-cause mortality in type 1 diabetes. *Sci Rep* (2019) 9:1–10. doi: 10.1038/s41598-019-52916-w

14. Mäkinen VP, Tynkkynen T, Soininen P, Forsblom C, Peltola T, Kangas AJ, et al. Sphingomyelin is associated with kidney disease in type 1 diabetes (The FinnDiane study). *Metabolomics* (2012) 8:369–75. doi: 10.1007/s11306-011-0343-y

15. Klein RL, Hammad SM, Baker NL, Hunt KJ, al Gadban MM, Cleary PA, et al. Decreased plasma levels of select very long chain ceramide species are associated with the development of nephropathy in type 1 diabetes. *Metabolism* (2014) 63:1287–95. doi: 10.1016/j.metabol.2014.07.001

16. Mäkinen VP, Tynkkynen T, Soininen P, Peltola T, Kangas AJ, Forsblom C, et al. Metabolic diversity of progressive kidney disease in 325 patients with type 1 diabetes (the FinnDiane study). *J Proteome Res* (2012) 11:1782–90. doi: 10.1021/pr201036j

17. Kim JK, Shin S-Y, Moon JS, Li L, Cho SK, Kim T-J, et al. Isolation of dextran-hydrolyzing intestinal bacteria and characterization of their dextranolytic activities. *Biopolymers* (2015) 103:321–7. doi: 10.1002/bip.22615

18. Vieira-Silva S, Falony G, Darzi Y, Lima-Mendez G, Garcia Yunta R, Okuda S, et al. Species-function relationships shape ecological properties of the human gut microbiome. *Nat Microbiol* (2016) 1:16088. doi: 10.1038/nmicrobiol.2016.88

19. Argelaguet R, Arnol D, Bredikhin D, Deloro Y, Velten B, Marioni JC, et al. MOFA+: A statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol* (2020) 21:1–17. doi: 10.1186/s13059-020-02015-1

20. Durack J, v. Lynch S. The gut microbiome: Relationships with disease and opportunities for therapy. *J Exp Med* (2019) 216:20–40. doi: 10.1084/jem.20180448

21. Parada Venegas D, de la Fuente MK, Landskron G, González MJ, Quera R, Dijkstra G, et al. Corrigendum: Short chain fatty acids (SCFAs)-mediated gut epithelial and immune regulation and its relevance for inflammatory bowel diseases. *Front Immunol* (2019) 10:1486. doi: 10.3389/fimmu.2019.01486

22. Qin Q, Yan S, Yang Y, Chen J, Li T, Gao X, et al. A metagenome-wide association study of the gut microbiome and metabolic syndrome. *Front Microbiol* (2021) 12:682721. doi: 10.3389/fmicb.2021.682721

23. Bell A, Juge N. Mucosal glycan degradation of the host by the gut microbiota. *Glycobiology* (2021) 31:691–6. doi: 10.1093/glycob/cwaa097

24. Yao Y, Yan L, Chen H, Wu N, Wang W, Wang D. Cyclocarya paliurus polysaccharides alleviate type 2 diabetic symptoms by modulating gut microbiota and short-chain fatty acids. *Phytomedicine* (2020) 77:153268. doi: 10.1016/j.phymed.2020.153268

25. Shkoporov AN, v. Chaplin A, Shcherbakova VA, Suzina NE, Kafarskaia LI, Bozhenko VK, et al. Ruthenibacterium lactatiformans gen. nov., sp. nov., an anaerobic, lactate-producing member of the family ruminococcaceae isolated from human faeces. *Int J Syst Evol Microbiol* (2016) 66:3041–9. doi: 10.1099/ijsem.0.001143

26. Brouwers MCGJ, Ham JC, Wisse E, Misra S, Landewe S, Rosenthal M, et al. Elevated lactate levels in patients with poorly regulated type 1 diabetes and glycogenic hepatopathy: A new feature of mauriac syndrome. *Diabetes Care* (2015) 38:e11–2. doi: 10.2337/dc14-2205

27. Li H, Xu H, Li Y, Jiang Y, Hu Y, Liu T, et al. Alterations of gut microbiota contribute to the progression of unruptured intracranial aneurysms. *Nat Commun* (2020) 11:3218. doi: 10.1038/s41467-020-16990-3

28. Kim KS, Oh DH, Kim JY, Lee BG, You JS, Chang KJ, et al. Taurine ameliorates hyperglycemia and dyslipidemia by reducing insulin resistance and leptin level in otsuka long-Evans tokushima fatty (OLETF) rats with long-term diabetes. *Exp Mol Med* (2012) 44:665–73. doi: 10.3858/emm.2012.44.11.075

29. Ito T, Schaffer SW, Azuma J. The potential usefulness of taurine on diabetes mellitus and its complications. *Amino Acids* (2012) 42:1529–39. doi: 10.1007/s00726-011-0883-5

30. Johnson EL, Heaver SL, Waters JL, Kim BI, Bretin A, Goodman AL, et al. Sphingolipids produced by gut bacteria enter host metabolic pathways impacting ceramide levels. *Nat Commun* (2020) 11:1–11. doi: 10.1038/s41467-020-16274-w

31. Chaurasia B, Tippetts TS, Monibas RM, Liu J, Li Y, Wang L, et al. Targeting a ceramide double bond improves insulin resistance and hepatic steatosis. *Science* (2019) 365:386–92. doi: 10.1126/science.aav3722

32. Xie C, Jiang C, Shi J, Gao X, Sun D, Sun L, et al. An intestinal farnesoid x receptor-ceramide signaling axis modulates hepatic gluconeogenesis in mice. *Diabetes* (2017) 66:613–26. doi: 10.2337/db16-0663

33. Cohen LJ, Esterhazy D, Kim SH, Lemetre C, Aguilar RR, Gordon EA, et al. Commensal bacteria make GPCR ligands that mimic human signalling molecules. *Nature* (2017) 549:48–53. doi: 10.1038/nature23874

34. Heaver SL, Johnson EL, Ley RE. Sphingolipids in host–microbial interactions. *Curr Opin Microbiol* (2018) 43:92–9. doi: 10.1016/j.mib.2017.12.011

35. Karaki SI, Mitsui R, Hayashi H, Kato I, Sugiya H, Iwanaga T, et al. Short-chain fatty acid receptor, GPR43, is expressed by enteroendocrine cells and mucosal mast cells in rat intestine. *Cell Tissue Res* (2006) 324:353–60. doi: 10.1007/s00441-005-0140-x

36. Tofte N, Suvitaival T, Trost K, Mattila IM, Theilade S, Winther SA, et al. Metabolomic assessment reveals alteration in polyols and branched chain amino acids associated with present and future renal impairment in a discovery cohort of 637 persons with type 1 diabetes. *Front Endocrinol (Lausanne)* (2019) 10:818. doi: 10.3389/fendo.2019.00818

37. Hansen CS, Suvitaival T, Theilade S, Mattila I, Lajer M, Trošt K, et al. Cardiovascular autonomic neuropathy in type 1 diabetes is associated with disturbances in TCA, lipid, and gucose metabolism. *Front Endocrino* (2022) 13, 831793. doi: 10.3389/fendo.2022.831793

38. Maier L, Pruteanu M, Kuhn M, Zeller G, Telzerow A, Anderson EE, et al. Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature* (2018) 555:623–8. doi: 10.1038/nature25979

39. Zimmermann M, Zimmermann-Kogadeeva M, Wegmann R, Goodman AL. Mapping human microbiome drug metabolism by gut bacteria and their genes. *Nature* (2019) 570:462–7. doi: 10.1038/s41586-019-1291-3

40. Hiltunen TP, Rimpelä JM, Mohney RP, Stirdivant SM, Kontula KK. Effects of four different antihypertensive drugs on plasma metabolomic profiles in patients with essential hypertension. *PloS One* (2017) 12:1–16. doi: 10.1371/journal.pone.0187729

41. Leshem A, Segal E, Elinav E. The gut microbiome and individual-specific responses to diet. *mSystems* (2020) 5:1–12. doi: 10.1128/mSystems.00665-20

42. Mailing LJ, Allen JM, Buford TW, Fields CJ, Woods JA. Exercise and the gut microbiome: A review of the evidence, potential mechanisms, and implications for human health. *Exerc Sport Sci Rev* (2019) 47:75–85. doi: 10.1249/JES.0000000000000183

43. Pedersen HK, Gudmundsdottir V, Nielsen HB, Hyotylainen T, Nielsen T, Chatelier L, et al. Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature* (2016) 535:376–81. doi: 10.1038/nature18646

44. Nielsen HB, Almeida M, Juncker AS, Rasmussen S, Li J, Sunagawa S, et al. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat Biotechnol* (2014) 32:822–8. doi: 10.1038/nbt.2939

45. Vandeputte D, Kathagen G, D'Hoe K, Vieira-Silva S, Valles-Colomer M, Sabino J, et al. Quantitative microbiome profiling links gut community variation to microbial load. *Nature* (2017) 551:507–11. doi: 10.1038/nature24460

46. Camarillo-Guerrero LF, Almeida A, Rangel-Pineros G, Finn RD, Lawley TD. Massive expansion of human gut bacteriophage diversity. *Cell* (2021) 184:1098–1109.e9. doi: 10.1016/j.cell.2021.01.029

47. Castillo S, Mattila I, Miettinen J, Orešič M, Hyötyläinen T. Data analysis tool for comprehensive two-dimensional gas chromatography/time-of-flight mass spectrometry. *Anal Chem* (2011) 83:3058–67. doi: 10.1021/ac103308x

48. Bowden JA, Heckert A, Ulmer CZ, Jones CM, Koelmel JP, Abdullah L, et al. Harmonizing lipidomics: NIST interlaboratory comparison exercise for lipidomics using SRM 1950-metabolites in frozen human plasma. *J Lipid Res* (2017) 58:2275–88. doi: 10.1194/jlr.M079012

49. Pang Z, Chong J, Zhou G, de Lima Morais DA, Chang L, Barrette M, et al. MetaboAnalyst 5.0: Narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res* (2021) 49:W388–96. doi: 10.1093/nar/gkab382

50. Folch J, Lees M, Stanley GHS. A simple method for the isolation and purification of total lipides from animal tissues. *J Biol Chem* (1957) 226:497–509. doi: 10.1016/S0021-9258(18)64849-5

51. Luukkonen PK, Zhou Y, Nidhina Haridas PA, Dwivedi OP, Hyötyläinen T, Ali A, et al. Impaired hepatic lipid synthesis from polyunsaturated fatty acids in TM6SF2 E167K variant carriers with NAFLD. *J Hepatol* (2017) 67:128–36. doi: 10.1016/j.jhep.2017.02.014

52. O'Gorman A, Suvitaival T, Ahonen L, Cannon M, Zammit S, Lewis G, et al. Identification of a plasma signature of psychotic disorder in children and adolescents from the Avon longitudinal study of parents and children (ALSPAC) cohort. *Transl Psychiatry* (2017) 7:e1240. doi: 10.1038/tp.2017.211

53. Pluskal T, Castillo S, Villar-Briones A, Orešič M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinf* (2010) 11:395. doi: 10.1186/1471-2105-11-395

54. Forslund SK, Chakaroun R, Zimmermann-Kogadeeva M, Markó L, Aron-Wisnewsky J, Nielsen T, et al. Combinatorial, additive and dose-dependent drug–microbiome associations. *Nature* (2021) 600:500–5. doi: 10.1038/s41586-021-04177-9

frontiers | Frontiers in Endocrinology

TYPE Original Research
PUBLISHED 16 December 2022
DOI 10.3389/fendo.2022.1054206

OPEN ACCESS

EDITED BY
Lu Hua,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

REVIEWED BY
Ya-Li Chen,
Second Hospital of Hebei Medical University, China
Ping Liu,
The Second Hospital of Shandong University, China

*CORRESPONDENCE
Jia Tang
tj676533@163.com
Wenting Wang
wangwentingtn@sina.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Bone Research, a section of the journal Frontiers in Endocrinology

RECEIVED 26 September 2022
ACCEPTED 23 November 2022
PUBLISHED 16 December 2022

CITATION
Yuan J, Xiong X, Zhang B, Feng Q, Zhang J, Wang W and Tang J (2022) Genetically predicted C-reactive protein mediates the association between rheumatoid arthritis and atlantoaxial subluxation. *Front. Endocrinol.* 13:1054206. doi: 10.3389/fendo.2022.1054206

Genetically predicted C-reactive protein mediates the association between rheumatoid arthritis and atlantoaxial subluxation

Jiaqin Yuan[1,2†], Xiaoqin Xiong[3†], Bin Zhang[4†], Qingyuan Feng[5], Jinglin Zhang[6], Wenting Wang[7*] and Jia Tang[8*]

[1]Department of Orthopedics, The Second People's Hospital of Yibin, Yibin, Sichuan, China, [2]Department of Orthopedics, Yibin Hospital, West China Hospital of Sichuan University, Yibin, Sichuan, China, [3]Department of Pediatrics, The Affiliated Hospital of Southwest Medical University, Luzhou, China, [4]Rheumatism Immunity Branch, Weifang People's Hospital, Weifang, Shandong, China, [5]Wuxi School of Medicine, Jiangnan University, Wuxi, Jiangsu, China, [6]Department of Occupational Disease, Yibin Center for Disease Control and Prevention, Yibin, Sichuan, China, [7]Department of Anesthesiology, The Second Affiliated Hospital of Hainan Medical University, Haikou, China, [8]Department of Pediatrics, Daping Hospital, Army Medical University, Chongqing, China

Objective: Investigating the causal relationship between rheumatoid arthritis (RA) and atlantoaxial subluxation (AAS) and identifying and quantifying the role of C-reactive protein (CRP) as a potential mediator.

Methods: Using summary-level data from a genome-wide association study (GWAS), a two-sample Mendelian randomization (MR) analysis of genetically predicted rheumatoid arthritis (14,361 cases, and 43,923 controls) and AAS (141 cases, 227,388 controls) was performed. Furthermore, we used two-step MR to quantitate the proportion of the effect of c-reactive protein-mediated RA on AAS.

Results: MR analysis identified higher genetically predicted rheumatoid arthritis (primary MR analysis odds ratio (OR) 0.61/SD increase, 95% confidence interval (CI) 1.36-1.90) increased risk of AAS. There was no strong evidence that genetically predicted AAS had an effect on rheumatoid arthritis risk (OR 1.001, 95% CI 0.97-1.03). The proportion of genetically predicted rheumatoid arthritis mediated by C-reactive protein was 3.7% (95%CI 0.1%–7.3%).

Conclusion: In conclusion, our study identified a causal relationship between RA and AAS, with a small proportion of the effect mediated by CRP, but a majority of the effect of RA on AAS remains unclear. Further research is needed on additional risk factors as potential mediators. In clinical practice, lesions of the upper cervical spine in RA patients need to be given more attention.

KEYWORDS

Mendelian randomization, rheumatoid arthritis, C-reactive protein, atlantoaxial subluxation, upper cervical instability

Introduction

Rheumatoid arthritis (RA) is a chronic inflammatory immune system disease characterized by synovitis and cartilage destruction, which mainly affects the synovial membrane, tendon sheaths and synovial bursae of the joints (1). It mainly manifests as clinical symptoms such as joint pain, stiffness, swelling, deformity, and dysfunction (2). Its global prevalence is approximately 1% and ranks 42nd among disabling diseases worldwide (3). As the global population ages, the prevalence continues to increase. Due to RA's high mortality and morbidity rates, patients' quality of life is lower, and the economic burden on society is greater. The National Audit Office (NAO) reports that RA costs the UK approximately £560 million a year in health care costs, not including the cost of sick leave and work-related disability (4).

The active segment of the cervical vertebra is the basic functional unit of the cervical spine. It consists of two adjacent cervical vertebrae and their attached soft tissues and is the smallest functional unit of the cervical vertebra. Cervical instability refers to excessive or abnormal cervical spine movement that cannot maintain the normal position between the vertebral bodies under physiological loads (5). Atlantoaxial subluxation (AAS) in RA patients mostly involves the atlantoaxial joint, which may be caused by head and neck trauma, congenital diseases (bone dysplasia), autoimmune diseases (rheumatoid arthritis), etc. However, the exact reason is not yet clear. Observational studies have shown that upper cervical instability occurs in 29.6% of RA patients, of which atlantoaxial subluxation accounts for 24.6% (6). However, epidemiological studies may suffer from measurement error, uncontrolled confounding factors, and reverse causality. Ultimately, the results may be subject to various biases. Therefore, a design is needed to avoid or reduce some biases further to demonstrate the causal relationship between RA and AAS.

Moreover, potential pathways related to RA and AAS have not been investigated. Previous studies have provided evidence that C-reactive protein (CRP) is elevated in both RA and AAS (7, 8). Consequently, CRP might be a potential mediator between RA and AAS.

Mendelian randomization (MR) is a potential causal inference method that uses genetic variation as an instrumental variable to obtain the effect of exposure factors on outcomes from observational data (9). MR can reduce the effects of nonmeasurement errors or confounding factors while avoiding reverse causality through Mendelian inheritance laws (9). Therefore, we aimed to (i) determine whether RA is causally related to AAS and (ii) assess the extent to which CRP mediates the effects of RA on AAS.

Methods

Study design

The data used in our analysis were publicly available and were approved by the institutional review committee in the respective studies. Therefore, no further sanctions were needed. All generated results are presented in the article and its supplements.

In this study, we explored the reciprocal causal relationship between rheumatoid arthritis and atlantoaxial subluxation by two-sample, bidirectional mendelian randomization. In our study, single nucleotide polymorphisms (SNPs) were defined as instrumental variables (IVs) (10).

GWAS summary data sources

The data used in our study were all publicly available, and the participants in the GWAS were of European ancestry. The genetic associations of RA were derived from a GWAS meta-analysis by Ha and colleagues (11), which included 14,361 RA case and 43,923 controls. All cases met the 1987 American College of Rheumatology criteria or were diagnosed as RA by a rheumatologist. ninety-one percent of individuals were serologically positive for anti-CCP antibodies or rheumatoid factor. Additional details are shown in Supplementary Table S1.

Data on AAS were drawn from the GWAS summary data sources on the FennGenn consortium, which is available at https://www.finngen.fi/en (AAS including 141 cases and 227,388 participants) (12). Individuals with ICD codes [ICD-10 M43.3 "Recurrent atlantoaxial subluxation with myelopathy" and ICD-10 M43.4 "Other recurrent atlantoaxial subluxation"] were characterized as AAS cases.

Summary statistics on CRP levels were obtained from a published GWAS meta-analysis that included 78 studies of European ancestry, with the largest sample size thus far (sample size = 204,402) (13). The study design, such as sample collection, quality control procedures, and imputation methods, were described in the original publication. Additional details are shown in Supplementary Table S2. All GWAS data are from different consortia or organizations, and thus there is no sample overlap.

Instrumental variable selection and data harmonization

We included SNPs that were genome-wide significant ($P < 5 \times 10^{-8}$). If there were no significant genome-wide SNPs as IVs, SNPs with less than a genome-wide significance level ($P < 5 \times 10^{-6}$) were used as candidate IVs. Then, these SNPs were

clustered based on linkage disequilibrium (window size = 10,000 kb and $r^2 < 0.001$). Estimated levels of linkage disequilibrium from the 1000 Genomes Project based on European samples (14). If a particular exposed SNP was not present in the outcome dataset, proxy SNPs were used by LD tagging. Palindromic and ambiguous SNPs were excluded from IVs for Mendelian randomization analysis (15). The F statistic was calculated by the variance explained by SNPs for each exposure, i.e. $[(N - K - 1)/K]/[R^2/(1 - R^2)]$, where K is the number of genetic variants, N is the sample size. We removed weak instrumental variables (F-statistics < 10) (16, 17).

Statistical analysis

We performed MR analysis using R software (version 4.2.0, http://www.r-project.org) and the "Two-Sample MR" package (version 0.5.6) (18). MR-Pleiotropy RESidual Sum and Outlier (MR-PRESSO) and robust adjusted profile score (MR.RAPS) were performed using the R packages "MRPRESSO" and "MR.raps", respectively. Calculation of statistical power for Mendelian randomization was performed using mRnd (https://cnsgenomics.shinyapps.io/mRnd/). And we applied a PhenoScanner search to assess all known phenotypes related to the considered genetic instruments in our analyses.

Primary analysis

Figure 1 shows a schematic summary of the analysis. We conducted a two-sample bidirectional MR to evaluate the mutual causality between RA and AAS (Figure 1A), which was designated as the total effect.

Inverse variance weighting (IVW) uses meta-analysis to combine the Wald ratios of causal effects for each SNP (15, 19). Then, MR-Egger (20) and weighted-median (21) methods were used as a complement to IVW. Different methods adapted to different validity assumptions were applied to obtain MR estimates. The application of IVW is based on the premise that all SNPs are valid instrumental variables. Therefore, this method can obtain accurate estimation results. MR-Egger assesses directional pleiotropy for instrumental variables, where the intercept can be interpreted as an estimate of the average pleiotropy of genetic variation. The weighted median has the advantage of maintaining higher precision (smaller standard deviation) compared to the MR-Egger analysis. In the presence of horizontal pleiotropy, the weighted median provides a consistent estimate even if 50% of the genetic variants are invalid IVs (22).

Mediation analysis

We further performed a mediation analysis using a two-step MR design to explore whether CRP mediates the causal pathway from RA to AAS outcome (Figure 1B). The overall effect can be decomposed into an indirect effect (through mediators) and a direct effect (without mediators) effect (23). The total effect of RA on AAS was decomposed into 1) direct effects of RA on AAS (*c'* in Figure 1B) and 2) indirect effects mediated by RA through the mediator ($a \times b$ in Figure 1B). We calculated the percentage mediated by the mediating effect by dividing the indirect effect

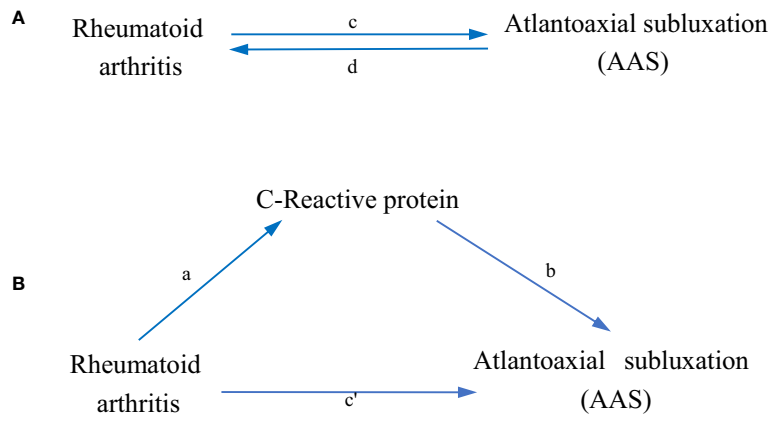


FIGURE 1
Diagrams illustrating associations examined in this study. **(A)** The total effect between rheumatoid arthritis(RA) and atlantoaxial subluxation(AAS). *c* is the total effect using genetically predicted RA as exposure and AAS as outcome. *d* is the total effect using genetically predicted AAS as exposure and RA as outcome. **(B)** The total effect was decomposed into: (i) indirect effect using a two-step approach (where *a* is the total effect of RA on CRP, and *b* is the effect of CRP on AAS) and the product method ($a \times b$) and (ii) direct effect ($c' = c - a \times b$). Proportion mediated was the indirect effect divided by the total effect.

by the total effect. Meanwhile, 95% confidence intervals were calculated with the delta method (24).

Sensitivity analysis

The causal direction of each extracted SNP to exposure and outcome was tested by using MR Steiger filtering (25). This method calculates the variance explained in exposure and results from the instrumental SNPs and tests whether the variance in the results is less than the exposure. "TRUE" MR Steiger results indicate causality in the expected direction, while "FALSE" results indicate causality in the opposite direction. We excluded SNPs with 'FALSE' results, indicating that it showed evidence of a major effect on the outcome rather than exposure.

Heterogeneity between SNPs was assessed using Cochran's Q statistic and funnel plots (26, 27). Horizontal pleiotropy was detected using the MR-Egger intercept (20) method and the MR-PRESSO (28) method. If outliers were detected, they were removed, and we re-evaluated the MR causal estimates. If heterogeneity remained high after removal, the stability of the results was assessed using a random effects model, which is less susceptible to weaker SNP exposure associations. Finally, leave-one-out analysis was used to validate the effect of each SNP on the overall causal estimates.

Results

Association of RA with AAS

After removing palindromic and ambiguous SNPs, SNPs without proxy and SNPs with wrong causal directions identified by MR Steiger filtering, there were 80 SNPs in RA and 4 SNPs in AAS as instrumental variables (Supplementary Tables S3, S4). Since AAS did not reach the level of gene-wide significance for SNPs, SNPs with less than genome-wide significance ($P < 5 \times 10^{-6}$) were used as instrumental variables. The variance explained by and the F-statistic for SNPs instrumenting RA exposure were 5.8% and 45, respectively. Our study provides 100% power to detect the causal effect of RA on AAS risk.

IVW, MR-Egger, and weighted median regression were used to estimate the causal relationship between genetically predicted RA and AAS (Figures 2, 3). Across all three MR methods, there was broad and consistent support for the positive association of RA with AAS (IVW odds ratio [OR] per SD increase in RA = 1.61 [95% CI, 1.36-1.90], P < 0.0001; MR-Egger OR per SD increase in RA = 1.66 [95% CI, 1.30-2.13], P < 0.001; weighted median OR per SD increase in RA = 1.80 [95% CI, 1.39-2.34], P < 0.0001). However, the results of our MR analysis showed no reverse causality for genetically predicted RA on AAS (i.e., no causality for genetically predicted AAS on RA.). The OR was 1.001 [95% CI, 0.97-1.03; p = 0.87] by using the IVW method. The results are shown in Figure 3.

Association of RA with CRP

We extracted a total of 52 genome-wide significant SNPs as instrumental variables after removing palindromic and ambiguous SNPs, SNPs without proxies, and SNPs in the wrong causal direction identified by MR Steiger filtering (Supplementary Table S5). The variance explained by and F-statistic for SNPs instrumenting RA exposure were 3.3% and 39, respectively.

According to the IVW, MR–Egger and weighted median methods, genetically predicted RA was found to be positively associated with CRP risk (IVW method, OR, 1.02; [95% CI, 1.01-1.03], P<0.0001; MR-Egger method, OR, 1.04; [95% CI, 1.02-1.05], P<0.0001; weighted median method, OR, 1.03; [95% CI, 1.02-1.04], P<0.0001). The results are shown in Figure 3.

Association of CRP with AAS

Genetic instruments for CRP explained 1.3% of its variance, with an F-statistic of 54. As shown in Supplementary Table S6, we presented all genetic instruments associated with CRP at the genome-wide significance level ($P < 5 \times 10^{-8}$). As shown in Figure 3, genetically predicted CRP was significantly positively correlated with AAS [OR=2.38, 95% CI, 1.03-5.51; P=0.04] by using the IVW method. The estimation directions of these three methods, IVW, MR-Egger and weighted median, were consistent.

Proportion of the association between RA and AAS mediated by CRP

We analyzed CRP as a mediator of the pathway from RA to AAS. We found that RA was associated with increased CRP, which in turn was associated with an increased risk of AAS. As shown in Figure 4, our study showed that CRP accounted for 3.7% of the increased risk of AAS associated with RA (proportion mediated: 3.7%; 95% CI = 0.1%–7.3%).

Sensitivity analysis

Several sensitivity analyses were used to examine and correct for the presence of pleiotropy in causal estimates. Cochran's Q-test and funnel plot showed no evidence of heterogeneity and asymmetry between these SNPs in the causal relationship between these SNPs (Supplementary Table S7 and Supplementary Figure S1). In our study, the MR-Egger intercept showed weak evidence of pleiotropy

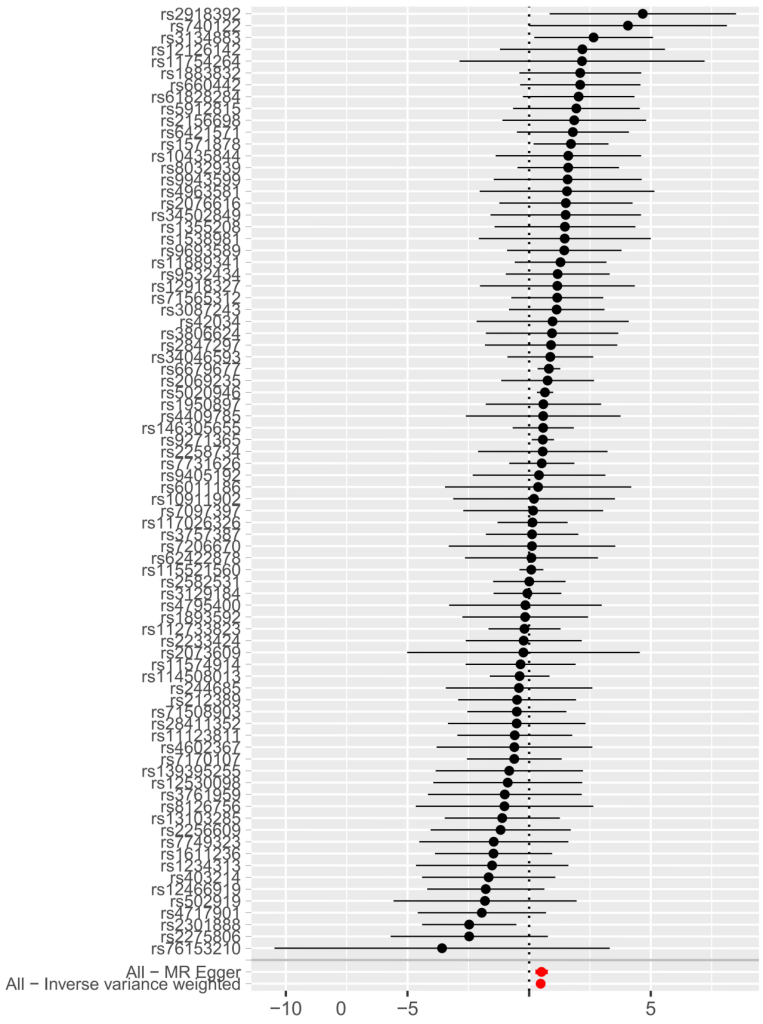


FIGURE 2
Forest plot to visualize causal effect of each single SNP on total AAS risk.

Exposure	MR method	No.of SNPs	Forest plot	OR (95% CI)	P-value
RA on CRP	MR–Egger	52		1.04(1.02-1.05)	<0.0001
	Weighted median	52		1.03(1.02-1.04)	<0.0001
	Inverse variance weighted	52		1.02(1.01-1.03)	<0.0001
AAS on RA	MR–Egger	4		0.98(0.94-1.01)	0.31
	Weighted median	4		0.99(0.96-1.02)	0.90
	Inverse variance weighted	4		1.01(0.97-1.03)	0.86
RA on AAS	MR–Egger	80		1.66(1.30-2.13)	<0.001
	Weighted median	80		1.80(1.39-2.34)	<0.0001
	Inverse variance weighted	80		1.61(1.36-1.90)	<0.0001
CRP on AAS	MR–Egger	49		2.82(0.63-12.6)	0.18
	Weighted median	49		2.28(0.69-7.57)	0.17
	Inverse variance weighted	49		2.38(1.03-5.51)	0.04

FIGURE 3
Forest plot to visualize the causal effects of CRP with RA and AAS.

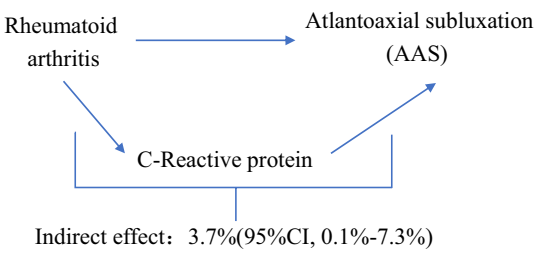


FIGURE 4
Schematic diagram of the CRP mediation effect.

at the directional level of the RA instrument (OR=0.996; 95% CI, 0.994-0.998; P=0.0001)), although no other MR-Egger intercepts did (Supplementary Table S7). We did not detect potential horizontal pleiotropy by using the MR-PRESSO global test (Supplementary Table S7). The effect of each SNP on the overall causal estimates was verified by leave-one-out analysis (Supplementary Figure S2). After removing each SNP, we systematically performed the MR analysis again for the remaining SNPs. The results remained consistent, indicating that all SNPs were calculated to make the causal relationship significant.

Discussion

Recent studies (6, 29, 30) have examined the relationship between RA and AAS. However, the current evidence is limited to observational studies, and the results may be influenced by confounding factors. Our study aimed to illustrate the causal effects between RA and AAS. We used MR analysis to investigate the association between RA and AAS based on existing GWAS and to demonstrate whether the causal relationship between them is mediated through CRP. Our results suggested that genetically predicted RA was associated with an increased risk of AAS (61% increased risk of AAS for every 1 SD increase in RA), and 3.7% of this effect was mediated through CRP.

To date, we are the first to investigate the causal relationship between RA and the risk of instability in the upper cervical spine by MR methods, while also demonstrating CRP as their mediator. Our findings are consistent with previous findings from traditional observational designs. Yurube et al. (31) showed that atlantoaxial instability occurred in 43.6% of RA patients, with atlantoaxial subluxation being the most common (at least 32.1%) through a prospective follow-up study of RA patients with no initial cervical involvement for at least 5 years. Similarly, in a retrospective study of 240 RA patients, Kotecki et al. (32) showed that the incidence of cervical spine involvement in RA patients was 75%, the most common lesion was anterior atlantoaxial subluxation (approximately 58%), and C-reactive protein levels increased (OR, 19.0; 95% CI, 7.0–32.0; P = 0.016). However, both studies were of observational design. First, they had low response rates between the two groups, and second, their results were more influenced by reverse causality or other potential mixed effects than MR analyses.

Atlantoaxial instability refers to the structural changes and dysfunction of the atlanto-occipital joint and atlantoaxial joint between the atlas, the axis and the base of the skull due to various reasons (e.g., deformity, trauma, degeneration, tumor and infection), which lead to excessive or abnormal activity or abnormal position under physiological load. Atlantoaxial joint instability and dislocation are rare in normal adults and are mostly secondary to trauma and disease. Cervical instability can be further divided into atlantoaxial subluxation (AAS), vertical subluxation (VS) and subaxial subluxation (SAS). AAS is the most common, followed by VS, and SAS is less common (6). VS usually occurs after AAS. VS is considered a serious condition in RA patients because it can lead to sudden death (33). Synovitis is the initial link of rheumatoid arthritis, and it is also the basic pathological change. The characteristics of multiple synovial sacs of the atlantoaxial joint provide conditions for its involvement. At the same time, synovial tissue macrophages produce tumor necrosis factor to promote the inflammatory response of the atlantoaxial joint (34). Sorimachi et al. (35) have suggested that synovitis invades the atlantoaxial joint in three stages. First, the medial and lateral atlantoaxial joints are invaded, the joint capsule is destroyed, and the joint capsule is swollen and exuded; then, the synovium begins to proliferate, and the ligaments are edematous and destroyed, after which finally, it erodes hyaline cartilage and penetrates into subchondral bone to produce bone tissue destruction. In addition, the stability of the atlantoaxial joint mainly depends on the maintenance of the transverse ligament and other ligaments, which are characterized by high stiffness and insufficient willfulness. Another characteristic product of RA is pannus, which not only blocks the bone from obtaining nutrition through the synovium but also grows to the cartilage surface in the joint cavity, produces adhesions, and locally releases more inflammatory factors, proteolytic enzymes, etc. (36). When inflammation involves the transverse ligament, it not only destroys the fibrous structure and relaxes the ligament but also erodes the odontoid process and causes erosion and rupture near

the attachment point, which finally leads to the instability of the atlantoaxial joint (37).

CRP, a member of the pentraxin family of proteins, consists of five 23 kDa subunits that can be increased 1,000-fold or more during infection, inflammation and tissue damage. Although hepatocytes are the main source of CRP, other cells, such as monocytes and lymphocytes, also produce CRP (38). Fang et al. (39) suggested that synovial tissue from RA patients also produces CRP. Therefore, one of the reasons for the increased CRP concentrations in synovial fluid and serum CRP levels in RA patients may be the local production of CRP in inflammatory synovial tissue (40). On the one hand, the interaction of CRP with Fcγ receptor I and FcγRIIA promotes the production of proinflammatory cytokines, leading to an amplification loop of the inflammatory response; on the other hand, CRP initiates bone destruction by inducing the receptor activator of nuclear factor-κB ligand protein and directly stimulating osteoclast generation, thus causing a vicious cycle between inflammation and bone destruction in RA (41). Therefore, CRP contributes to atlantoaxial joint instability by mediating synovial inflammation and bone destruction in RA.

Our findings also suggest that RA may increase the risk of atlantoaxial subluxation through other important mediators. Zhang et al. (6) showed that low hemoglobin levels may be associated with atlantoaxial instability in RA. This may be because low hemoglobin levels are partly a chronic inflammatory manifestation of the disease and are thought to be associated with joint damage in RA (42), thus showing a correlation with cervical instability. In addition, $CD5^+$ B cells in RA patients can produce IgG with the help of T lymphocytes, and rheumatoid factor and IgG form immune complexes deposited in the synovium, which are blocked during the clearance process, resulting in bone destruction and fusion.

This study has several limitations. First, our analysis was performed using the European population, which limits its prevalence (43). Second, the smaller number of cases in AAS is in the GWAS dataset of AAS, and it is hoped that larger GWAS data will be available for validation in the future. Third, even if we took steps to identify and eliminate outlier variants, we cannot exclude the possibility that horizontal pleiotropy influenced our results. Fourth, we used summary-level statistics in our study, not individual-level data. Therefore, we cannot further explore causal links between subgroups such as females and males. Fifth, our study demonstrates that genetic prediction of rheumatoid arthritis mediated by C-reactive protein is 3.7%, which is very low. Thus, more studies are needed to quantify other mediators.

Conclusion

In conclusion, our study identified a causal relationship between RA and atlantoaxial subluxation, with a small proportion of the effect mediated by CRP, but a majority of the effect of RA on atlantoaxial subluxation remains unclear. Further research is needed on additional risk factors as potential mediators. In clinical practice, lesions of the upper cervical spine in RA patients need to be given more attention.

Data availability statement

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

Author contributions

All authors designed this study. JT, WW, and XX performed the catalog and literature search and data extraction with suggestions and help from JY. JY, JZ, and XX performed the statistical analyses. All authors contributed to the data interpretation and manuscript writing. All authors contributed to the article and approved the submitted version.

Acknowledgments

Summary statistics for the genetic associations with RA, CRP, and AAS were obtained from GWAS respectively by Ha et al., Ligthart et al. and FennGenn consortium. We thank all investigators for sharing the genome-wide summary statistics.

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/fendo.2022.1054206/full#supplementary-material

References

1. Smolen JS, Aletaha D, McInnes IB. Rheumatoid arthritis. *Lancet (London England)* (2016) 388(10055):2023–38. doi: 10.1016/S0140-6736(16)30173-8

2. Sharif K, Sharif A, Jumah F, Oskouian R, Tubbs RS. Rheumatoid arthritis in review: Clinical, anatomical, cellular and molecular points of view. *Clin Anat (New York NY)* (2018) 31(2):216–23. doi: 10.1002/ca.22980

3. Cross M, Smith E, Hoy D, Carmona L, Wolfe F, Vos T, et al. The global burden of rheumatoid arthritis: estimates from the global burden of disease 2010 study. *Ann Rheum Dis* (2014) 73(7):1316–22. doi: 10.1136/annrheumdis-2013-204627

4. Fazal SA, Khan M, Nishi SE, Alam F, Zarin N, Bari MT, et al. A clinical update and global economic burden of rheumatoid arthritis. *Endocr Metab Immune Disord Drug Targets* (2018) 18(2):98–109. doi: 10.2174/1871530317666171114122417

5. Ng HW, Teo EC, Lee KK, Qiu TX. Finite element analysis of cervical spinal instability under physiologic loading. *J Spinal Disord Tech* (2003) 16(1):55–65. doi: 10.1097/00024720-200302000-00010

6. Zhang L, Hu XH, Wang QW, Cai YM, Zhao JX, Liu XY. Population distribution and clinical characteristics in rheumatoid arthritis patients with cervical spine instability. *Beijing Da Xue Xue Bao Yi Xue Ban* (2020) 52 (6):1034–9. doi: 10.19723/j.issn.1671-167X.2020.06.008

7. Magarelli N, Simone F, Amelia R, Leone A, Bosello S, D'Antona G, et al. MR imaging of atlantoaxial joint in early rheumatoid arthritis. *Radiol Med* (2010) 115 (7):1111–20. doi: 10.1007/s11547-010-0574-4

8. Ahn JK, Hwang JW, Oh JM, Lee J, Lee YS, Jeon CH, et al. Risk factors for development and progression of atlantoaxial subluxation in Korean patients with rheumatoid arthritis. *Rheumatol Int* (2011) 31(10):1363–8. doi: 10.1007/s00296-010-1437-y

9. Emdin CA, Khera AV, Kathiresan S. Mendelian randomization. *JAMA* (2017) 318(19):1925–6. doi: 10.1001/jama.2017.17219

10. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* (2014) 23(R1):R89–98. doi: 10.1093/hmg/ddu328

11. Ha E, Bae SC, Kim K. Large-Scale meta-analysis across East Asian and European populations updated genetic architecture and variant-driven biology of rheumatoid arthritis, identifying 11 novel susceptibility loci. *Ann Rheum Dis* (2021) 80(5):558–65. doi: 10.1136/annrheumdis-2020-219065

12. Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K , Donner K, et al. FinnGen: Unique genetic insights from combining isolated population and national health register data. (2022) 2022:. doi: 10.1101/2022.03.03.22271360

13. Ligthart S, Vaez A, Võsa U, Stathopoulou MG, de Vries PS, Prins BP, et al. Genome analyses of >200,000 individuals identify 58 loci for chronic inflammation and highlight pathways that link inflammation and complex disorders. *Am J Hum Genet* (2018) 103(5):691–706. doi: 10.1016/j.ajhg.2018.09.009

14. 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A , Brooks LD , Durbin RM, et al. A map of human genome variation from population-scale sequencing. *Nature* (2010) 467(7319):1061–73. doi: 10.1038/nature09534

15. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-base platform supports systematic causal inference across the human phenome. *eLife* (2018) 7:e34408. doi: 10.7554/eLife.34408

16. Li B, Martin EB. An approximation to the f distribution using the chi-square distribution. *Comput Stat Data Anal* (2002) 40(1):21–6. doi: 10.1016/s0167-9473(01)00097-4

17. Burgess S, Thompson SGCRP CHD Genetics Collaboration. Avoiding bias from weak instruments in mendelian randomization studies. *Int J Epidemiol* (2011) 40(3):755–64. doi: 10.1093/ije/dyr036

18. Broadbent JR, Foley CN, Grant AJ, Mason AM, Staley JR, Burgess S. MendelianRandomization v0.5.0: Updates to an r package for performing mendelian randomization analyses using summarized data. *Wellcome Open Res* (2020) 5:252. doi: 10.12688/wellcomeopenres.16374.2

19. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* (2013) 37(7):658–65. doi: 10.1002/gepi.21758

20. Burgess S, Thompson SG. Interpreting findings from mendelian randomization using the MR-egger method. *Eur J Epidemiol* (2017) 32(5):377–89. doi: 10.1007/s10654-017-0255-x

21. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol* (2016) 40(4):304–14. doi: 10.1002/gepi.21965

22. Zhang Y, Liu Z, Choudhury T, Cornelis MC, Liu W. Habitual coffee intake and risk for nonalcoholic fatty liver disease: A two-sample mendelian randomization study. *Eur J Nutr* (2021) 60(4):1761–7. doi: 10.1007/s00394-020-02369-z

23. Carter AR, Sanderson E, Hammerton G, Richmond RC, Davey Smith G, Heron J, et al. Mendelian randomisation for mediation analysis: Current methods and challenges for implementation. *Eur J Epidemiol* (2021) 36(5):465–78. doi: 10.1007/s10654-021-00757-1

24. Lynch M, Walsh B. Genetics and analysis of quantitative traits. *Sunderland MA: Sinauer* (1998) 1:535–57.

25. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PloS Genet* (2017) 13(11):e1007081. doi: 10.1371/journal.pgen.1007081

26. Tan JS, Liu NN, Guo TT, Hu S, Hua L. Genetically predicted obesity and risk of deep vein thrombosis. *Thromb Res* (2021) 207:16–24. doi: 10.1016/j.thromres.2021.08.026

27. Tan JS, Ren JM, Fan L, Wei Y, Hu S, Zhu SS, et al. Genetic predisposition of anti-cytomegalovirus immunoglobulin G levels and the risk of 9 cardiovascular diseases. *Front Cell Infect Microbiol* (2022) 12:884298. doi: 10.3389/fcimb.2022.884298

28. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from mendelian randomization between complex traits and diseases. *Nat Genet* (2018) 50(5):693–8. doi: 10.1038/s41588-018-0099-7

29. Ferrante A, Ciccia F, Giammalva GR, Iacopino DG, Visocchi M, Macaluso F, et al. The craniovertebral junction in rheumatoid arthritis: State of the art. *Acta Neurochir Suppl* (2019) 125:79–86. doi: 10.1007/978-3-319-62515-7_12

30. Shlobin NA, Dahdaleh NS. Cervical spine manifestations of rheumatoid arthritis: A review. *Neurosurg Rev.* (2021) 44(4):1957–65. doi: 10.1007/s10143-020-01412-1

31. Yurube T, Sumi M, Nishida K, Miyamoto H, Kohyama K, Matsubara T, et al. Incidence and aggravation of cervical spine instabilities in rheumatoid arthritis: A prospective minimum 5-year follow-up study of patients initially without cervical involvement. *Spine* (2012) 37(26):2136–44. doi: 10.1097/BRS.0b013e31826def1c

32. Kotecki M, Gasik R, Głuszko P, Sudoł-Szopińska I. Radiological evaluation of cervical spine involvement in rheumatoid arthritis: A cross-sectional retrospective study. *J Clin Med* (2021) 10(19):4587. doi: 10.3390/jcm10194587

33. Paus AC, Steen H, Røislien J, Mowinckel P, Teigland J. High mortality rate in rheumatoid arthritis with subluxation of the cervical spine: A cohort study of operated and nonoperated patients. *Spine* (2008) 33(21):2278–83. doi: 10.1097/BRS.0b013e31817f1a17

34. Kurowska-Stolarska M, Alivernini S. Synovial tissue macrophages in joint homeostasis, rheumatoid arthritis and disease remission. *Nat Rev Rheumatol* (2022) 18(7):384–97. doi: 10.1038/s41584-022-00790-8

35. Sorimachi Y, Iizuka H, Ara T, Nishinome M, Iizuka Y, Nakajima T, et al. Atlanto-axial joint of atlanto-axial subluxation patients due to rheumatoid arthritis before and after surgery: Morphological evaluation using CT reconstruction. *Eur Spine J* (2011) 20(5):798–803. doi: 10.1007/s00586-010-1611-7

36. Nazarinia M, Jalli R, Kamali Sarvestani E, Farahangiz S, Ataollahi M. Asymptomatic atlantoaxial subluxation in rheumatoid arthritis. *Acta Med Iran.* (2014), 462–6.

37. Liao S, Jung MK, Hörnig L, Grützner PA, Kreinest M. Injuries of the upper cervical spine-how can instability be identified? *Int Orthop* (2020) 44(7):1239–53. doi: 10.1007/s00264-020-04593-y

38. Lv JM, Chen JY, Liu ZP, Yao ZY, Wu YX, Tong CS, et al. Celluar folding determinants and conformational plasticity of native c-reactive protein. *Front Immunol* (2020) 11:583. doi: 10.3389/fimmu.2020.00583

39. Fang Z, Lv J, Wang J, Qin Q, He J, Wang M, et al. C-reactive protein promotes the activation of fibroblast-like synoviocytes from patients with rheumatoid arthritis. *Front Immunol* (2020) 11:958. doi: 10.3389/fimmu.2020.00958

40. Nasonov EL, Chichasova NV, Baranov AA, Imametdinova GR, Tishchenko VA, Kuznetsova TB, et al. The clinical role of c-reactive protein in rheumatoid arthritis. *Klin Med (Mosk)* (1997) 75(7):26–30.

41. Kim KW, Kim BM, Moon HW, Lee SH, Kim HR. Role of c-reactive protein in osteoclastogenesis in rheumatoid arthritis. *Arthritis Res Ther* (2015) 17(1):41. doi: 10.1186/s13075-015-0563-z

42. Möller B, Everts-Graber J, Florentinus S, Li Y, Kupper H, Finckh A. Low hemoglobin and radiographic damage progression in early rheumatoid arthritis: Secondary analysis from a phase III trial. *Arthritis Care Res (Hoboken)* (2018) 70 (6):861–8. doi: 10.1002/acr.23427

43. Tan JS, Yan XX, Wu Y, Gao X, Xu XQ, Jiang X, et al. Rare variants in MTHFR predispose to occurrence and recurrence of pulmonary embolism. *Int J Cardiol* (2021) 2021:331:236–242. doi: 10.1016/j.ijcard.2021.01.073

frontiers | Frontiers in Endocrinology

TYPE Original Research
PUBLISHED 22 December 2022
DOI 10.3389/fendo.2022.1037289

OPEN ACCESS

EDITED BY
Tony Yuen,
Icahn School of Medicine at Mount Sinai, United States

REVIEWED BY
Elisa Wirthgen,
University Hospital Rostock, Germany
Rongxiu Zheng,
Tianjin Medical University Eye Hospital, China

*CORRESPONDENCE
Yan Ni
yanni617@zju.edu.cn
Junfen Fu
fjf68@zju.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Pediatric Endocrinology, a section of the journal Frontiers in Endocrinology

RECEIVED 05 September 2022
ACCEPTED 02 December 2022
PUBLISHED 22 December 2022

CITATION
Zhang J, Wu W, Huang K, Dong G, Chen X, Xu C, Ni Y and Fu J (2022) Untargeted metabolomics reveals gender- and age- independent metabolic changes of type 1 diabetes in Chinese children.
Front. Endocrinol. 13:1037289.
doi: 10.3389/fendo.2022.1037289

Untargeted metabolomics reveals gender- and age-independent metabolic changes of type 1 diabetes in Chinese children

Jianwei Zhang[1,2†], Wei Wu[1†], Ke Huang[1], Guanping Dong[1], Xuefeng Chen[1], Cuifang Xu[1], Yan Ni[1*] and Junfen Fu[1*]

[1]Department of Endocrinology, The Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Hangzhou, China, [2]Department of Paediatrics, Shaoxing Women and Children Hospital, Shaoxing, China

Introduction: Type 1 diabetes (T1D) is a chronic condition associated with multiple complications that substantially affect both the quality of life and the life-span of children. Untargeted Metabolomics has provided new insights into disease pathogenesis and risk assessment.

Methods: In this study, we characterized the serum metabolic profiles of 76 children with T1D and 65 gender- and age- matched healthy controls using gas chromatography coupled with timeof-flight mass spectrometry. In parallel, we comprehensively evaluated the clinical phenome of T1D patients, including routine blood and urine tests, and concentrations of cytokines, hormones, proteins, and trace elements.

Results: A total of 70 differential metabolites covering 11 metabolic pathways associated with T1D were identified, which were mainly carbohydrates, indoles, unsaturated fatty acids, amino acids, and organic acids. Subgroup analysis revealed that the metabolic changes were consistent among pediatric patients at different ages or gender but were closely associated with the duration of the disease.

Discussion: Carbohydrate metabolism, unsaturated fatty acid biosynthesis, and gut microbial metabolism were identified as distinct metabolic features of pediatric T1D. These metabolic changes were also associated with T1D, which may provide important insights into the pathogenesis of the complications associated with diabetes.

KEYWORDS

type 1 diabetes, metabolomics, children, serum, duration

Introduction

Type 1 diabetes mellitus (T1D), a chronic disease caused by destruction of pancreatic cells and decreased insulin secretion, is accompanied by various complications, and has serious effects on the quality of life and life span (1). According to the newly released Diabetes Atlas from the International Diabetes Federation in 2021, a total of 1,211,900 children and adolescents world-wide had T1D (2) and China is ranked 4th globally for the number of children with T1D (www.diabetesatlas.org).

Recent increases in the incidence of T1D in children and adolescents highlight the importance of environmental factors in disease development. Metabolomic analysis serves as an excellent approach to explore the integrated response of an organism toward environmental changes. Previous metabolomic studies have demonstrated the crucial role of metabolic profiling in discovering biomarker discovery that are predictive of disease incidence and development and potentially its pathogenesis (3–8). For example, Balderas et al. compared the urine and serum metabolome of 34 children with diabetes and 15 non-diabetic controls and discovered that children with T1D had altered bile acid profiles (9). Bile acids are absorbed during enterohepatic circulation, and thus alterations in bile acid profiles may reflect the T1D-associated changes in the gut microbiome. Changes in lipids that play a role in cellular signaling and metabolism in the body during progression to T1D were also noted by several lipidomic studies (5, 9–16). Suvitaival et al. (17) concluded that levels of triacylglycerols, phosphatidylcholines, sphingomyelins, and ceramides were reduced in the plasma of TID children before diagnosis. However, most of the previous metabolomics studies have focused on specific varieties of candidate metabolites, thus failing to effectively provide a complete understanding of the metabolic pathogenesis in TID. Furthermore, these studies have been conducted predominantly in Western populations (18). Data from Asian populations are sparse and limited mainly to cross-sectional studies (19, 20). It is well known that T1D in children and adolescents is affected by age and gender. However, current studies have not stratified the metabolomics spectrum of the disease from the perspective of age and gender.

In the present study, we applied an untargeted metabolomics approach to measure the metabolic profiles of pediatric patients with T1D in the Chinese population, as compared to their healthy controls. The untargeted metabolomics approach was performed using gas chromatography coupled with time-of-flight mass spectrometry (GC-TOFMS). In parallel, we comprehensively evaluated the changes of clinical biomarkers of patients with T1D, including metabolic biomarkers from blood biochemistry, inflammatory cytokines, antibodies, immunoglobulins, and trace elements. This study aims to investigate the metabolic phenome of pediatric T1D and its association with the duration of the disease, and validate the consistency of metabolic changes among male and female patients or at different ages.

Materials and methods

Study subjects

The study was approved by the Institutional Review Board of The Children's Hospital of Zhejiang University, School of Medicine (Approval Number: 2016-JRB-018). Written informed consent was obtained from the guardians of all recruited children, and the study was performed in accordance with the principles of the Declaration of Helsinki. A total of 141 participants including 76 T1D patients and 65 healthy controls were enrolled in this study (Table 1). Patients were diagnosed with T1D during their stay at the Children's Hospital of Zhejiang University School of Medicine and were enrolled in the study between 2016 and 2020. The median duration of the disease since diagnosis was 12 months (range: 1–72 months) and based on this, T1D patients were divided into short-term group (<3 months), mid-term group (3–12 months), and long-term group (>12 months), respectively. T1D was diagnosed based on clinical and biochemical features, specifically elevated blood glucose at presentation (a random measurement of > 11.1 mmol/l and/or fasting blood glucose level of > 7.1 mmol/l), and classical symptoms of diabetes. Furthermore, all patients met at least one of the following criteria: 1) diabetic ketoacidosis (DKA); 2) presence of T1D-associated antibodies (glutamic acid decarboxylase, islet antigen 2, islet cell, or insulin autoantibodies); and/or 3) on-going requirement for insulin therapy. Healthy control refers to the group of children who visited the hospital for routine physical examination, had no disease state, and were enrolled on a voluntary basis.

Clinical measurements

The medical records and routine laboratory biochemistry data of the participants were summarized in Table 2. Serum lipid profiles (i.e., total cholesterol, HDL-cholesterol, LDL-cholesterol, and triacylglycerols), total protein, apolipoprotein A1, apolipoprotein B, lipoprotein A, carboxyhemoglobin, high-sensitivity C-reactive protein, albumin, globulin, glycated hemoglobin (HbA1C), insulin-like growth factor 1 (IGF-1), insulin-like growth factor-binding protein 3, alanine aminotransferase (ALT), alkaline phosphatase (ALP), aspartate aminotransferase (AST), lactate dehydrogenase (LDH), total bilirubin, direct bilirubin, and indirect bilirubin were measured. Human inflammatory cytokine multiple ELISA kit was used to quantitatively measure cytokine levels including interferon-gamma (IFN-γ), interleukin (IL)-10, IL-2, IL-4, IL-6, and tumor necrosis factor-alpha (TNF-α). Multiple monoclonal antibodies that recognize a common cell-surface antigen are combined to

10.3389/fendo.2022.1037289

TABLE 1 Baseline characteristics of the study participants.

Name	T1D (n=76)	Control (n=65)	*p*
Age (in months)	109.2 ± 47.07	110.9 ± 41.33	0.1325
Gender			
Male	29 (44.6%)	31 (47.7%)	
Female	47 (55.4%)	34 (52.3%)	
Height, cm	132.1 ± 24.93	131.3 ± 20.75	0.09
Weight, kg	30.0 ± 13.76	29.4 ± 11.58	0.20
BMI, kg/m^2	16.4 ± 2.69	16.3 ± 1.90	0.14
Duration (in months)	16.7 ± 18.9	—	—

Normally distributed variables were analyzed using student t-test and presented as mean ± standard deviation (SD), while non-normally distributed variables were performed by non-parametric Mann-Whitney U test and presented as medians and interquartile range (25th–75th percentiles). T1D, type 1 diabetes; BMI, body mass index.

form clusters of differentiation. The clusters are numbered sequentially with respect to when they were discovered and defined. The cell-surface reactivity of monoclonal antibodies to each CD antigen was detected by flow cytometry.

Metabolomics

Blood samples were collected after fasting overnight for at least 8 hours and centrifuged to obtain serum prior to storage in -80°C freezer. The untargeted metabolomics profiling of serum samples was performed on a GC-TOFMS system (Pegasus BT, Leco Corp., St. Joseph, MO, USA) equipped with an Agilent 7890B gas chromatograph and a Gerstel multipurpose sampler with dual heads (Gerstel, Muehlheim, Germany). The procedure was performed as described in a previously published paper with minor modifications (21). Briefly, each aliquot of 50 µL serum sample was mixed with 10 µL of internal standard, to which 175 µL of pre-chilled methanol/chloroform (v/v=3/1) was added for metabolite extraction. After centrifugation at 13,500 rpm for 20 min at 4°C (Microfuge 20R, Beckman Coulter, Inc., Indianapolis, IN, USA), the supernatant was carefully transferred to an autosampler vial. The samples in autosampler vials were then evaporated briefly to remove chloroform using a CentriVap vacuum concentrator (Labconco, Kansas City, MO, USA), and further lyophilized with a FreeZone freeze dryer equipped with a stopping tray dryer (Labconco, Kansas City, MO, USA). The sample derivatization was performed by a robotic multipurpose sampler with dual heads (Gerstel, Muehlheim, Germany). Specifically, the dried sample was derivatized with 50 µL of methoxyamine (20 mg/mL in pyridine) at 37.5°C for 2 hr., followed by incubation at 37.5°C for 1 hr after the addition of 50 µL of MSTFA (1% TMCS) containing FAMEs as retention indices. Separation and sample derivatization were performed as parallel operations. A Rxi-5 ms capillary column (30 m × 250 µm i.d., 0.25 µm film thickness; Restek corporation, Bellefonte, PA, USA) was used for metabolite separation. The temperature was initially held at 80°C for 2 min, then ramped up to 300°C at the rate of 12°C/min, held for 4.5 min, then further ramped up to 320°C at the rate of 40°C/min, and finally, held for 1 min. Helium was used as the carrier gas at a constant flow rate of 1.0 mL/min. The temperature of the injector and the transfer interface were both set to 270°C and the injection volume was 0.5 µL in spitless mode. Measurements were made using electron impact ionization (70 eV) in the full scan mode (m/z 50–500). Instrument optimization was performed as needed.

Metabolite annotation

Metabolite annotation was performed by comparing the retention indices and mass spectral data with those previously generated from reference standards in the in-house library (22). The reference chemicals were commercially purchased from Sigma-Aldrich (St. Louis, MO, USA), Santa Cruz (Dallas, TX, USA), and Nu-Chek Prep (Elysian, MN, USA). Commercial libraries such as NIST library 2017 and LECO/Fiehn Metabolomics library for GC-TOFMS were used for cross-validation analysis. The direct relationship of two adjacent metabolites from the known metabolic relation network (KEGG) was used to indicate the alteration of specific metabolic enzyme, thereby providing complementary biological information for metabolite interactions. Metabolites were annotated in the serum samples with those of pure chemical standards. Metabolites that did not pass our QC criteria (CV>20%) were removed from further statistical analysis, as the purpose of this project was to provide data for a further validation study, rather than making a simple biomarker discovery. Missing values were initially imputed using QRILC method reported in our previous work (23).

Data analysis

The medical records and the routine laboratory biochemistry data were statistically analyzed using R packages ver. 4.0.2. The details of statistical methods applied in this study, R functions and packages were summarized in Supplementary Table 1. Specifically, normally distributed variables were analyzed using student *t*-test and presented as mean ± standard deviation (SD), while non-normally distributed variables were performed by non-parametric Mann-Whitney U test and presented as medians and interquartile range (25th–75th percentiles). The raw metabolomic data generated by GC-TOFMS were processed using ADAP software (24). To reduce bias caused by the high blood glucose levels in diabetic patients,

10.3389/fendo.2022.1037289

TABLE 2 Characteristics of clinical measurements.

Name	HC Median (reference range)	T1D (Median ± SD)	Median	FC
TSH (uIU/ml)	2.645 (0.35–4.94)	2 ± 1.12	2	0.636
Potassium (mmol/L)	4.5 (3.5–5.5)	4 ± 0.35	4	0.842
Sodium (mmol/L)	140 (135–145)	138 ± 4.94	138	0.987
Glucose (mmol/L)	4.85 (3.6–6.1)	12 ± 5.79	12	2.515
Total Bile Acids (μmol/L)	6.0 (0.0–12.0)	5 ± 22.19	5	0.783
Uric acid (μmol/L)	256 (155–357)	220 ± 90.1	220	0.859
25(OH)D (ng/ml)	88.75 (27.5–150.0)	48 ± 19.31	48	0.544
Hemoglobin (g/L)	140 (120–160)	133 ± 11.68	133	0.951
PLT (10^9/L)	250 (100–400)	291 ± 82.44	291	1.163
WBC (10^9/L)	8.0 (4.0–12.0)	7 ± 4.07	7	0.865
Insulin (μg/L)	12.45 (1.9–23)	4 ± 23.88	4	0.321
HOMA-IR (%)		2 ± 14.38	2	
ALT (U/L)	<50	13 ± 10.05	13	
Urea (μmol/L)	4.11 (1.79–6.43)	5 ± 1.7	5	1.128
Cholesterol (mmol/L)	4.35 (3.00–5.70)	4 ± 1.83	4	0.961
CKMB	<25	23 ± 13.36	23	
Creatinine (μmol/L)	46 (15–77)	58 ± 14.44	58	1.261
GGT (U/L)	32.5 (8–57)	12 ± 3.4	12	0.369
TC (mmol/L)	<1.70	1 ± 7.08	1	
HDLC (mmol/L)	>1.04	1 ± 0.4	1	
LDLC (mmol/L)	<3.37	2 ± 1.03	2	
HbA1c (%)	5.4 (4.5–6.3)	8 ± 2.86	8	1.463
HsCRP (mg/L)	4 (0–8)	4 ± 6.36	4	1
β2-MG (mg/L)	0.15 (0.00–0.30)	0 ± 2.45	0	0.715
C3 (g/L)	1 (0.50–1.50)	1 ± 0.26	1	1
C4 (g/L)	0.25 (0.10–0.40)	0 ± 0.1	0	0.8
IgG (g/L)	10.2 (6.36–14.04)	11 ± 2.42	11	1.058
IgM (g/L)	0.75 (0.29–1.21)	1 ± 0.48	1	1.44
Urinary Creatinine (μmol/L)	11275 (2550–20000)	4781 ± 3491.06	4781	0.424
Uridine triphosphate (mg)	<100.0	12 ± 270.56	12	
Urine α1-microglobulin (mg/L)	<12.00	7 ± 13.63	7	

FC is the fold change ratios by calculating the median value of each clinical marker in the T1D group vs. the reference range. HC, healthy control; T1D, type 1 diabetes; FC, fold change; TSH, thyroid stimulating hormone; PLT, total platelet count; WBC, white blood cell count; HOMA-IR, homeostatic model assessment for assessing insulin resistance; ALT, alanine transaminase; CKMB, creatinine kinase myocardial band; GGT, gamma-glutamyl transferase; TC, total count; HDLC, high density lipoprotein cholesterol; LDLC, low density lipoprotein cholesterol; Hb1Ac, hemoglobin A1c; HsCRP, high-sensitivity C-reactive protein; β2-MG, Beta-2 microglobulin; C3, complement component 3; C4, complement component 4; IgG, immunoglobulin G; IgM, immunoglobulin M.

glucose was excluded from the final data set. The metabolome data were further standardized before statistical modeling. Unsupervised principal component analysis (PCA) was used to evaluate the natural clustering between patients with T1D and their healthy controls. Each dot represented an individual subject and color-coded based on their grouping. To overcome the complexity of biological samples, a widely-used supervised orthogonal partial least square discriminant analysis (OPLS-DA)

model was applied to capture the differential metabolites between the two groups. The OPLS-DA model was constructed using 1/7-fold cross-validation. Metabolic pathway enrichment analysis was done using the hypergeometric algorithm deployed in MetaboAnalyst (25). The significance of the metabolic pathways associated with T1D was determined by the cutoff p-value of 0.10. All the *p* values were adjusted by Benjamini & Hochberg method. Spearman correlation analysis was used to evaluate the correlation between each differential metabolite and disease duration.

Results

Clinical characteristics and metabolic profiles of T1D patients

The basic demographic information of the participants is summarized in Table 1. No significant differences were observed in gender, age, and BMI between patients with T1D and healthy controls. Serum lipid profiles are summarized in Table 2. The mass spectrometry-based metabolomics study detected 282 circulating metabolites that were present across all the study samples, with a low median process variability among QC samples (<20%). A total of 51 metabolite-metabolite ratios were also calculated according to their metabolic reactions. The identified serum metabolome covered over 60 biochemical pathways of human metabolism and included a wide range of metabolite classes such as amino acids, organic acids, fatty acids, alcohols and sugar derivatives, lipids, nucleotides, indoles, and phenols. The OPLS-DA scores plot depicts the distinct metabolic profiles associated with T1D patients versus healthy controls (Figure 1A). A total of 70 differential metabolites and 14 metabolite ratios were obtained between T1D patients and healthy controls with FDR-corrected *p* value ($p < 0.01$). Carbohydrates and organic acids altered apparently, and the most significantly altered metabolites were 1,5-anhydrosorbitol, α-lactose, indole acetic acid, arachidic acid, and so forth. The majority of these metabolites were significant after age- and/or gender adjusted (Supplemental Table 2). Based on these differential metabolites, the metabolic enrichment analysis indicated that 11 pathways were significantly perturbed in T1D patients as compared to controls (Figure 1B), including glucose metabolism, glutathione metabolism, arginine and proline metabolism, branched chain amino acid (BCAA) metabolism, etc. Among these, galactose metabolism was the most significantly altered metabolic pathway associated with T1D, and specifically increased α-lactose, sorbitol, myoinositol, sucrose, glycerol, and reduced d-mannose and d-galactose were identified in this study. Additionally, we examined the clinical characteristics of T1D patients, including blood biochemistry, complete blood count (CBC), cytokines, hormones, proteins, trace elements, and urine tests. An integrative view of both clinical characteristics and metabolic changes of T1D patients is illustrated in Figure 1C. Clinical markers that were different included elevated levels of urine transferrin, IgE, IL-4, tartrate resistant acid phosphatase (TRAcP-5b), adenosine deaminase (ADA), and glucose levels, as well as reduced levels of urine creatinine, IGF-1, adrenocorticotropic hormone (ACTH), IFN-r, eosinophils, insulin, and r-glutamyl-transferase (GGT).

Association of age and gender factors with circulating metabolome

Although age- matched controls were used in this study, age was believed to be a confounding factor of host metabolism that deserves thorough investigation, particularly for children. The OPLS regression model revealed that metabolic profile variations correlated closely with the age of T1D patients (Figure 2A, $r = 0.98$, $p = 1.96e\text{-}50$). The patients were then further stratified into three different subgroups according to their age: young (1–84 months), middle (85–120 months), and old (121–185 months). The baseline characteristics of patients in these three subgroups are shown in Table 1. The metabolic profile of patients in each subgroup was compared with age-matched healthy controls, and each comparison consistently showed an apparent separation between the two groups according to the OPLS-DA score plot (Figures 2B–D). The heatmap of z-score values derived from each comparison (FDR-corrected *p* value < 0.01) showed the relative expressions of differential metabolites among patients and healthy controls, and indicated whether specific metabolic changes were consistent across different age groups (Figure 2E). For example, 1,5-anhydrosorbitol and indoleacetic acid were significantly reduced in T1D patients of all different ages (Figures 2F, G).

Similarly, we also examined the impact of gender on metabolism. The OPLS-DA model with information on the gender of the participants did not show any obvious internal variations among patients with T1D or healthy controls (Figure 3A). Moreover, to eliminate the influence of gender, OPLS-DA model was applied to compare the metabolic variations between male and female T1D patients (29 and 49, respectively) and healthy controls (31 and 34, respectively), separately (Figures 3B, C). The differential metabolites, 1,5-anhydrosorbitol and indoleacetic acid, were significantly reduced in both male and female patients with T1D (Figures 3D, E). The enrichment pathway analysis based on male and female differential metabolites further validated that most of differential metabolic pathways were consistent in both male and female patients. However, we found that four pathways namely, (i) phenylalanine, tyrosine and tryptophan biosynthesis, (ii) pentose phosphate pathway, (iii) purine metabolism, and (iv) glyoxylate and dicarboxylate metabolism, could be affected by the gender (Figure 3F).

10.3389/fendo.2022.1037289

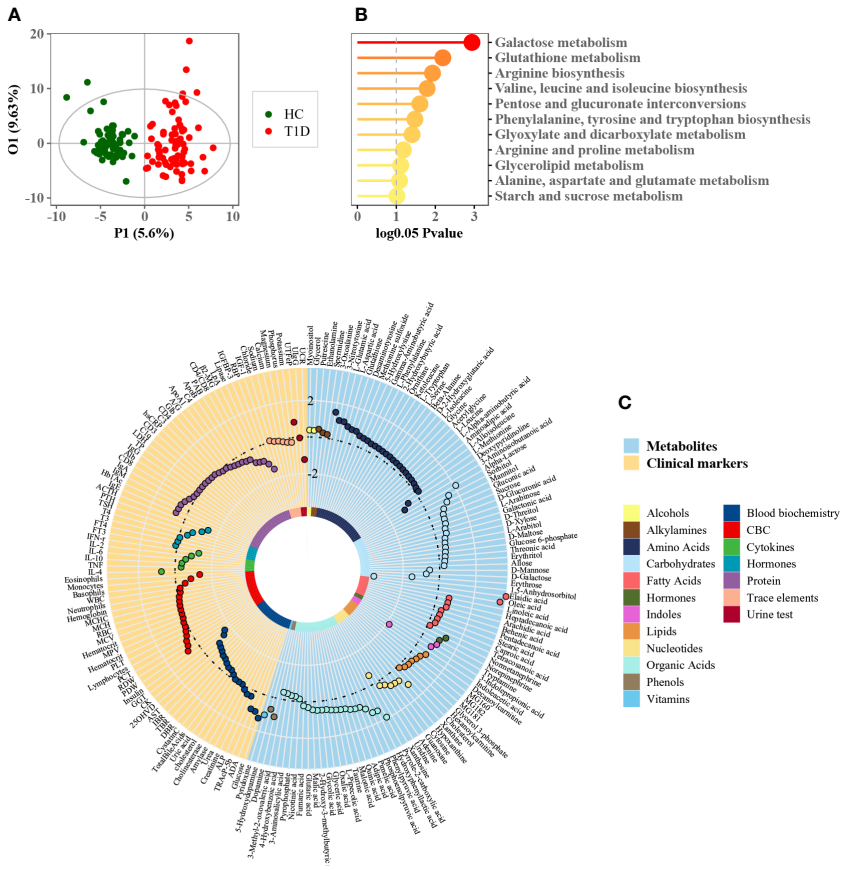


FIGURE 1
Phenome and metabolome analysis of T1D patients. **(A)** OPLS-DA score plot of patients with T1D and healthy controls (each green dot represents a healthy subject while each red dot denotes a T1D patient). **(B)** Metabolic enrichment pathway analysis using MetaboAnalyst. **(C)** Circle plot of phenome and metabolome fold changes (T1D versus healthy controls). Each dot was color-coded based on their chemical classes or clinical diagnostic purpose.

Association of disease duration with circulating metabolome

To determine whether there is an association of disease duration and serum metabolome, the orthogonal partial least squares regression (PLSR) analysis was performed (Figure 4A). This analysis depicted an obvious linearity between phenome and disease duration of T1D patients. The patients were further divided into three major clusters: initial progression from disease onset (short, 1–6 months), moderate phase (medium, 7–18 months), and advanced phase (long, 1.5-6 years). As shown in Figure 4B, the metabolic profiles of the three subgroups were clearly separated according to the score plot of PLS-DA model. The heat map indicated a significant Spearman's correlation between metabolites belonging to nine chemical classes and disease duration (Figure 4C). Of them, 1,5-anhydrosorbitol, pyruvic acid, and adenine had the strongest positive correlations with disease duration.

Discussion

The pathogenesis of diabetes in children is complicated due to the frequent occurrence of future complications. Insulinopenia and hyperglycemia, characteristics of the T1D milieu, profoundly alter metabolic homeostasis. Altered metabolites affected by the variable insulin and glycemic levels may theoretically increase the risk of long-term complications. While diabetes is primarily characterized by hyperglycemia, other nutrient metabolic pathways like amino acid and tricarboxylic acid cycle (TCA) are also profoundly perturbed. However, a comprehensive metabolic signature for T1D, especially in Chinese children, was not previously established. Hence, we performed a serum metabolomics study of Chinese children with T1D. Utilizing both multivariate and univariate statistical analyses, a unique metabolic pattern was observed to be related to T1D. This included 70 differently expressed metabolites that were associated with 11 specific altered metabolic pathways. These metabolic changes were also

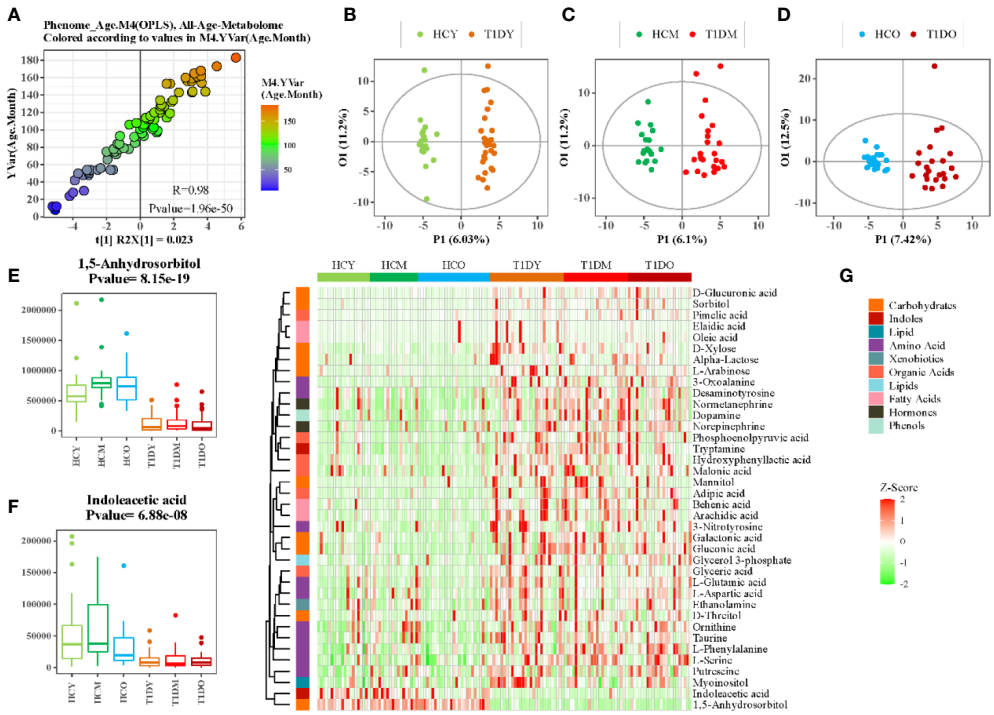


FIGURE 2
Association analysis between age and serum metabolome. **(A)** Scatter plot of OPLS-DA PC1 scores and ages of patients. **(B–D)** OPLS-DA score plot of patients with T1D and healthy controls at different ages, B:(1-84 months), C:(85-120months), and D:(121-185months). **(E, F)** Box plots showing the concentration of selected metabolites at different ages. **(G)** Heatmap of z-score values for differential metabolites between patients with T1D and healthy controls at different ages.

investigated by sub-dividing the patients based on different ages and gender. In addition, an integrative analysis of clinical features and metabolic profiles was performed, which provided us a comprehensive view of pediatric T1D. An interesting finding was that a group of differential metabolites were closely associated with the time elapsed since diagnosis of the disease, which might provide important insights into the pathogenesis of the complications associated with diabetes.

Dysregulated carbohydrate metabolism was an obvious metabolic feature of T1D observed in this study. Particularly, galactose metabolism was the most significantly altered metabolic pathway associated with T1D (Figure 1B). This alteration remained significant across different ages and in both genders of T1D patients compared to healthy controls, but it was similar between young and old participants (Figure 2G), or between male and female counterparts (Figure 3F). We also observed that glutathione, arginine, and BCAA metabolism were significantly altered between the T1D and control groups. In normal physiological conditions, glutathione has antioxidative and free radical-scavenging roles, thereby maintaining the metabolism and homeostasis of cells (26, 27). Metabolic disorder of the glutathione pathway increases oxidative stress, which may damage kidneys and blood vessels, and cause neurodegeneration. Arginine is the precursor for oxide synthesis, which can be converted to vasodilating factors in the body. On one hand, arginine can stimulate insulin secretion, but on the other hand, arginine produces NO, which is involved in response to oxidative stress in organisms, and participates in glutathione metabolism. Disorders of arginine metabolism may affect endothelial cell function, leading to insulin resistance and disturbances in metabolism and hemodynamics (28). BCAAs include leucine, isoleucine, and valine, which are essential amino acids that provide energy to the body (29). Disordered BCAA metabolites can block insulin signaling and disturb lipid metabolism, resulting in insulin resistance and excessive lipid accumulation, respectively (30). BCAA metabolism disorder is a biomarker of cardiovascular metabolic diseases (31).

Among the differential carbohydrates, 1,5-anhydroglucitol (1,5-AG) was the most significant marker that was consistently lower in T1D patients, regardless of gender or age. Serum 1,5-AG has been considered a potential marker of short-term glycemic control and can be used for T1D diagnosis or the screening of high-risk patients. Moreover, it was found to be less influenced by diet or physical activity as compared to point glycemic markers (32). Thus, 1,5-AG can be an effective supplementary marker to hemoglobin A1c. Low plasma levels of 1,5-AG are associated with

10.3389/fendo.2022.1037289

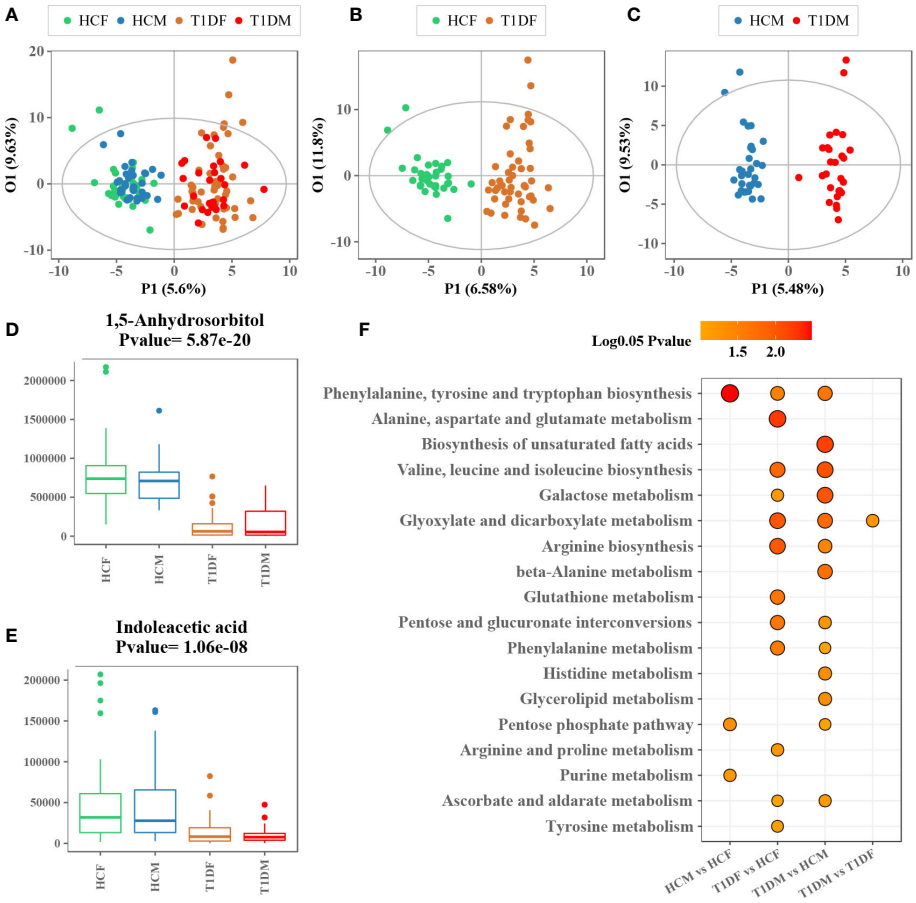


FIGURE 3
Association analysis between gender and serum metabolome. **(A)** OPLS-DA score plot of patients with T1D and healthy controls labeled with information on gender of the participants. **(B, C)** OPLS-DA score plot of patients with T1D and healthy controls for males and females. **(D, E)** Box plots showing the concentration of representative metabolites across different groups. **(F)** Metabolic enrichment pathway analysis for different group comparisons.

decreased pancreatic β-cell function (33). Sorbitol takes part in the polyol pathway through the reduction of intracellular glucose to sorbitol. The polyol pathway gets activated when excess glucose is present within the cells and thus, a hyperglycemic state might accelerate intracellular accumulation of sorbitol. Furthermore, excessive sorbitol in the cells has been associated with a pro-oxidative environment, which is known to increase diabetes-related complications (34, 35). Renal tubular reabsorption of 1,5-AG is inhibited when there is excess glucose in the plasma. Studies have found that 1,5-AG levels decrease in patients with diabetes and this decrease is related to kidney damage owing to high blood glucose levels (36).

We found that compared with the healthy group, TCA cycle metabolites (pyruvate, fumarate, malate, and linoleic acid) were significantly increased in the T1D group. Higher concentrations of pyruvate appear to be necessary for anaplerosis (37). The increased linoleic acid might be related to the insulin resistance in patients with T1D. The increase of fatty acids can lead to the slowing down of the tricarboxylic acid cycle e.g., the accumulation of fumarate and malate, which can block the oxidation of glucose. Some studies have considered that high physiological levels of exogenous insulin and hyperglycemia could be responsible for insulin resistance in T1D patients (38, 39).

In the present study, indoleacetic acid was significantly reduced in the diabetic group, suggesting that T1D was associated with the metabolism of indole and its derivatives. Dietary tryptophan can be metabolized into IAA by gut microbiota through the indole-3-acetamide pathway under the catalysis of tryptophan monooxygenase and indole-3-acetamide hydrolase (40). Recent work has demonstrated that gut microbiota is an essential modulator of T1D susceptibility and the reduced IAA levels were associated with intestinal mucosal barrier integrity impairment (41). Dysfunction of the intestinal mucosal barrier can increase intestinal permeability and trigger an immunological response, contributing to the development of

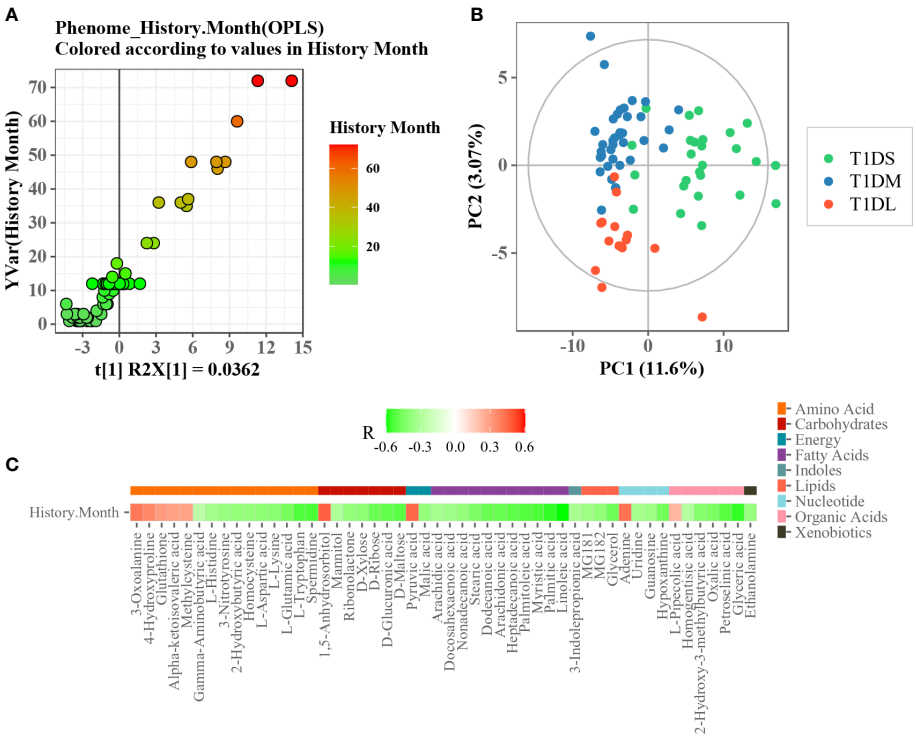


FIGURE 4
Association analysis between disease duration and serum metabolome. **(A)** Scatter plot of OPLS PC1 scores and disease duration. **(B)** PLS-DA score plot of T1D patients with different disease durations. **(C)** Heatmap of Spearman's correlation between differential metabolites and disease durations. The metabolite was assigned to each metabolite class (plotted with different color).

various autoimmune diseases, Including T1D (42). Indole acetic acid is also involved in the metabolism of purines, and this metabolic cycle is closely related to the pathogenesis of diabetic nephropathy.

We also found that adenine, 1,5-AG, and pyruvate correlated strongly and positively with the course of the disease. Clinically, accumulation of adenine in the blood results in insoluble crystal (2, 8-dihydroxyadenine) precipitates in the renal tubules and obstruction of tubular flow, which initiates renal injury (43, 44). A similar disease phenotype can be induced in rodents by adenine-feeding (45, 46). Adenine feeding-induced chronic kidney disease in rodents is characterized by elevated plasma concentrations of urea and creatinine, proteinuria, interstitial fibrosis, extensive tubular dilation, degeneration of the proximal tubular epithelium with loss of the brush border and inflammatory cell infiltration (47, 48).

In summary, metabolomics is a powerful tool for investigation of the nature-nurture relationships involved in the development of pediatric diabetes. Overall, in this study, carbohydrate metabolism, unsaturated fatty acid biosynthesis, and gut microbial metabolism were identified as distinct metabolic features of pediatric T1D. These metabolic changes were also associated with T1D, which may provide important insights into the pathogenesis of the complications associated with diabetes. The limitation of this study is the lack of fecal samples for the analysis of gut microbiome to confirm the role of microbial metabolites in T1D, e.g., IAA metabolism. Further studies are needed to explore the complex relationship between gut microbiome and metabolism in the pathogenesis of T1D. In the future, larger studies are needed to determine whether these metabolic markers can add to the prediction of long-term T1D.

Data availability statement

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

Ethics statement

The study was approved by the Institutional Review Board of The Children's Hospital of Zhejiang University, School of Medicine (Approval Number: 2016-JRB-018). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

JF and JZ conceived and designed the study. WW, KH collected the plasma samples. GD and XC provided patient care and were responsible for communication with the parents. YN conducted the metabolomics analysis. YN and CX performed the data analysis. JZ and YN wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study is funded by the National Key Research and Development Program of CHINA (2021YFC2701901, 2016YFC1305301), National Natural Science Fund (81570759), Zhejiang Province Key Disciplines of Medicine (Innovation discipline, 11-CX24), Zhejiang Provincial Key Science and Technology Project (LGF21H070004) and Health Commission of Zhejiang Province (2021KY1155).

Acknowledgments

The authors would like to thank all pediatric patients and their parents. The authors thank Dr. Qing Ye and the nurse co-workers for sample collection. The authors also thank Dr. Irwin Kurland and Dr. Yunping Qiu at Albert Einstein College of Medicine, and Dr. Wei Jia and Dr. Cynthia Rajani at the University of Hawaii Cancer Center for providing invaluable insights on the initial draft of the manuscript.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fendo.2022.1037289/full#supplementary-material

References

1. DiMeglio LA, Evans-Molina C, Oram RA. Type 1 diabetes. *Lancet* (2018) 391:2449–62. doi: 10.1016/S0140-6736(18)31320-5

2. Home. *IDF diabetes atlas 2021 | IDF diabetes atlas* . Available at: https://diabetesatlas.org/atlas/tenth-edition/.

3. Li Q, Parikh H, Butterworth MD, Lernmark Å, Hagopian W, Rewers M, et al. Longitudinal metabolome-wide signals prior to the appearance of a first islet autoantibody in children participating in the TEDDY study. *Diabetes* (2020) 69:465–76. doi: 10.2337/db19-0756

4. Lamichhane S, Kemppainen E, Trost K, Siljander H, Hyöty H, Ilonen J, et al. Circulating metabolites in progression to islet autoimmunity and type 1 diabetes. *Diabetologia* (2019) 62:2287–97. doi: 10.1007/s00125-019-04980-0

5. Sen P, Dickens AM, López-Bascón MA, Lindeman T, Kemppainen E, Lamichhane S, et al. Metabolic alterations in immune cells associate with progression to type 1 diabetes. *Diabetologia* (2020) 63:1017–31. doi: 10.1007/s00125-020-05107-6

6. Sysi-Aho M, Ermolov A, Gopalacharyulu PV, Tripathi A, Seppänen-Laakso T, Maukonen J, et al. Metabolic regulation in progression to autoimmune diabetes. *PloS Comput Biol* (2011) 7:e1002257. doi: 10.1371/journal.pcbi.1002257

7. Oresic M. Metabolomics in the studies of islet autoimmunity and type 1 diabetes. *Rev Diabetes Stud* (2012) 9:236–47. doi: 10.1900/RDS.2012.9.236

8. Knebel B, Strassburger K, Szendroedi J, Kotzka J, Scheer M, Nowotny B, et al. Specific metabolic profiles and their relationship to insulin resistance in recent-onset type 1 and type 2 diabetes. *J Clin Endocrinol Metab* (2016) 101:2130–40. doi: 10.1210/jc.2015-4133

9. Balderas C, Rupérez FJ, Ibañez E, Señorans J, Guerrero-Fernández J, Casado IG, et al. Plasma and urine metabolic fingerprinting of type 1 diabetic children. *Electrophoresis* (2013) 34:2882–90. doi: 10.1002/elps.201300062

10. Pflueger M, Seppänen-Laakso T, Suortti T, Hyötyläinen T, Achenbach P, Bonifacio E, et al. Age- and islet autoimmunity-associated differences in amino acid and lipid metabolites in children at risk for type 1 diabetes. *Diabetes* (2011) 60:2740–7. doi: 10.2337/db10-1652

11. Lamichhane S, Ahonen L, Dyrlund TS, Dickens AM, Siljander H, Hyöty H, et al. Cord-blood lipidome in progression to islet autoimmunity and type 1 diabetes. *Biomolecules* (2019) 9:33. doi: 10.3390/biom9010033

12. La Torre D, Seppänen-Laakso T, Larsson HE, Hyötyläinen T, Ivarsson SA, Lernmark A, et al. Decreased cord-blood phospholipids in young age-at-onset type 1 diabetes. *Diabetes* (2013) 62:3951–6. doi: 10.2337/db13-0215

13. Orešič M, Simell S, Sysi-Aho M, Näntö-Salonen K, Seppänen-Laakso T, Parikka V, et al. Dysregulation of lipid and amino acid metabolism precedes islet autoimmunity in children who later progress to type 1 diabetes. *J Exp Med* (2008) 205:2975–84. doi: 10.1084/jem.20081800

14. Johnson RK, Vanderlinden L, DeFelice BC, Kechris K, Uusitalo U, Fiehn O, et al. Metabolite-related dietary patterns and the development of islet autoimmunity. *Sci Rep* (2019) 9:14819. doi: 10.1038/s41598-019-51251-4

15. Sorensen CM, Ding J, Zhang Q, Alquier T, Zhao R, Mueller PW, et al. Perturbations in the lipid profile of individuals with newly diagnosed type 1 diabetes mellitus: lipidomics analysis of a diabetes antibody standardization program sample subset. *Clin Biochem* (2010) 43:948–56. doi: 10.1016/j.clinbiochem.2010.04.075

16. Overgaard AJ, Weir JM, Jayawardana K, Mortensen HB, Pociot F, Meikle PJ. Plasma lipid species at type 1 diabetes onset predict residual beta-cell function after 6 months. *Metabolomics* (2018) 14:158. doi: 10.1007/s11306-018-1456-3

17. Suvitaival T. Lipidomic abnormalities during the pathogenesis of type 1 diabetes: a quantitative review. *Curr Diabetes Rep* (2020) 20:46. doi: 10.1007/s11892-020-01326-8

18. Wu F, Liang P. Application of metabolomics in various types of diabetes. *Diabetes Metab Syndr Obes* (2022) 15:2051–9. doi: 10.2147/DMSO.S370158

19. Xu F, Tavintharan S, Sum CF, Woon K, Lim SC, Ong CN. Metabolic signature shift in type 2 diabetes mellitus revealed by mass spectrometry-based metabolomics. *J Clin Endocrinol Metab* (2013) 98:E1060–1065. doi: 10.1210/jc.2012-4132

20. Tillin T, Hughes AD, Wang Q, Würtz P, Ala-Korpela M, Sattar N, et al. Diabetes risk and amino acid profiles: cross-sectional and prospective analyses of ethnicity, amino acids and diabetes in a south Asian and European cohort from the SABRE (Southall and Brent REvisited) study. *Diabetologia* (2015) 58:968–79. doi: 10.1007/s00125-015-3517-8

21. Zhou K, Xie G, Wang J, Zhao A, Liu J, Su M, et al. Metabonomics reveals metabolite changes in biliary atresia infants. *J Proteome Res* (2015) 14:2569–74. doi: 10.1021/acs.jproteome.5b00125

22. Ni Y, Su M, Qiu Y, Jia W, Du X. ADAP-GC 3.0: Improved peak detection and deconvolution of Co-eluting metabolites from GC/TOF-MS data for metabolomics studies. *Anal Chem* (2016) 88:8802–11. doi: 10.1021/acs.analchem.6b02222

23. Pang Z, Chong J, Zhou G, de Lima Morais DA, Chang L, et al. MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res* (2021) 49:W388–96. doi: 10.1093/nar/gkab382

24. Ni Y, Qiu Y, Jiang W, Suttlemyre K, Su M, Zhang W, et al. ADAP-GC 2.0: Deconvolution of coeluting metabolites from GC/TOF-MS data for metabolomics studies. *Anal Chem* (2012) 84:6619–29. doi: 10.1021/ac300898h

25. Wei R, Wang J, Su M, Jia E, Chen S, Chen T, et al. Missing value imputation approach for mass spectrometry-based metabolomics data. *Sci Rep* (2018) 8:663. doi: 10.1038/s41598-017-19120-0

26. Song J, Kang SM, Lee WT, Park KA, Lee KM, Lee JE. Glutathione protects brain endothelial cells from hydrogen peroxide-induced oxidative stress by increasing Nrf2 expression. *Exp Neurobiol* (2014) 23:93–103. doi: 10.5607/en.2014.23.1.93

27. Xiao Z, La Fontaine S, Bush AI, Wedd AG. Molecular mechanisms of glutaredoxin enzymes: Versatile hubs for thiol–disulfide exchange between protein thiols and glutathione. *J Mol Biol* (2019) 431:158–77. doi: 10.1016/j.jmb.2018.12.006

28. Cleland SJ, Petrie JR, Small M, Elliott HL, Connell JMC. Insulin action is associated with endothelial function in hypertension and type 2 diabetes. *Hypertension* (2000) 35:507–11. doi: 10.1161/01.HYP.35.1.507

29. Grajeda-Iglesias C, Aviram M. Specific amino acids affect cardiovascular diseases and atherogenesis *via* protection against macrophage foam cell formation: Review article. *Rambam Maimonides Med J* (2018) 9:e0022. doi: 10.5041/RMMJ.10337

30. Newgard CB, An J, Bain JR, Muehlbauer MJ, Stevens RD, Lien LF, et al. A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell Metab* (2009) 9:311–26. doi: 10.1016/j.cmet.2009.02.002

31. Bhattacharya S, Granger CB, Craig D, Haynes C, Bain J, Stevens RD, et al. Validation of the association between a branched chain amino acid metabolite profile and extremes of coronary artery disease in patients referred for cardiac catheterization. *Atherosclerosis* (2014) 232:191–6. doi: 10.1016/j.atherosclerosis.2013.10.036

32. Yamanouchi T, Akanuma H, Nakamura T, Akaoka I, Akanuma Y. Reduction of plasma 1,5-anhydroglucitol (1-deoxyglucose) concentration in diabetic patients. *Diabetologia* (1988) 31:41–5. doi: 10.1007/BF00279131

33. Shen Y, Si Y, Lu J, Ma X, Zhang L, Mo Y, et al. Association between 1,5-anhydroglucitol and acute c peptide response to arginine among patients with type 2 diabetes. *J Diabetes Res* (2020) 2020:4243053. doi: 10.1155/2020/4243053

34. Lorenzi M. The polyol pathway as a mechanism for diabetic retinopathy: Attractive, elusive, and resilient. *Exp Diabetes Res* (2007) 2007:61038. doi: 10.1155/2007/61038

35. Srivastava SK, Ramana KV, Bhatnagar A. Role of aldose reductase and oxidative damage in diabetes and the consequent potential for therapeutic options. *Endocrine Rev* (2005) 26:380–92. doi: 10.1210/er.2004-0028

36. Suhre K, Meisinger C, Döring A, Altmaier E, Belcredi P, Gieger C, et al. Metabolic footprint of diabetes: A multiplatform metabolomics study in an epidemiological setting. *PloS One* (2010) 5:e13953. doi: 10.1371/journal.pone.0013953

37. Gibala MJ, MacLean DA, Graham TE, Saltin B. Tricarboxylic acid cycle intermediate pool size and estimated cycle flux in human muscle during exercise. *Am J Physiology-Endocrinology Metab* (1998) 275:E235–42. doi: 10.1152/ajpendo.1998.275.2.E235

38. Yki-Jarvinen H, Helve E, Koivisto VA. Hyperglycemia decreases glucose uptake in type I diabetes. *Diabetes* (1987) 36:5. doi: 10.2337/diab.36.8.892

39. Yki-Järvinen H, Mott D, Young AA, Stone K, Bogardus C. Regulation of glycogen synthase and phosphorylase activities by glucose and insulin in human skeletal muscle. *J Clin Invest.* (1987) 80:95–100. doi: 10.1172/JCI113069

40. Hubbard TD, Murray IA, Perdew GH. Indole and tryptophan metabolism: Endogenous and dietary routes to ah receptor activation. *Drug Metab Dispos* (2015) 43:1522–35. doi: 10.1124/dmd.115.064246

41. Dong F, Hao F, Murray IA, Smith PB, Koo I, Tindall AM, et al. Intestinal microbiota-derived tryptophan metabolites are predictive of ah receptor activity. *Gut Microbes* (2020) 12:1–24. doi: 10.1080/19490976.2020.1788899

42. Arrieta MC, Bistritz L, Meddings JB. Alterations in intestinal permeability. *Gut* (2006) 55:1512–20. doi: 10.1136/gut.2005.085373

43. Kaartinen K, Hemmilä U, Salmela K, Räisänen-Sokolowski A, Kouri T, Mäkelä S. Adenine phosphoribosyltransferase deficiency as a rare cause of renal allograft dysfunction. *J Am Soc Nephrol* (2014) 25:671–4. doi: 10.1681/ASN.2013090960

44. Nasr SH, Sethi S, Cornell LD, Milliner DS, Boelkins M, Broviac J, et al. Crystalline nephropathy due to 2,8-dihydroxyadeninuria: an under-recognized cause of irreversible renal failure. *Nephrol Dialysis Transplant* (2010) 25:1909–15. doi: 10.1093/ndt/gfp711

45. Rahman A, Yamazaki D, Sufiun A, Kitada K, Hitomi H, Nakano D, et al. A novel approach to adenine-induced chronic kidney disease associated anemia in rodents. *PloS One* (2018) 13:e0192531. doi: 10.1371/journal.pone.0192531

46. Fong D, Ullah MM, Lal JG, Abdelkader A, Ow CPC, Hilliard LM, et al. Renal cellular hypoxia in adenine-induced chronic kidney disease. *Clin Exp Pharmacol Physiol* (2016) 43:896–905. doi: 10.1111/1440-1681.12621

47. Santana AC, Degaspari S, Catanozi S, Dellê H, de Sá Lima L, Silva C, et al. Thalidomide suppresses inflammation in adenine-induced CKD with uraemia in mice. *Nephrol Dialysis Transplant* (2013) 28:1140–9. doi: 10.1093/ndt/gfs569

48. Milman Z, Axelrod JH, Heyman SN, Nachmansson N, Abramovitch R. Assessment with unenhanced MRI techniques of renal morphology and hemodynamic changes during acute kidney injury and chronic kidney disease in mice. *Am J Nephrol* (2014) 39:268–78. doi: 10.1159/000360093

frontiers | Frontiers in Endocrinology

TYPE Original Research
PUBLISHED 20 January 2023
DOI 10.3389/fendo.2023.1080138

OPEN ACCESS

EDITED BY
Marcel Twickler,
Monica Hospital Antwerp, Belgium

REVIEWED BY
Mehdi Sadeghian,
Ahvaz Jundishapur University of Medical Sciences, Iran
Seyedeh Parisa Moosavian,
Shiraz University of Medical Sciences, Iran
Reza Hashemi,
Urmia University of Medical Sciences, Iran

*CORRESPONDENCE
Mehran Rahimlou
Rahimlum@gmail.com

SPECIALTY SECTION
This article was submitted to Cardiovascular Endocrinology, a section of the journal Frontiers in Endocrinology

RECEIVED 25 October 2022
ACCEPTED 04 January 2023
PUBLISHED 20 January 2023

CITATION
Amirkhizi F, Khademi Z, Hamedi–Shahraki S and Rahimlou M (2023) Vitamin D insufficiency and its association with adipokines and atherogenic indices in patients with metabolic syndrome: A case-control study.
Front. Endocrinol. 14:1080138.
doi: 10.3389/fendo.2023.1080138

Vitamin D insufficiency and its association with adipokines and atherogenic indices in patients with metabolic syndrome: A case-control study

Farshad Amirkhizi[1], Zeinab Khademi[2], Soudabeh Hamedi–Shahraki[3] and Mehran Rahimlou[4]*

[1]Department of Nutrition, Faculty of Public Health, Zabol University of Medical Sciences, Zabol, Iran, [2]Department of Public Health, Sirjan School of Medical Sciences, Sirjan, Iran, [3]Department of Epidemiology and Biostatistics, Faculty of Public Health, Zabol University of Medical Sciences, Zabol, Iran, [4]Department of Nutrition, Faculty of Medicine, Zanjan University of Medical Sciences, Zanjan, Iran

Introduction: Vitamin D deficiency is one of the most common nutritional disorders in most countries of the world. The present study was designed and implemented with the aim of investigating the relationship between vitamin D deficiency and the level of adipokines, atherogenesis indicators and factors related to metabolic syndrome.

Methods: This case-control study was done on 195 patients with metabolic syndrome aged 20-50 y who attended the health centers in Zabol County, northeast Iran, between April 2021 and January 2022. Anthropometric and biochemical parameters were measured for all subjects with standard methods. To determine serum 25(OH)D levels, we used enzymatic linked immunosorbent assay (ELISA) kits. Atherogenic index of plasma (AIP) was calculated as log (TG/HDL-c). The visceral adiposity index (VAI) and the lipid accumulation product (LAP) were estimated according to standard formulas.

Results and Discussion: Participants in the case group had lower serum levels of 25(OH)D compared to controls (19.8 $\pm$ 6.2 ng/ml vs. 41.2 $\pm$ 9.7ng/ml, P<0.001). We found that the mean serum levels of fasting blood sugar (P=0.023) and TG (P=0.008) as well as HOMA-IR (P=0.023) were significantly higher in the cases compared to controls. Also, patients with MetS and vitamin D insufficiency (cases) had higher AIP (P=0.040) and LAP (P=0.012) than controls. Furthermore, serum 25 (OH)D levels showed significant inverse correlations with serum RBP-4 and a positive correlation with serum omentin-1 concentrations. The results of the present study showed that vitamin D deficiency correlated with some of the cardiometabolic risk factors among the patients with MetS.

KEYWORDS

vitamin D, metabolic syndrome, cardio metabolic, adipokines, cardiovacsular disease(s)

10.3389/fendo.2023.1080138

Introduction

Several studies have shown a rise in the percentage of overweight and obese adults in recent decades. According to the world health organization (WHO) reports, 39% of adults aged 18 years and over were overweight in 2016, and 13% were obese. Also, it has been reported that 39 million children under the age of 5 were overweight or obese in 2020 (1).

Adults, teenagers, and children are all affected by the issue (2–4). Numerous systemic illnesses, such as metabolic syndrome (MetS), type 2 diabetic mellitus (T2DM), atherosclerosis, cardiovascular problems, and cancers, are caused on by obesity (5). Currently, the prevalence of T2DM is rising, and by 2040, it has been estimated that over than 642 million persons (10% of the population) will have this disease. People who have low levels of serum 25-hydroxyvitamin D3 [25(OH) D3] have a considerably greater risk of developing T2DM and MetS (6, 7). Vitamin D deficiency have been linked to a higher odds ratio of MetS and T2DM, according to some of cohort studies (8, 9). It has been reported that vitamin D receptor (VDR) changes involved in the pathogenesis of some chronic disorders such as diabetes (10), autoimmune diseases (11), nonalcoholic liver disease (12), cardiovascular disease(CVD) (13), and cancer (14).

Individuals with risk factors for CVD frequently have low serum 25(OH)D3 concentration. It has been reported that there was a significant inverse association between serum levels of 25(OH)D3 and some cardiometabolic risk factors such as fasting blood sugar (FBS), hemoglobin A1c (HbA1c), total cholesterol (TC), triglyceride (TG), body mass index (BMI), waist circumference (WC), and atherogenic indices (Castelli Risk Index I (CRI I), Castelli Risk Index II (CRI II), and Atherogenic index of plasma (AIP) (15).

Obesity, especially central obesity, is one of the most important risk factors for T2DM, which causes insulin resistance and inflammation due to the increase of fat tissue in the body, which also increases the risk of CVD (16). The mechanisms involved in the relationship between obesity, adipose tissue dysfunction and MetS have not yet been precisely identified (17). It has been reported in some pervious research that adipose tissue metabolic alterations, such as dysregulated adipokine production, may mediate the association between various obesity phenotypes and vulnerability to some chronic disorders such as MetS (18). Adipokines have long been recognized as significant hunger and satiety moderators as well as critical regulators of energy homeostasis, inflammation, immunological function, blood pressure, vascular function, insulin levels, and glucose and lipid metabolism (19).

In recent years, some studies have shown that there is a bidirectional relationship between vitamin D and obesity (20, 21). In obese people, due to the higher fat mass, high amounts of vitamin D are trapped in fat tissues, and therefore, these people may need higher amounts of vitamin D. On the other hand, some findings have revealed that vitamin D supplementation may have a positive effect on obesity prevention and treatment (22, 23). In addition, some researchers have suggested that some effects of vitamin D deficiency in the etiology of obesity and T2DM are exerted through disturbances in the concentration of adipokines such as leptin and adiponectin (22, 24).

Therefore, considering the importance of vitamin D in the prevention and treatment of metabolic syndrome and the existence of conflicting evidence regarding the effective mechanisms in the results observed in previous studies, the present study designed to investigate the relationship between vitamin D deficiency with adipokines and atherogenic indices in patients with metabolic syndrome.

Material and methods

Study participants

This case-control study was done on 195 patients with Mets aged 20-50 y who attended the health centers in Zabol County, southeast Iran, between April 2021 and January 2022. Serum omentin-1 levels as a key variable obtained from Zorlu et al. study (25), was used to estimate the sample size. Considering the study power of 80%, a type I error of 5%, and ratio of controls to cases as 2, we required 65 cases and 130 controls for this study.

Cases were MetS patients with vitamin D insufficiency and controls were MetS patients with optimal vitamin D status. Controls were frequency-matched with cases on age (± 2 years), sex, and BMI (± 1 kg/m^2) and were selected from patients visiting the same health centers. Vitamin D insufficiency was defined, according to the previous studies and Endocrine Society clinical practice guidelines (26, 27), as serum 25(OH)D levels < 30 ng/ml and those who had serum 25(OH)D levels ≥ 30 ng/ml were considered sufficient.

MetS was diagnosed based on the National Cholesterol Education Program-Adult Treatment Panel III (NCEP-ATP III) consensus (28) as the presence of at least three of the following criteria: (1) a waist circumference (WC) > 102 cm in males and > 88 cm in females, (2) a fasting blood sugar (FBS) ≥ 100 mg/dl, (3) a high serum triglyceride (≥150 mg/dl), (4) a low serum HDL-c (< 40 mg/dl in males and < 50 mg/dl in females), (5) a high blood pressure (SBP ≥ 130 mmHg or DBP ≥ 80 mmHg).

In this study, individuals with a prior history of cardiovascular diseases, endocrine disorders such as diabetes and hypo-/hyperthyroidism, cancer, renal or liver dysfunction were excluded. The individuals were also excluded if taking antioxidant supplements like selenium, carotenoids, and vitamins E and C, as well as any medications known to induce metabolic or hormonal changes such as estrogens, corticosteroid drugs and lipid-lowering medications within three months before enrollment the study. Following a specific diet, using fish oil supplement and taking anti-inflammatory medications in the past three months were among the exclusion criteria as well. Eligible individuals, including 65 cases and 130 controls were recruited for the study.

All study participants signed an informed consent form after explaining the aims and the study methodology. The study protocol was approved by the Ethics Committee of Zabol University of Medical Sciences (Approval code: IR.ZBMU.REC.1399.156).

Assessment of anthropometric variables, blood pressure and physical activity

Anthropometric parameters were measured for all subjects. Height was measured without shoes in an upright position using a fixed non-stretchable tape with a precision of 0.1 cm. Weight was measured on light clothing by a Seca scale to the nearest 0.1 kg. Body

mass index (BMI) was calculated as weight (kg) divided by squared height (m^2). Finally, the waist circumference (WC) was measured between the lower rib margin and the iliac crest at the end of a normal expiration. A bioelectrical impedance analysis (BIA) system (InBody S10, JMW140, Korea) was applied to measure the percentage of body fat mass (%FM) and visceral fat level (%VF).

Blood pressure was measured in a sitting position, after a 15-min rest, using a mercury sphygmomanometer twice with 5-min intervals. The mean of two measurements was considered as the final systolic (SBP) and diastolic (DBP) blood pressure.

To evaluate the physical activity levels of the participants, we applied a short form of International physical activity questionnaire (IPAQ) and then classified into 3 categories of "low," "moderate," and "vigorous" activity according to the IPAQ scoring guideline (29).

Laboratory measurements

Blood samples were collected from all patients after almost 12 h of overnight fasting and centrifuged at 3500 rpm for 10 min to separate the sera. Then, the separated sera were immediately distributed in aliquots and stored at -70°C until analysis.

Fasting blood sugar (FBS) as well as serum levels of total cholesterol (TC), high-density lipoprotein cholesterol (HDL-c), and low-density lipoprotein cholesterol (LDL-c) were determined using the standard enzymatic-colorimetric method on an automatic biochemical Hitachi 717 analyzer (Hitachi, Boehringer Mannheim, Japan) through commercial kits (Pars-Azmoon Co., Tehran, Iran) with inter- and intra-assay coefficient variances (CVs) of lower than 5%. Non-HDL-cholesterol was calculated as the TC minus HDL-c. Serum insulin levels were determined based on the radioimmunoassay method using the commercial kit (Immunotech, Prague, Czech Republic), which is sensitive to 0.5 mU/ml variations in serum levels of insulin and its intra- and inter-assay coefficients of variation were 3.8% and 6.2%, respectively. Insulin resistance was estimated with the homeostasis model assessment method (HOMA-IR) using the suggested equation of Matthews et al. (30): HOMA-IR= [fasting insulin (U/l) × fasting glucose (mg/dl)]/405.

Atherogenic index of plasma (AIP) was calculated as log (TG/HDL-c) (31). The visceral adiposity index (VAI) and the lipid accumulation product (LAP) were estimated according to suggested formulas for each gender as follows:

$$\text{Males}: \text{VAI} = [\text{WC}/(39.68 + 1.88 \times \text{BMI})] \times (\text{TG}/1.03) \times (1.31/\text{HDL-c}) \quad (32)$$

$$\text{Females}: \text{VAI} = [\text{WC}/(36.58 + 1.89 \times \text{BMI})] \times (\text{TG}/0.81) \times (1.52/\text{HDL-c})$$

$$\text{Males}: \text{LAP} = [\text{WC (cm)} - 65] \times [\text{TG (mmol/l)}] \quad (33)$$

$$\text{Females}: \text{LAP} = [\text{WC (cm)} - 58] \times [\text{TG (mmol/l)}]$$

To determine serum 25(OH)D levels, we used enzymatic linked immunosorbent assay (ELISA) kits (DIAsource Immunoassays SA, Louvain-laNeuve, Belgium) according to the manufacturer's instructions. Serum levels of omentin-1, chemerin, vaspin, and retinol binding protein 4 (RBP-4) were measured using the Human ELISA kits (Shanghai Crystal Day Biotech Co., Ltd.) according to the manufacturer's instructions. The intra- and inter-assay CV of adipokines was lower than 4.7% and 7.8%, respectively.

Statistical analysis

The results were presented as mean ± standard deviation for normally distributed continuous data and frequency (percent) for categorical data. The non-normally distributed data were expressed as the median and interquartile range (IQR). General characteristics were compared between cases and controls using an independent samples t-test and Pearson chi-squared test, as appropriate. In addition, between-groups differences in normally and non-normally distributed cardiometabolic parameters and adipokines were investigated using independent samples t-test and non-parametric Mann–Whitney U-test, where appropriate. To examine the association between serum levels of 25(OH)D with adipokines and cardiometabolic parameters, multiple linear regression was applied by adjusting for age, sex, smoking, physical activity level, and BMI. Data were analyzed using IBM SPSS version 25 (IBM Corp., Armonk, NY, USA) and statistical significance was considered as a p-value< 0.05.

Results

Characteristics of cases and controls

The study participants consisted of 195 patients with MetS (65 cases and 130 controls). The mean age of cases and controls was 38.0 ± 5.5 and 37.5 ± 5.6 y, respectively that was non-significant. Totally, 34% of patients participating in this study were male and 66% were female. No significant differences were found between cases and controls with respect to age, BMI, WC, fat mass, and visceral fat (Table 1) (P>0.05). In addition, the distribution of participants with regard to sex, smoking status, PAL, and education level was not significantly different between the study groups (P>0.05). As shown in Table 1, participants in the case group had lower serum levels of 25(OH)D compared to controls (19.8 ± 6.2 ng/ml vs. 41.2 ± 9.7*ng/ml*, P<0.001).

Cardiometabolic parameters in cases and controls

Table 2 shows the differences in the cardiometabolic parameters between the cases and controls.

The mean serum levels of FBS (*P*=0.023) and TG (*P*=0.008) as well as HOMA-IR (*P*=0.023) were significantly higher in the cases compared to controls. Furthermore, patients with MetS and vitamin D insufficiency (cases) had higher AIP (*P*=0.040) and LAP (*P*=0.012) than controls, whereas, there were no significant differences in serum levels of insulin, TC, LDL-c, HDL-c, and non-HDL as well as VAI, SBP, and DBP.

10.3389/fendo.2023.1080138

TABLE 1 General characteristics of cases and controls.

Variables	Cases (n=65)	Controls (n=130)	*P*-value
Male, *n* (%)	21 (32.3)	45 (34.6)	0.748†
Age (years)	38.0 ± 5.5	37.5 ± 5.6	0.556‡
Weight (kg)	89.7 ± 11.6	90.5 ± 12.1	0.673‡
BMI (kg/m^2)	32.0 ± 2.1	31.6 ± 1.9	0.289‡
WC (cm)	96.7 ± 8.0	94.7 ± 9.4	0.141‡
Fat mass (%)	44.3 ± 7.2	42.8 ± 7.3	0.185‡
Visceral fat (%)	13.2 ± 4.1	12.8 ± 4.0	0.534‡
PAL, *n* (%)			
Low	47 (72.3)	90 (69.2)	0.904†
Moderate	15 (23.1)	33 (25.4)	
Vigorous	3 (4.9)	7 (5.4)	
Smokers, *n* (%)	14 (21.5)	31 (23.8)	0.718†
Education level, *n* (%)			
Primary school	24 (36.9)	49 (37.7)	0.776†
Secondary & high school	26 (40.0)	46 (35.4)	
Diploma & university	15 (23.1)	35 (26.9)	
25(OH)D (ng/ml)	19.8 ± 6.2	41.2 ± 9.7	<0.001‡

BMI, body mass index; WC, waist circumference; PAL, physical activity level.
Values are reported as mean ± standard deviation for continuous variables and number (%) for categorical variables.
P-values<0.05 were considered significant.
†Pearson chi-square test.
‡Independent samples t-test.

TABLE 2 Comparison of cardiometabolic parameters between cases and controls.

Variables	Cases (n=65)	Controls (n=130)	*P*-value†
FBS (mg/dL)	106.1 ± 13.6	101.9 ± 11.1	0.023
Insulin (μU/mL)	18.1 ± 5.8	16.8 ± 5.2	0.116
HOMA-IR	4.8 ± 2.1	4.2 ± 1.5	0.032
TG (mg/dL)	190.0 ± 27.7	180.0 ± 23.1	0.008
TC (mg/dL)	197.5 ± 23.1	191.8 ± 23.9	0.113
LDL-c (mg/dL)	137.3 ± 27.2	132.5 ± 23.4	0.201
HDL-c (mg/dL)	46.7 ± 5.8	47.1 ± 6.4	0.719
non-HDLC	150.8 ± 23.3	144.8 ± 24.3	0.101
AIP	0.61 ± 0.08	0.58 ± 0.08	0.040
LAP	78.25 ± 21.4	69.79 ± 22.20	0.012
VAI	0.99 ± 0.08	0.97 ± 0.09	0.239
SBP (mmHg)	127.2 ± 3.8	126.5 ± 4.1	0.304
DBP (mmHg)	83.7 ± 3.2	82.9 ± 2.7	0.083

FBS, fasting blood sugar; HOMA-IR, homeostasis model assessment of insulin resistance; TG, triglyceride; TC, total cholesterol; LDL-c, low-density lipoprotein cholesterol; HDL-c, high-density lipoprotein cholesterol; AIP, atherogenic index of plasma; LAP, lipid accumulation product; VAI, visceral adiposity index; SBP, systolic blood pressure; DBP, diastolic blood pressure.
Values are reported as mean ± standard deviation.
P-values<0.05 were considered significant.
†Independent samples t-test.

Serum adipokines in cases and controls

The comparisons of the serum levels of adipokines between the cases and controls are shown in Figure 1. The mean serum levels of omentin-1 were significantly higher (P=0.007) and serum levels of RBP-4 (P=0.007) were significantly lower in the cases compared to controls. However, we failed to find any significant difference in mean serum levels of vaspin and chemerin between the two groups.

Association of serum vitamin D with serum adipokines and cardiometabolic parameters

The results of multiple linear regression models, which investigated the associations between serum 25(OH)D levels with serum adipokines and cardiometabolic parameters are reported in Table 3. Serum 25(OH)D levels showed significant inverse correlations with serum RBP-4 (β=-0.305, P<0.001), FBS (β=-0.189, P=0.007), insulin (β=-0.170, P=0.017), HOMA-IR (β=-0.206, P=0.004), TG (β=-0.431, P<0.001) as well as AIP (β=-0.360, P<0.001) and LAP (β=-0.281, P<0.001) after adjustment for potential confounders including age, sex, PAL, and BMI. There were significant positive correlations between serum 25(OH)D levels and serum levels of omentin-1 (β=0.197, P=0.007). When regression analyses were performed by adjustment for potential confounders, serum 25(OH)D levels did not show any significant correlations with serum chemerin, vaspin, TC, LDL-c, HDL-c, non-HDL, SBP, DBP as well as VAI.

Discussion

The results of the present study were showed that participants with MetS and vitamin D insufficiency had higher levels of some risk factors for CVD. Also, we found an inverse association between serum levels of 25(OH)D and some CVD related biomarkers.

Vitamin D insufficiency is currently a major global health issue due to its high incidence (34). According to the results of a meta-analysis study, the prevalence of vitamin D deficiency in the Iranian population is 56% (64% in women and 44% in men) (35). Given that the majority of body cells have vitamin D receptors and para/autocrine vitamin D metabolic activity, vitamin D effects extend beyond the regulation of bone tissue (36, 37). The potential therapeutic benefits of vitamin D3 can be obtained by keeping vitamin levels in adults in the range of 30-80 ng/mL (75-200 nm/L), according to expert recommendations that published in 2018 (38). We found that the serum levels of FBS, TG and HOMAIR in the participants with vitamin D deficiency were significantly higher than

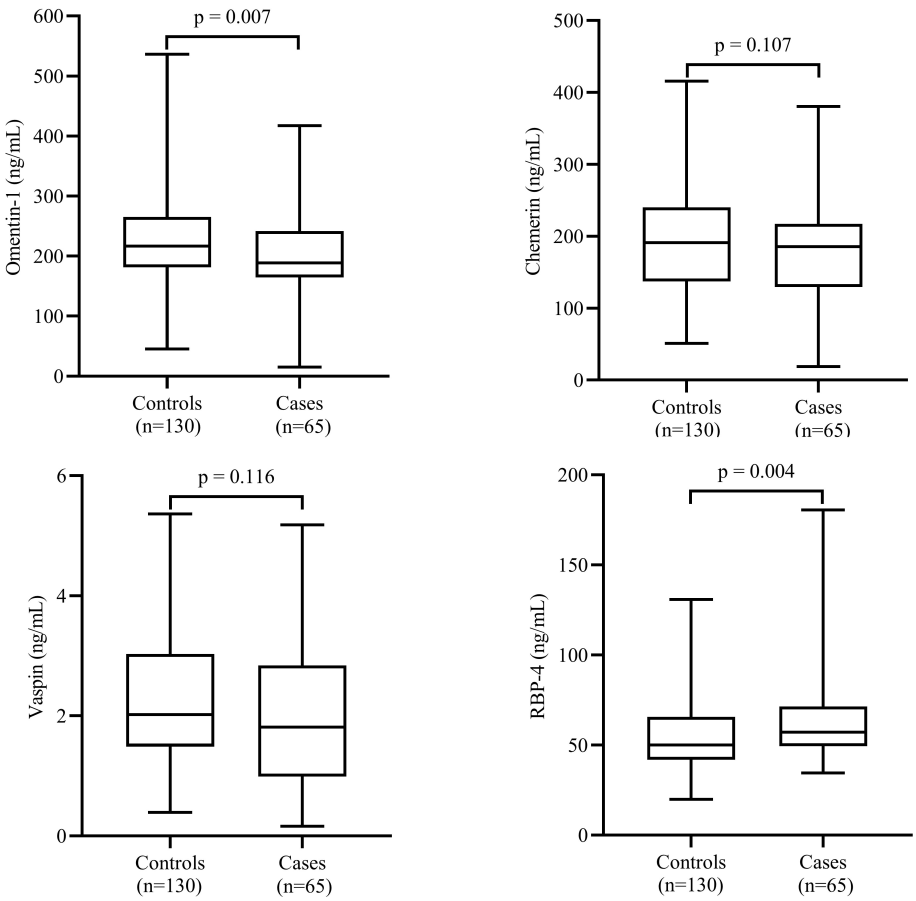


FIGURE 1
Comparison of serum levels of adipokines between cases and controls. Data are shown as median (interquartile range). P-values obtained from Mann–Whitney U test. P-values < 0.05 were considered significant.

10.3389/fendo.2023.1080138

TABLE 3 Multiple linear regression analysis for the association between serum 25(OH)D concentrations with adipokines and cardiometabolic parameters (*n*=195).

Model	B	S.E.	β	*P*-value†
Omentin-1 (ng/mL)	1.028	0.376	0.197	0.007
Chemerin (ng/mL)	0.433	0.341	0.094	0.205
Vaspin (ng/mL)	0.010	0.006	0.118	0.110
RBP-4 (ng/mL)	-0.476	0.110	-0.305	<0.001
FBS (mg/dL)	-0.172	0.063	-0.189	0.007
Insulin (μU/mL)	-0.071	0.030	-0.176	0.017
HOMA-IR	-0.027	0.009	-0.206	0.004
TG (mg/dL)	-0.811	0.125	-0.431	<0.001
TC (mg/dL)	-0.188	0.130	-0.106	0.150
LDL-c (mg/dL)	-0.158	0.136	-0.085	0.248
HDL-c (mg/dL)	-0.037	0.033	-0.081	0.266
non-HDLC	-0.225	0.132	-0.125	0.089
AIP	-0.003	0.001	-0.360	<0.001
LAP	-0.468	0.109	-0.281	<0.001
VAI	-0.002	0.001	-0.116	0.115
SBP (mmHg)	-0.033	0.022	-0.110	0.134
DBP (mmHg)	-0.019	0.016	-0.089	0.223

RBP-4, retinol binding protein 4; FBS, fasting blood sugar; HOMA-IR, homeostasis model assessment of insulin resistance; TG, triglyceride; TC, total cholesterol; LDL-c, low-density lipoprotein cholesterol; HDL-c, high-density lipoprotein cholesterol; AIP, atherogenic index of plasma; LAP, lipid accumulation product; VAI, visceral adiposity index; SBP, systolic blood pressure; DBP, diastolic blood pressure; B, unstandardized coefficient; S.E., standard error.
†Adjusted for age, sex, smoking, physical activity level, and BMI.
P-values<0.05 were considered significant.

control group. Also, we found a significant inverse association between 25(OH)D levels with serum levels of FBS, TG and HOMAIR. In line with our findings, Schmitt et al. were reported that in postmenopausal women, vitamin D deficiency was correlated with higher prevalence of MetS, as well as hypertriglyceridemia (39). In other study, Lee et al. found that lower serum levels of 25(OH)D was associated with higher waist circumference, TG and insulin resistance (40). Likewise, Zhu and Heil reported that 25(OH)D levels were inadequate in 50% of the study population, formed by residents of Shanghai, China, aged 19–70 years and were associated with the presence of MetS, and they found that higher serum levels of vitamin D was associated with lower LDL and total cholesterol concentration (41). Increasing evidence points to a strong association between vitamin D insufficiency and a decrease in insulin secretion in both people and animal models (42). Numerous studies have suggested that reduced levels of vitamin D disrupt insulin sensitivity, beta-cell function, or both to cause the onset of T2DM and insulin resistance (43–46). On the other hand, some studies evaluated the effect of vitamin D supplementation in people suffering from vitamin D deficiency and its effect on factors related to metabolic syndrome. In research that was conducted among the Asian women with insulin resistance and baseline 25(OH)D levels below 20 ng/mL, it has been reported that vitamin D supplementation with dose of 4,000 IU led to a significant improvement in the insulin sensitivity (47). Also, in a systematic review and meta-analysis study on 19 randomized controlled trials (RCT), it has been reported that compared with the control group, the short-term vitamin D supplementation group had a significant reduction in HbA1c, insulin resistance, and insulin concentration (48). The mechanisms of vitamin D reducing the risk of T2D include, improved insulin sensitivity, and reduced insulin resistance (49–51). It has also been shown that vitamin D reduces inflammation as one of the main factors involved in the pathogenesis of T2DM due the dysfunction in the insulin sensitivity and pancreatic beta-cell function (52–54).

In the present study, we found that participants in the vitamin D deficiency group had higher AIP and LAP than control group. Wang et al. in a population-based study on 1475 participants showed that serum 25(OH)D concentration was inversely correlated with TG, LDL-C and AIP (55). In other study, Mahmoodi et al. in a case control study among the participants with T2DM found that all of the atherogenic indices including AIP, CRII, and Atherogenic Coefficient (AC) significantly reduced with improved serum vitamin D status (56). It has been reported in several studies that vitamin D deficiency increase risk of dyslipidemia (57, 58). According to results from the Karhapää et al. investigation, in middle-aged Finnish men, serum 25(OH)D levels are inversely related to TC, TG, and LDL-C (59). The results of a cross-sectional study on the Danish population showed that a 10-unit increase in vitamin D concentration was associated with a significant decrease in the level of dyslipidemia-related factors (60). AIP value, which is obtained by applying a

logarithmic transformation to the result of dividing the plasma TG value by the plasma HDL value, has recently been discovered to be a reliable indicator of the risk of atherosclerosis and CVD (61). In our study, AIP value in the case group was 0.61 ± 0.08 and in control group was 0.58 ± 0.08, which was significantly different. It has been reported in some previous studies that AIP value between 0.1–0.24 shows medium cardiac risk (62). Wu et al. were reported that AIP was an independent predictor of CAD (63). Compared to LDL-C or TC, AIP has been found to be a more helpful indicator of atherogenicity and CVD risk (64). Because vitamin D signaling reduces the expression of TNF-alpha, IL-6, IL-1, and IL-8 in isolated blood monocytes, it may have an impact on the pathogenesis of atherosclerosis (65, 66). By increasing nuclear factor kappa B (NF-κB) activation, vitamin D insufficiency was found to hasten the development of coronary artery disease in pigs, thereby demonstrating the anti-inflammatory properties of vitamin D (67). A characteristic of the development of atherosclerosis is the creation of foam cells originating from macrophages (68). It has been reported in previous studies that vitamin D can decrease cholesterol accumulation in macrophages and LDL uptake in Atheromas (69). Additionally, it alters the expression of tissue factor and thrombomodulin in monocytes, which has an impact on platelet aggregation and thrombogenic activity (70). In the Nakagawa et al. study, 1,25(OH)2D decreased matrix metalloproteinase (MMP)-2 and MMP-9 expression in cell culture, potentially avoiding plaque instability, luminal rupture, and thrombosis (71). Furthermore, 1,25 (OH)2D inhibited foam cell production in macrophages isolated from patients with hypertension, diabetes, and obesity (72).

We also found that serum 25(OH)D levels showed significant inverse correlations with serum RBP-4 and positive correlation with omentin-1. In line with our findings, Dikker et al. found that omentin concentration had a positive correlation with vitamin D levels among the postmenopausal women (73). Also, Nazar et al. in another case control study found a linear correlation between vitamin D status and omentin-1 levels and also reported that vitamin D and omentin-1 deficiency maybe increase the risk of coronary artery disease (74). In contrast with these findings, Zorlu et al. in a cross-sectional study among the 77 female volunteers were reported that Omentin levels correlated significantly and negatively with the vitamin D (25). Omentin-1 is a 34-kDa, anti-inflammatory, circulating adipocytokine, has been considered to have a significant role in endothelial dysfunction, atherosclerosis and myocardial remodeling (75). It has been reported that omentin-1 exert its anti-inflammatory effects by suppressing some of the cytokines cascades and factors such as TNF-alpha. Also, omentin-1 induce the production of activated B cells in endothelial cells *via* nuclear factor kappa-light-chain. Moreover, the protein kinase (5'AMP) that activated by Omentin-1 can suppress the expression of vascular adhesion molecule E-selectin (74). Many researchers found that the serum levels of RBP-4 are associated with risk of metabolic syndrome (76, 77). Jialal and cols. reported that serum RBP-4 concentration would be independent predictors of CVD in diabetes (78).

Based on our knowledge, the present study is the first study that evaluated the association between vitamin D status with several cardiometabolic factors among the patients with MetS. The current study had several strengths, including the evaluation of several biochemical factors related to the risk of cardiovascular diseases, strong methodology and controlling the covariates between case and control. However, there were some limitations in the present study that should be considered in interpreting the results. Firstly, because of the study design, the cause-and-effect association will not be possible. Second, although the sample size in this study is 195, some potential associations between vitamin D and individual biomarkers of MetS might have not surfaced. Higher sample size might have discovered some additional relationships between vitamin D and markers of MetS.

Conclusion

The results of the present study showed that participants with vitamin D deficiency had higher concentration of some MetS and CVD related biomarkers. Also, we found a significant association between 25(OH)D status and some of the adipokines and atherogenic indices in patients with MetS. Given the widespread vitamin D deficiency among the Iranian participants specially women, it has been suggested that fortification of some staple foods such as vegetable oils or cereals and dairy products with this vitamin can be appropriate therapeutic strategies to improve the vitamin D status in society. More studies with a higher sample size, especially clinical trials with strong methodology, are needed to re-evaluate the effect of vitamin D supplementation on the factors investigated in this study.

Data availability statement

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

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee of Zabol University of Medical Sciences. The patients/participants provided their written informed consent to participate in this study.

Author contributions

FA, MR, and ZK designed the study and analyzed data. SH-S, FA, and ZK cooperated in the implementation of the study. All authors contributed to the article and approved the submitted version.

Conflict of interest

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

Publisher's note

References

1. Chooi YC, Ding C, Magkos F. The epidemiology of obesity. *Metabolism* (2019) 92:6–10. doi: 10.1016/j.metabol.2018.09.005

2. James WP. The epidemiology of obesity: the size of the problem. *J Intern Med* (2008) 263:336–52. doi: 10.1111/j.1365-2796.2008.01922.x

3. De Onis M, Blössner M, Borghi E. Global prevalence and trends of overweight and obesity among preschool children. *Am J Clin Nutr* (2010) 92:1257–64. doi: 10.3945/ajcn.2010.29786

4. Park MH, Falconer C, Viner RM, Kinra S. The impact of childhood obesity on morbidity and mortality in adulthood: A systematic review. *Obes Rev* (2012) 13:985–1000. doi: 10.1111/j.1467-789X.2012.01015.x

5. Eckel RH, Grundy SM, Zimmet PZ. The metabolic syndrome. *Lancet* (2005) 365:1415–28. doi: 10.1016/S0140-6736(05)66378-7

6. Gao Y, Zheng T, Ran X, Ren Y, Chen T, Zhong L, et al. Vitamin d and incidence of prediabetes or type 2 diabetes: a four-year follow-up community-based study. *Dis Markers* (2018) 2018:192-201. doi: 10.1155/2018/1926308

7. Akter S, Kuwahara K, Matsushita Y, Nakagawa T, Konishi M, Honda T, et al. Serum 25-hydroxyvitamin D3 and risk of type 2 diabetes among Japanese adults: the Hitachi health study. *Clin Nutr* (2020) 39:1218–24. doi: 10.1016/j.clnu.2019.05.010

8. Paari NA. Comparative study of serum vitamin d levels in type 2 diabetes mellitus patients and age matched controls.

9. González-Molero I, Rojo-Martínez G, Morcillo S, Gutiérrez-Repiso C, Rubio-Martín E, Almaraz MC, et al. Vitamin d and incidence of diabetes: a prospective cohort study. *Clin Nutr* (2012) 31:571–3. doi: 10.1016/j.clnu.2011.12.001

10. Isnuwardana R, Bijukchhe S, Thadanipon K, Ingsathit A, Thakkinstian A. Association between vitamin d and uric acid in adults: A systematic review and meta-analysis. *Hormone Metab Res* (2020) 52:732–41. doi: 10.1055/a-1240-5850

11. Dankers W, González-Leal C, Davelaar N, Asmawidjaja PS, Mus A, Hazes JM, et al. 1, 25 (OH) 2D3 and dexamethasone additively suppress synovial fibroblast activation by CCR6+ T helper memory cells and enhance the effect of tumor necrosis factor alpha blockade. *Arthritis Res Ther* (2018) 20:1–10. doi: 10.1186/s13075-018-1706-9

12. Pacifico L, Osborn JF, Bonci E, Pierimarchi P, Chiesa C. Association between vitamin d levels and nonalcoholic fatty liver disease: potential confounding variables. *Mini Rev Medicinal Chem* (2019) 19:310–32. doi: 10.2174/1389557518666181025153712

13. Mohammad S, Mishra A, Ashraf MZ. Emerging role of vitamin d and its associated molecules in pathways related to pathogenesis of thrombosis. *Biomolecules* (2019) 9:649. doi: 10.3390/biom9110649

14. Jeon S-M, Shin E. Exploring vitamin d metabolism and function in cancer. *Exp Mol Med* (2018) 50:1–14. doi: 10.1038/s12276-018-0038-9

15. Barbalho SM, Tofano RJ, De Campos AL, Rodrigues AS, Quesada K, Bechara MD, et al. Association between vitamin d status and metabolic syndrome risk factors. *Diabetes Metab Syndrome: Clin Res Rev* (2018) 12:501–7. doi: 10.1016/j.dsx.2018.03.011

16. Frydrych LM, Bian G, O'lone DE, Ward PA, Delano MJ. Obesity and type 2 diabetes mellitus drive immune dysfunction, infection development, and sepsis mortality. *J Leukocyte Biol* (2018) 104:525–34. doi: 10.1002/JLB.5VMR0118-021RR

17. Böhm A, Halama A, Meile T, Zdichavsky M, Lehmann R, Weigert C, et al. Metabolic signatures of cultured human adipocytes from metabolically healthy versus unhealthy obese individuals. *PloS One* (2014) 9:e93148. doi: 10.1371/journal.pone.0093148

18. Alizadeh S, Mirzaei K, Mohammadi C, Keshavarz SA, Maghbooli Z. Circulating omentin-1 might be associated with metabolic health status in different phenotypes of body size. *Arch Endocrinol Metab* (2017) 61:567–74. doi: 10.1590/2359-3997000000269

19. Moradi S, Mirzaei K, Abdurahman AA, Keshavarz SA, Hossein-Nezhad A. Mediatory effect of circulating vaspin on resting metabolic rate in obese individuals. *Eur J Nutr* (2016) 55:1297–305. doi: 10.1007/s00394-015-0948-4

20. Vranić L, Mikolašević I, Milić S. Vitamin d deficiency: Consequence or cause of obesity? *Medicina* (2019) 55:541. doi: 10.3390/medicina55090541

21. Bennour I, Haroun N, Sicard F, Mounien L, Landrier J-F. Vitamin d and Obesity/Adiposity–a brief overview of recent studies. *Nutrients* (2022) 14:2049. doi: 10.3390/nu14102049

22. Bassatne A, Chakhtoura M, Saad R, Fuleihan GE-H. Vitamin d supplementation in obesity and during weight loss: A review of randomized controlled trials. *Metabolism* (2019) 92:193–205. doi: 10.1016/j.metabol.2018.12.010

23. Migliaccio S, Di Nisio A, Mele C, Scappaticcio L, Savastano S, Colao A. Obesity and hypovitaminosis d: Causality or casualty? *Int J Obes Suppl* (2019) 9:20–31. doi: 10.1038/s41367-019-0010-8

24. Mousa A, Naderpoor N, Wilson K, Plebanski M, De Courten MP, Scragg R, et al. Vitamin d supplementation increases adipokine concentrations in overweight or obese adults. *Eur J Nutr* (2020) 59:195–204. doi: 10.1007/s00394-019-01899-5

25. Zorlu M, Kiskac M, Cakirca M, Karatoprak C, Güler E, Çelik K, et al. Evaluation of the relation between vitamin d and serum omentin and vaspin levels in women. *Exp Clin Endocrinol Diabetes* (2016) 124:440–3. doi: 10.1055/s-0042-108853

26. Aoun A, Maalouf J, Fahed M, El Jabbour F. When and how to diagnose and treat vitamin d deficiency in adults: A practical and clinical update. *J Dietary Suppl* (2020) 17:336–54. doi: 10.1080/19390211.2019.1577935

27. Amirkhizi F, Pishdadian A, Asghari S, Hamedi-Shahraki S. Vitamin d status is favorably associated with the cardiovascular risk factors in adults with obesity. *Clin Nutr ESPEN* (2021) 46:232–9. doi: 10.1016/j.clnesp.2021.10.003

28. Grundy SM, Cleeman JI, Daniels SR, Donato KA, Eckel RH, Franklin BA, et al. Diagnosis and management of the metabolic syndrome: an American heart Association/National heart, lung, and blood institute scientific statement. *Circulation* (2005) 112:2735–52. doi: 10.1161/CIRCULATIONAHA.105.169404

29. Hallal PC, Victora CG. Reliability and validity of the international physical activity questionnaire (IPAQ). *Med Sci Sports Exerc* (2004) 36:556. doi: 10.1249/01.MSS.0000117161.66394.07

30. Matthews DR, Hosker J, Rudenski A, Naylor B, Treacher D, Turner R. Homeostasis model assessment: insulin resistance and β-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* (1985) 28:412–9. doi: 10.1007/BF00280883

31. Niroumand S, Khajedaluee M, Khadem-Rezaiyan M, Abrishami M, Juya M, Khodaee G, et al. Atherogenic index of plasma (AIP): A marker of cardiovascular disease. *Med J Islamic Republic Iran* (2015) 29:240.

32. Amato MC, Giordano C, Galia M, Criscimanna A, Vitabile S, Midiri M, et al. Visceral adiposity index: a reliable indicator of visceral fat function associated with cardiometabolic risk. *Diabetes Care* (2010) 33:920–2. doi: 10.2337/dc09-1825

33. Kahn HS. The" lipid accumulation product" performs better than the body mass index for recognizing cardiovascular risk: a population-based comparison. *BMC Cardiovasc Disord* (2005) 5:1–10. doi: 10.1186/1471-2261-5-26

34. Palacios C, Gonzalez L. Is vitamin d deficiency a major global public health problem? *J Steroid Biochem Mol Biol* (2014) 144:138–45. doi: 10.1016/j.jsbmb.2013.11.003

35. Vatandost S, Jahani M, Afshari A, Amiri MR, Heidarimoghadam R, Mohammadi Y. Prevalence of vitamin d deficiency in Iran: A systematic review and meta-analysis. *Nutr Health* (2018) 24:269–78. doi: 10.1177/0260106018802968

36. Holick MF, Chen TC. Vitamin d deficiency: a worldwide problem with health consequences. *Am J Clin Nutr* (2008) 87:1080S–6S. doi: 10.1093/ajcn/87.4.1080S

37. Wacker M, Holick MF. Vitamin d–effects on skeletal and extraskeletal health and the need for supplementation. *Nutrients* (2013) 5:111–48. doi: 10.3390/nu5010111

38. Pilz S, Trummer C, Pandis M, Schwetz V, Aberer F, Gruebler M, et al. Vitamin d: current guidelines and future outlook. *Anticancer Res* (2018) 38:1145–51. doi: 10.21873/anticanres.12333

39. Schmitt EB, Nahas-Neto J, Bueloni-Dias F, Poloni PF, Orsatti CL. Vitamin d deficiency is associated with metabolic syndrome in postmenopausal women. *Maturitas* (2018) 107:97–102. doi: 10.1016/j.maturitas.2017.10.011

40. Lee SJ, Lee EY, Lee JH, Kim JE, Kim KJ, Rhee Y, et al. Associations of serum 25-hydroxyvitamin d with metabolic syndrome and its components in elderly men and women: the Korean urban rural elderly cohort study. *BMC geriatrics* (2019) 19:1–8. doi: 10.1186/s12877-019-1118-y7

41. Zhu W, Heil DP. Associations of vitamin d status with markers of metabolic health: A community-based study in shanghai, China. *Diabetes Metab Syndrome: Clin Res Rev* (2018) 12:727–32. doi: 10.1016/j.dsx.2018.04.010

42. Takiishi T, Gysemans C, Bouillon R, Mathieu C. Vitamin d and diabetes. *Rheum Dis Clinics* (2012) 38:179–206. doi: 10.1016/j.rdc.2012.03.015

43. Gagnon C, Daly RM, Carpentier A, Lu ZX, Shore-Lorenti C, Sikaris K, et al. Effects of combined calcium and vitamin d supplementation on insulin secretion, insulin sensitivity and β-cell function in multi-ethnic vitamin d-deficient adults at risk for type 2 diabetes: a pilot randomized, placebo-controlled trial. *PloS One* (2014) 9:e109607. doi: 10.1371/journal.pone.0109607

44. Gao Y, Wu X, Fu Q, Li Y, Yang T, Tang W. The relationship between serum 25-hydroxy vitamin d and insulin sensitivity and β-cell function in newly diagnosed type 2 diabetes. *J Diabetes Res* (2015) 2015:63-68. doi: 10.1155/2015/636891

45. Park S, Kim DS, Kang S. Vitamin d deficiency impairs glucose-stimulated insulin secretion and increases insulin resistance by reducing PPAR-γ expression in nonobese type 2 diabetic rats. *J Nutr Biochem* (2016) 27:257–65. doi: 10.1016/j.jnutbio.2015.09.013

46. Park JE, Pichiah PT, Cha Y-S. Vitamin d and metabolic diseases: growing roles of vitamin d. *J Obes Metab syndrome* (2018) 27:223. doi: 10.7570/jomes.2018.27.4.223

47. Von Hurst PR, Stonehouse W, Coad J. Vitamin d supplementation reduces insulin resistance in south Asian women living in new Zealand who are insulin resistant and vitamin d deficient - a randomised, placebo-controlled trial. *Br J Nutr* (2010) 103:549–55. doi: 10.1017/S0007114509992017

48. Hu Z, Jin'an Chen XS, Wang L, Wang A. Efficacy of vitamin d supplementation on glycemic control in type 2 diabetes patients: A meta-analysis of interventional studies. *Medicine* (2019) 98:149-170. doi: 10.1097/MD.0000000000014970

49. George P, Pearson E, Witham M. Effect of vitamin d supplementation on glycaemic control and insulin resistance: a systematic review and meta-analysis. *Diabetic Med* (2012) 29:e142–50. doi: 10.1111/j.1464-5491.2012.03672.x

50. Belenchia AM, Tosh AK, Hillman LS, Peterson CA. Correcting vitamin d insufficiency improves insulin sensitivity in obese adolescents: a randomized controlled trial. *Am J Clin Nutr* (2013) 97:774–81. doi: 10.3945/ajcn.112.050013

51. Poolsup N, Suksomboon N, Plordplong N. Effect of vitamin d supplementation on insulin resistance and glycaemic control in prediabetes: a systematic review and meta-analysis. *Diabetic Med* (2016) 33:290–9. doi: 10.1111/dme.12893

52. Sung C-C, Liao M-T, Lu K-C, Wu C-C. Role of vitamin d in insulin resistance. *J Biomed Biotechnol* (2012) 2012:63-72. doi: 10.1155/2012/634195

53. Pilz S, Kienreich K, Rutters F, De Jongh R, Van Ballegooijen AJ, Grübler M, et al. Role of vitamin d in the development of insulin resistance and type 2 diabetes. *Curr Diabetes Rep* (2013) 13:261–70. doi: 10.1007/s11892-012-0358-4

54. Mathieu C. Vitamin d and diabetes: where do we stand? *Diabetes Res Clin Pract* (2015) 108:201–9. doi: 10.1016/j.diabres.2015.01.036

55. Wang Y, Si S, Liu J, Wang Z, Jia H, Feng K, et al. The associations of serum lipids with vitamin d status. *PloS One* (2016) 11:e0165157. doi: 10.1371/journal.pone.0165157

56. Mahmoodi MR, Najafipour H. Associations between serum vitamin D3, atherogenic indices of plasma and cardiometabolic biomarkers among patients with diabetes in the KERCADR study. *BMC Endoc Disord* (2022) 22:1–12. doi: 10.1186/s12902-022-01043-1

57. Jiang X, Peng M, Chen S, Wu S, Zhang W. Vitamin d deficiency is associated with dyslipidemia: a cross-sectional study in 3788 subjects. *Curr Med Res Opin* (2019) 35:1059–63. doi: 10.1080/03007995.2018.1552849

58. Alquaiz AM, Kazi A, Youssef RM, Alshehri N, Alduraywish SA. Association between standardized vitamin 25 (OH) d and dyslipidemia: a community-based study in Riyadh, Saudi Arabia. *Environ Health Prev Med* (2020) 25:1–9. doi: 10.1186/s12199-019-0841-5

59. Karhapää P, Pihlajamäki J, Pörsti I, Kastarinen M, Mustonen J, Niemelä O, et al. Diverse associations of 25-hydroxyvitamin d and 1, 25-dihydroxy-vitamin d with dyslipidaemias. *J Internal Med* (2010) 268:604–10. doi: 10.1111/j.1365-2796.2010.02279.x

60. Skaaby T, Husemoen LLN, Pisinger C, Jørgensen T, Thuesen BH, Fenger M, et al. Vitamin d status and changes in cardiovascular risk factors: A prospective study of a general population. *Cardiology* (2012) 123:62–70. doi: 10.1159/000341277

61. Shen S, Lu Y, Qi H, Li F, Shen Z, Wu L, et al. Association between ideal cardiovascular health and the atherogenic index of plasma. *Medicine* (2016) 95:38-50. doi: 10.1097/MD.0000000000003866

62. Dobiasova M. AIP–atherogenic index of plasma as a significant predictor of cardiovascular risk: from research to practice. *Vnitrni lekarstvi* (2006) 52:64–71.

63. Wu T-T, Gao Y, Zheng Y-Y, Ma Y-T, Xie X. Atherogenic index of plasma (AIP): A novel predictive indicator for the coronary artery disease in postmenopausal women. *Lipids Health Dis* (2018) 17:1–7. doi: 10.1186/s12944-018-0828-z

64. Fernández-Macías JC, Ochoa-Martínez AC, Varela-Silva JA, Pérez-Maldonado IN. Atherogenic index of plasma: Novel predictive biomarker for cardiovascular illnesses. *Arch Med Res* (2019) 50:285–94. doi: 10.1016/j.arcmed.2019.08.009

65. Kassi E, Adamopoulos C, Basdra EK, Papavassiliou AG. Role of vitamin d in atherosclerosis. *Circulation* (2013) 128:2517–31. doi: 10.1161/CIRCULATIONAHA.113.002654

66. Shrivastava AK, Singh HV, Raizada A, Singh SK. C-reactive protein, inflammation and coronary heart disease. *Egyptian Heart J* (2015) 67:89–97. doi: 10.1016/j.ehj.2014.11.005

67. Chen S, Swier VJ, Boosani CS, Radwan MM, Agrawal DK. Vitamin d deficiency accelerates coronary artery disease progression in swine. *Arteriosclerosis thrombosis Vasc Biol* (2016) 36:1651–9. doi: 10.1161/ATVBAHA.116.307586

68. Sharma G, She Z-G, Valenta DT, Stallcup WB, Smith JW. Scavenger receptor-mediated targeting of macrophage foam cells in atherosclerotic plaque using oligonucleotide-functionalized nanoparticles. *Nano Life* (2010) 1:207–14. doi: 10.1142/S1793984410000183

69. Yin K, You Y, Swier V, Tang L, Radwan MM, Pandya AN, et al. Vitamin d protects against atherosclerosis *via* regulation of cholesterol efflux and macrophage polarization in hypercholesterolemic swine. *Arteriosclerosis thrombosis Vasc Biol* (2015) 35:2432–42. doi: 10.1161/ATVBAHA.115.306132

70. Koyama T, Shibakura M, Ohsawa M, Kamiyama R, Hirosawa S. Anticoagulant effects of 1α, 25-dihydroxyvitamin D3 on human myelogenous leukemia cells and monocytes. *Blood J Am Soc Hematol* (1998) 92:160–7.

71. Nakagawa K, Sasaki Y, Kato S, Kubodera N, Okano T. 22-Oxa-1α, 25-dihydroxyvitamin d 3 inhibits metastasis and angiogenesis in lung cancer. *Carcinogenesis* (2005) 26:1044–54. doi: 10.1093/carcin/bgi049

72. Oh J, Weng S, Felton SK, Bhandare S, Riek A, Butler B, et al. 1, 25 (OH) 2 vitamin d inhibits foam cell formation and suppresses macrophage cholesterol uptake in patients with type 2 diabetes mellitus. *Circulation* (2009) 120:687–98. doi: 10.1161/CIRCULATIONAHA.109.856070

73. Dikker O, Bekpinar S, Ozdemirler G, Uysal M, Vardar M, Atar S, et al. Evaluation of the relation between omentin-1 and vitamin d in postmenopausal women with or without osteoporosis. *Exp Clin Endocrinol Diabetes* (2018) 126:316–20. doi: 10.1055/s-0043-120110

74. Nazar S, Zehra S, Azhar A. Association of single nucleotide missence polymorphism Val109Asp of omentin-1 gene and coronary artery disease in Pakistani population: Multicenter study. *Pakistan J Med Sci* (2017) 33:1128-1133. doi: 10.12669/pjms.335.13110

75. Du Y, Ji Q, Cai L, Huang F, Lai Y, Liu Y, et al. Association between omentin-1 expression in human epicardial adipose tissue and coronary atherosclerosis. *Cardiovasc Diabetol* (2016) 15:1–9. doi: 10.1186/s12933-016-0406-5

76. Park SE, Lee NS, Park JW, Rhee E-J, Lee W-Y, Oh K-W, et al. Association of urinary RBP4 with insulin resistance, inflammation, and microalbuminuria. *Eur J Endocrinol* (2014) 171:443–9. doi: 10.1530/EJE-14-0247

77. Wang L, Song J, Wang C, Lin P, Liang K, Sun Y, et al. Circulating levels of betatrophin and irisin are not associated with pancreatic β-cell function in previously diagnosed type 2 diabetes mellitus patients. *J Diabetes Res* (2016) 2016:261-273. doi: 10.1155/2016/2616539

78. Jialal I, Adams-Huet B, Duong F, Smith G. Relationship between retinol-binding protein-4/adiponectin and leptin/adiponectin ratios with insulin resistance and inflammation. *Metab Syndrome Related Disord* (2014) 12:227–30. doi: 10.1089/met.2014.0013

frontiers | Frontiers in Endocrinology

TYPE Original Research
PUBLISHED 06 March 2023
DOI 10.3389/fendo.2023.1084327

OPEN ACCESS

EDITED BY
Prasoon Agarwal,
National Bioinformatics Infrastructure Sweden, Sweden

REVIEWED BY
SubbaRao V. Madhunapantula,
JSS Academy of Higher Education and Research, India
Houcemeddine Othman,
University of the Witwatersrand, South Africa
Shaimaa Hamza,
Kazan Federal University, Russia

*CORRESPONDENCE
Sangdun Choi
sangdunchoi@ajou.ac.kr

SPECIALTY SECTION
This article was submitted to Systems Endocrinology, a section of the journal Frontiers in Endocrinology

RECEIVED 30 October 2022
ACCEPTED 20 February 2023
PUBLISHED 06 March 2023

CITATION
Pirzada RH, Ahmad B, Qayyum N and Choi S (2023) Modeling structure–activity relationships with machine learning to identify GSK3-targeted small molecules as potential COVID-19 therapeutics. *Front. Endocrinol.* 14:1084327. doi: 10.3389/fendo.2023.1084327

Modeling structure–activity relationships with machine learning to identify GSK3-targeted small molecules as potential COVID-19 therapeutics

Rameez Hassan Pirzada[1,2], Bilal Ahmad[1], Naila Qayyum[1] and Sangdun Choi[1,2]*

[1]Department of Molecular Science and Technology, Ajou University, Suwon, Republic of Korea, [2]S&K Therapeutics, Ajou University Campus Plaza, Suwon, Republic of Korea

Coronaviruses induce severe upper respiratory tract infections, which can spread to the lungs. The nucleocapsid protein (N protein) plays an important role in genome replication, transcription, and virion assembly in SARS-CoV-2, the virus causing COVID-19, and in other coronaviruses. Glycogen synthase kinase 3 (GSK3) activation phosphorylates the viral N protein. To combat COVID-19 and future coronavirus outbreaks, interference with the dependence of N protein on GSK3 may be a viable strategy. Toward this end, this study aimed to construct robust machine learning models to identify GSK3 inhibitors from Food and Drug Administration–approved and investigational drug libraries using the quantitative structure–activity relationship approach. A non-redundant dataset consisting of 495 and 3070 compounds for GSK3α and GSK3β, respectively, was acquired from the ChEMBL database. Twelve sets of molecular descriptors were used to define these inhibitors, and machine learning algorithms were selected using the LazyPredict package. Histogram-based gradient boosting and light gradient boosting machine algorithms were used to develop predictive models that were evaluated based on the root mean square error and R-squared value. Finally, the top two drugs (selinexor and ruboxistaurin) were selected for molecular dynamics simulation based on the highest predicted activity (negative log of the half-maximal inhibitory concentration, pIC_{50} value) to further investigate the structural stability of the protein-ligand complexes. This artificial intelligence-based virtual high-throughput screening approach is an effective strategy for accelerating drug discovery and finding novel pharmacological targets while reducing the cost and time.

KEYWORDS

GSK3, machine learning, coronaviruses, QSAR, molecular descriptors

1 Introduction

Coronaviruses are a family of enveloped viruses known for their ability to infect humans, typically leading to respiratory illnesses (1, 2). To address the current epidemic of the novel coronavirus SARS-CoV-2 causing COVID-19, numerous approaches for virus identification and infection prevention and treatment are required. In this context, high-throughput screening has been performed to identify bioactive compounds that inhibit SARS-CoV-2 replication in tissue culture models (3–6). However, the mode of action and clinical efficacy of these candidates remain to be fully characterized, and additional targets need to be identified to further combat new and emerging SARS-CoV-2 variants. Among other structural and non-structural SARS-CoV-2 proteins, the nucleocapsid (N) protein is an essential RNA-binding protein that plays a crucial role in viral replication, transcription, and assembly (7–11). Inhibiting SARS-CoV-2 transcription will be a crucial objective, along with strengthening the immune response to the virus and reducing cytokine release syndrome linked to severe cases of COVID-19.

Glycogen synthase kinase 3 (GSK3) is a serine-threonine kinase signaling protein that plays a crucial role in a variety of biological processes, and its aberrant activity has been associated with diabetes, inflammation, and neurodegenerative and psychiatric disorders (12–15). GSK3 has two structurally identical isoforms (α and β), which are 97% similar in their catalytic domains but differs in their N and C (16–18). Both of these isoforms shows 98% of sequence identity (19). GSK3 (α and β) are required for the phosphorylation of SARS-CoV-2 N protein, and its inhibition blocks SARS-CoV-2–mediated infection in human lung epithelial tissues (16). In addition, knockout of GSK3α and GSK3β validates that it is vital for N phosphorylation (16). Inhibition of GSK3 can increase adaptive T cell and innate natural killer responses of $CD8^+$ T cells while also inhibiting SARS-CoV-2 viral replication (20). Moreover, GSK3 phosphorylates N proteins within the arginine-serine (RS) region of the JHM strain of mouse hepatitis virus (JHMV) and SARS-CoV, which caused the outbreak of severe acute respiratory syndrome in 2002–2004 (9, 10, 21, 22). Phosphorylation of the JHMV N protein is necessary for the recruitment of the ATP-dependent RNA helicase DDX1 for the transcription of long sub-genetic RNAs (10). The N protein from infectious bronchitis virus (IBV) and SARS-CoV interact directly with GSK3, and its knockdown was shown to disrupt the replication of IBV in the Vero cell line (23, 24). Moreover, GSK3 inactivators inhibit the coronavirus protease Mpro (or 3C-like protease), which cleaves SARS-CoV-2–encoded polyproteins (pp1a and pp1ab) required for viral replication and transcription. Additionally, GSK3β was also identified to control the autophagy pathway as it involves in the regulation of transcription factor EB (TFEB) nuclear expression mediated *via* mechanistic target of rapamycin complex 1 (mTORC1) dependent manner (25, 26). It also modulates TFEB through the signaling pathways of protein kinase C (PKC) and eukaryotic translation initiation factor 4A-3 (eIF4A3) (27, 28). Furthermore, GSK3β-induced phosphorylation of TFEB leads to its cytoplasmic retention and contributes to the blockage of the lysosomal-mediated autophagy pathway (29, 30). Autophagy plays a critical role in the degradation of dysfunctional cytoplasmic organelles and infectious pathogens, whereas defective autophagy has been observed in SARS-CoV-2 pathogenesis (31, 32). An interplay between viral infection and autophagy directs towards the development of an effective therapeutic approach for COVID-19 (31, 33). Therefore, it is reasonable to propose that inhibition of GSK3 using small-molecule inhibitors would effectively block SARS-CoV-2 replication as found for SARS-CoV-1.

Quantitative structure–activity relationship (QSAR) approaches are increasingly being applied to aid in the preclinical development of small-molecule drugs. QSAR models help predict the physicochemical properties, biological activity, toxicity, chemical reactivity, and metabolism of chemical compounds (34–36). The main goal of QSAR analysis is to link a set of predictor variables (X) to the response variable (Y). Techniques for linking X and Y and molecular descriptors have received substantial research attention. In this context, a key strategy in drug discovery is the development of machine learning (ML) techniques to estimate drug-target interactions. QSAR approaches employ a variety of linear and non-linear ML algorithms to produce predictive models for ligand binding to a biological receptor. The term "QSAR" refers to regression models that establish quantitative relationships between molecules' chemical structures and their physical, chemical, or biological characteristics. ML techniques such as gradient boosting, support vector machines, partial least squares, artificial neural networks, or linear regression use a set of molecular descriptors as input data to predict chemical features.

We attempted to develop ML-based QSAR models that could identify GSK3 inhibitors using the bioactivity data available in the ChEMBL and PubChem databases. ML models were developed using two algorithms, histogram-based gradient boosting (HGBM), and light gradient boost machine (LGBM), to prospectively identify GSK3 inhibitors from the Food and Drug Administration (FDA)–approved drug library. The rationale behind selecting GSK3 as a drug target is that most of the anti-viral therapies are primarily designed to target the viral structure which however is frequently associated with drawbacks such as drug resistance as a consequence of viral mutation (37, 38). Here drug-oriented machine learning-based repurposing approach was adopted based on the physicochemical and pharmacological properties of both active and inactive GSK3 inhibitors to build a model that can identify already approved drugs against the selected target (GSK3). Furthermore, this drug discovery strategy is highly efficient, saves time, and cost and proves to be a prospective approach towards finding already approved drugs against SARS-CoV-2 (39). Finally, molecular dynamics (MD) simulation was performed to further investigate the structural stability of the protein-ligand complexes.

2 Materials and methods

2.1 Data compilation and curation

The GSK3α (Target ID: CHEMBL2850) and GSK3β (Target ID: CHEMBL262) datasets used in this study were extracted from the ChEMBL database to generate ML models (40). These datasets are composed of a diverse set of molecules that have been

10.3389/fendo.2023.1084327

experimentally validated for their inhibitory activity against GSK3. A schematic illustration of the ML-based QSAR and structural dynamics workflow used in this study is shown in Figure 1. Initially, the total number of collected compounds with various bioactivity units, including IC_{50}, Ki, EC_{50}, and KD, were collected for GSK3α and GSK3β, consisting of 587 and 3637 molecules, respectively. The dataset was cleaned and preprocessed by applying the following filters: first, compounds with undefined activity were discarded; second, compounds containing salt or mixtures, along with overlapping compounds, were removed. Furthermore, in this study, the subset of bioactivity data obtained with IC_{50} as a unit was investigated for both GSK3α and GSK3β, consisting of 495 and 3070 unique bioactive compounds, respectively. As the objective of this study was to develop classification models of biologically active compounds, the activity dataset was divided into active and inactive compounds with IC_{50} thresholds of<1 and >10 μM, respectively,

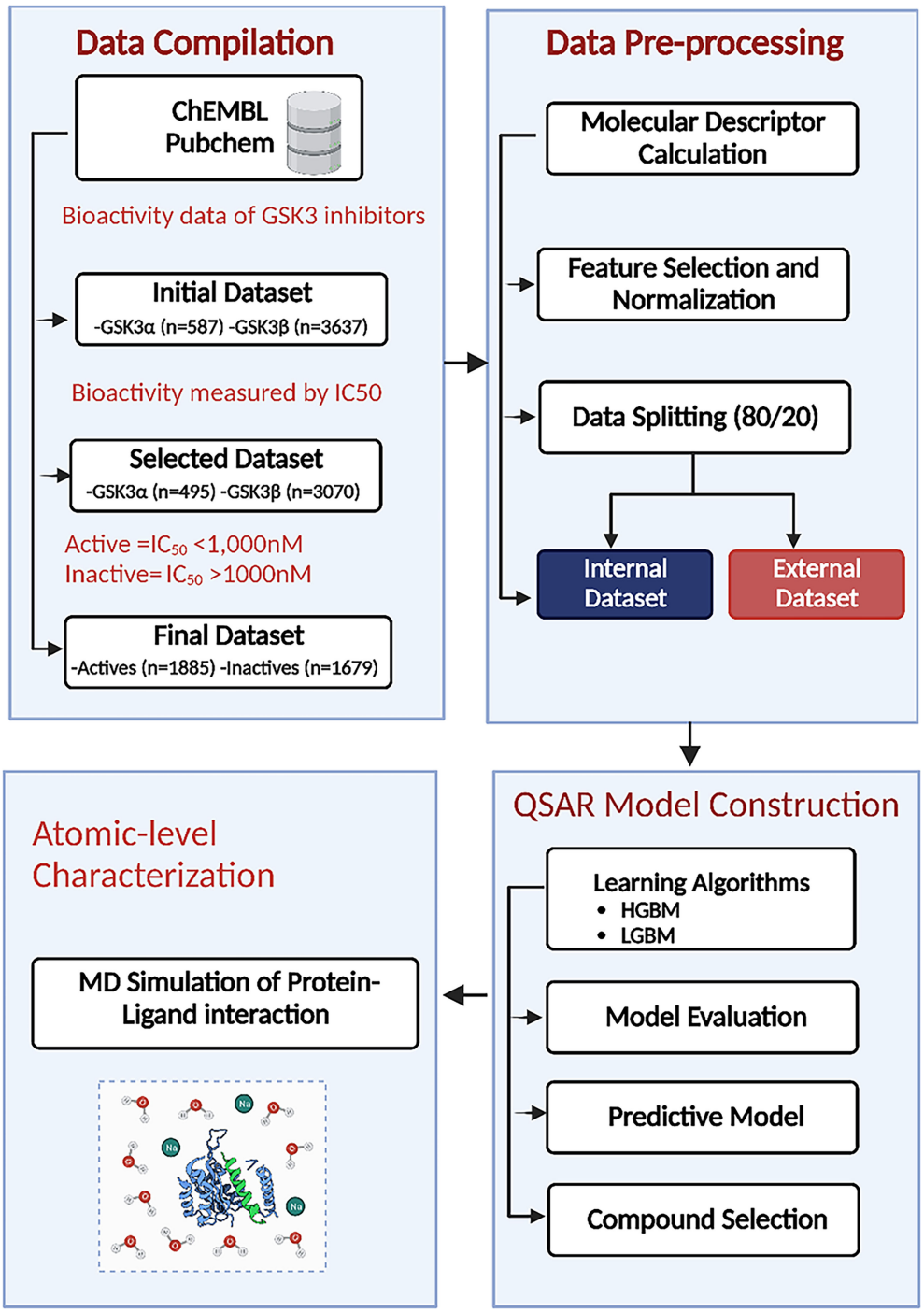


FIGURE 1
Machine learning (ML)–based quantitative structure–activity relationship (QSAR) and structural dynamics analysis workflow. HGBM, histogram-based gradient boosting ML model; LGBM, light gradient boosting ML model.

whereas compounds with intermediate activity levels were not considered. The selected compounds were filtered based on Lipinski's rule of five (RO5). Finally, after pre-and post-processing, two sets of non-redundant and curated compounds for GSK3 (α, β) were used for further investigation: 1885 active compounds and 1679 inactive compounds. All of the collected compounds were randomly segregated into a training set to generate the QSAR classification model and a test set to validate the model quality with a ratio of 80:20.

2.2 Molecular descriptors

A molecular descriptor is a mathematical (quantitative/or qualitative) representation of a molecule that is encoded with various sources of chemical information, which is converted and coded to deal with the biological, chemical, and pharmacological features of different small molecules that are used for model construction (41, 42). To develop robust non-linear binary QSAR classification models with better performance, various descriptors such as electronic, topological, geometrical, and spatial descriptors were computed for each molecule in the dataset. We used PaDEL-Descriptor software for molecular descriptor calculation based on a Python code to calculate 12 sets of molecular descriptors, as shown in Table 1 (43). The descriptors belong to nine different types: CDK fingerprint, CDK extended, CDK graph only, Klekota-Roth, AtomPairs 2D, MACCS, E-state, PubChem, and Substructure. These descriptor types can be further divided into two versions: binary and count versions. In this context, the descriptors Klekota-Roth, AtomPairs 2D, and Substructure belong to both versions and provide a detailed description of the substructural components of the studied molecules. The remaining descriptors (n = 9) belong to the binary version. We also computed Lipinski's RO5 molecular descriptors to be used as classification parameters for the identification of drug-like molecules.

2.3 Data filtering

Molecular fingerprints with redundant or constant variables were discarded to remove any inherent biases that could negatively impact the resulting model. Not all molecular descriptors are required to represent the features of inhibitors and non-inhibitors. Moreover, the model learns the biases in the data and continues to amplify them, which could lead to overfitting. A selection criterion is required to discard irrelevant descriptors that can measure the relevance of a specific descriptor to the output of any classifier (44). In this context, the VarianceThreshold class from Scikit-learn was implemented to remove the low-variance features with a threshold value higher than 0.1, and the remaining features were used for further analysis.

2.4 Data splitting and test selection

Following data filtering, the GSK3 (α, β) datasets were split using the Kennard–Stone algorithm, which separated the entire dataset into internal and external sets with a ratio of 80% and 20%, respectively. The internal datasets were used for training the ML model, and its capability to extrapolate to unknown molecules was simulated by analysis against the external dataset. Finally, the training set was used to estimate the performance of the model using a five-fold cross-validation scheme. The correlation plots for the experimental versus predicted pIC_{50} values for GSK3 inhibition in the training and test sets are shown in Figure 2.

TABLE 1 Calculated using the PaDEL-descriptor, 12 sets of fingerprint descriptors.

Fingerprint	Abbreviation	Number (bits)	Fingerprint Pattern Type	Description
CDK	FP	1024	Hash fingerprints	Fingerprint of length 1024
CDK extended	ExtendedFP	1024	Hash fingerprints	Adds fingerprint information about ring features.
CDK graph only	GraphOnlyFP	1024	Hash fingerprints	A distinct approach that considers connectivity and ignores bond order
MACCS	MaccsFP	166	Structural features	Chemical characteristics represented in binary using MACCS keys
Substructure	Substructure	307	Structural features	SMARTS patterns for functional groups are present.
Substructure count	nSubstructure	307	Structural features count	SMARTS patterns counted for functional groups
2D atom pairs Count	nAP2DC	780	Structural features count	Atom pair count at different topological distances
2D atom pairs	AP2D	780	Structural features	Atom pair presence at different topological distances
PubChem	PubChemFP	881	Structural features	Binary representation of the PubChem-defined substructures
Klekota–Roth	KRFP	4860	Structural features	Existence of chemical substructures
Klekota–Roth Count	nKRFP	4860	Structural features count	Substructure count for chemicals
E-state	EstateFP	79	Structural features	Types of atoms with respect to electrotopology

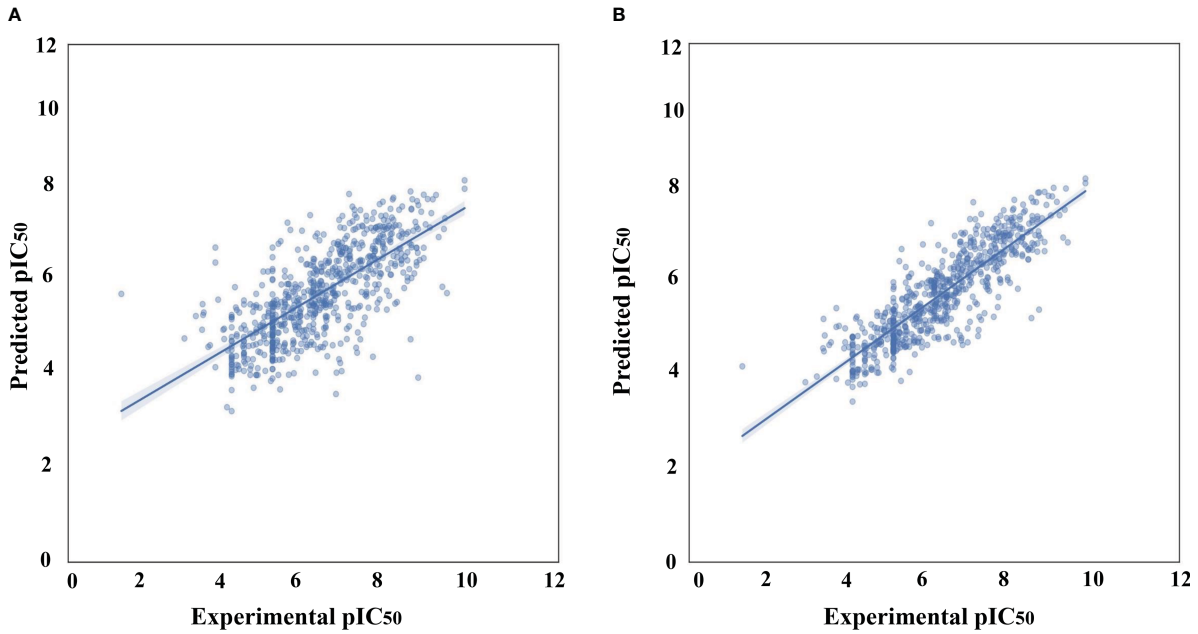


FIGURE 2
Correlation plots of experimental vs. predicted pIC_{50} values for GSK3 inhibition to the training and test set. **(A)** Histogram-based gradient boosting (HGBM) model. **(B)** Light gradient boosted machine (LGBM) model.

Additionally, the effectiveness of the ML model was evaluated using 20 compounds that were identified as inhibitors of this target in numerous previous studies (45–47). The model performance measures and the activity threshold for this external dataset were compared with the experimental IC_{50} values, as shown in the Supplementary Tables S1, S2.

2.5 ML-based QSAR classification

The QSAR classification model can represent molecular descriptors as a relationship between the dependent variable (IC_{50}) and independent variables (molecular descriptors), each demonstrating the category of the corresponding sample (GSK3 inhibitory activity). In structure–activity relationships, the association between the corresponding datasets is complex and non-linear; thus, a QSAR modeling–based approach was used because it has previously shown outstanding performance in this regard (48). In this context, ML algorithms can cluster instances or observations present in data into classes. A variety of ML algorithms have been employed to construct QSAR classification models from dataset activity labels and molecular descriptors. For example, support vector machines, naïve Bayes classifiers, neural networks, rule-based classifiers, and decision trees are various ML-based techniques used to elucidate the classification problem. In this study, the LazyPredict package was employed using a Python script for model selection, which generates a variety of ML algorithms and authenticates the best-performing algorithm, as shown in Supplementary Table S3 (49). The top models were selected based on the R-squared and root mean squared error (RMSE) values to train our regression model to precisely predict the activity of GSK3 inhibitors. These ML algorithms were implemented using Python software. Subsequently, to determine the optimal values, hyperparameter tuning of the selected models was performed using GridSearchCV implemented in Scikit-learn. The list of the best parameters selected for hyperparameter tuning is presented in Table 2.

2.6 Statistical assessment for model validation/performance

Model validation is a crucial step to ensure that a fitted model can accurately predict responses to unknown data. We used two statistical parameters, the Pearson correlation coefficient (r) and RMSE, to assess the performance of the QSAR models. The Pearson correlation coefficient is a common statistic used to describe the strength of the relationship between two variables of interest, which ranges from −1 to +1, with negative values denoting a negative correlation between two variables and positive values indicating a positive correlation. The relative error of the prediction model is frequently examined using the RMSE.

The Y-scrambling test, external validation, and 10-fold cross-validation were used to confirm the predictive capacity of the QSAR model. The 10-fold cross-validation method does not use the entire dataset to create a predictive model. As an alternative, it separates the data into training and testing datasets, enabling the model to be tested on the testing dataset using the training dataset as a basis. By repeating the 10-fold validation, the average accuracy was used to examine the performance of the prediction model. To further assess the performance and prediction accuracy of the model on external/benchmarking datasets, mean absolute percentage error (MAPE)

was evaluated, a smaller value indicates better model performance.

$$MAPE = \frac{1}{N}\sum_{K=1}^{N}\frac{|X_k - \hat{X}_k|}{X_k} \times 100\,\%$$

Xk is the actual value, $\hat{X}k$ is the predicted value of the model, $|X_k - \hat{X}_k|$ represents the absolute error, and N is the number of incidence data points.

2.7 Docking and MD simulation

The selected FDA-approved and natural product compounds were selected based on the predicted IC_{50} values produced by our QSAR model and docked to the active-site of GSK3 using a protocol described in our previous report (50). From the top 10 docked conformations the best pose was selected based on the MOE docking score (S), protein-ligand interaction fingerprint (PLIF) calculation, including hydrogen and ionic interactions. These selected best poses of selinexor, ruboxistaurin alongwith GSK3 apo and PF-04802367 (PDB ID: 5K5N) were subjected to MD simulation (200 ns) using GROMACS. Furthermore, the dynamic behavior and stability of each system was validated. The detailed methodology adopted here can be found in our previous study (50).

2.8 Binding free energy calculations

Molecular mechanics Poisson–Boltzmann surface area (MMPBSA) method was used to measure the binding free energy between GSK3 and ligand complexes. The Poisson-Boltzmann equation was used to determine the effects of the solvent's polar and nonpolar components on the free energy.

$$\Delta G_{bind=} G_{complex} - (G_{receptor} + G_{ligand})$$

$$\Delta G_{bind=} \Delta E_{MM} + \Delta\Delta G_{sol} - T\Delta S$$

In the above equation, ΔGbind represents binding free energy, ΔEMM is the intermolecular energy difference, ΔΔGsol shows the difference in solvation energy, whereas T and ΔS stands for absolute temperature and change in entropy. The study was performed using the gmx_MMPBSA (51). Frames were extracted throughout the trajectory with an interval of 30-frames. The detailed protocol is described in our previous study (52).

TABLE 2 The best parameters of machine learning algorithms following parameter adjustment using GridSearchCV.

Machine learning methods	Tuning parameter	Model performance R^2
HGBM	n_estimators: 800 random_state: 100 learning_rate: 0.1 subsample: 1.0	0.72
LGBM	n_estimators: 100 max_depth: 9 learning_rate: 0.1 gamma: 0.1	0.70

3 Results and discussion

The computational workflow for elucidating the underlying basis of the bioactivity of GSK3 is summarized in Figure 1. To gain a deeper understanding of the dataset, a standard chemical space analysis was performed on the investigated compounds. The preprocessed dataset was utilized to create predictive classification and regression models using the HGBM and LGBM models after rigorous data curation. Subsequently, hyperparameter optimization was performed to determine the optimal parameter configuration of the model. The best-performing model was used to evaluate the predictive capability after training the selected algorithms to gain biological insight. Finally, MD simulations were performed on selected FDA-approved drugs based on the pIC_{50} value to further assess the structural dynamics and stability of the protein–drug complexes.

3.1 Exploratory chemical space analysis of GSK3 inhibitors

A set of 3,565 compounds tested against GSK3 (α, β) were extracted from public databases. This dataset included seven bioassay formats characterized using the BioAssays Ontology number (53). However, 95% of the dataset was linked to the same bioassay (BAO_0000357) connected to the single-protein affinity format, which could infer the homogeneity of the dataset.

Chemical space analysis of GSK3 inhibitors was explored using Lipinski's RO5 descriptors to gain an understanding of structure–activity relationships. Chemical space analysis can provide considerable understanding of the general characteristics of compounds that define their inhibitory activity. Exploratory data analysis was performed using RO5 descriptors, including the number of hydrogen bond acceptors (nHBAcc), number of hydrogen bond donors (nHBDon), molecular weight (MW), and octanol/water partition coefficient (LogP). The MW of a chemical compound is often used to compute its size, as it facilitates the analysis and prediction of the appropriate size of a drug, which is critical for its transport across a lipid membrane (54). Molecular hydrophobicity (lipophilicity) is usually computed as LogP, which is an important estimator of chemical membrane penetration and permeability (55). Moreover, nHBDon and nHBAcc were computed to estimate the hydrogen bond-forming capacity of a chemical compound.

Initially, the analysis was carried out by visualizing the distribution of active and inactive compounds as determined by the scatter plot of MW vs. LogP, followed by a comparative analysis of active and inactive compounds as a function of Lipinski's RO5 descriptors (56).

Figure 3 shows the chemical space distribution of the training set in a scatter plot of the MW versus the logarithm of the LogP. The MW and LogP of the active and inactive compounds showed almost

10.3389/fendo.2023.1084327

identical distributions, with the majority of the compounds having an MW falling between approximately 200 and 600 Da and a LogP falling between 0 and 7. Most oral drugs are more likely to have optimal physicochemical and absorption, distribution, metabolism, and excretion properties between 1 and 4 (57, 58), which is also evident in Figure 3. Furthermore, Figure 3 shows that MW and ALogP cannot be used to discriminate between active and inactive compounds because of their shared chemical space.

Figure 4A illustrates the total distribution of the compounds. As can be inferred from the median values in the box plot, active compounds tended to have a slightly higher MW of approximately 380 Da than that inactive compounds, which was approximately 350 Da. Similarly, inactive compounds had LogP values of approximately 3.8, which was higher than that of active compounds approximately 3.5 (Figure 4B). The distribution of nHBAcc, as deduced from the median, shows that the active compounds have a higher number of hydrogen bond acceptors (n ≈ 7) than inactive compounds (n ≈ 5) (Figure 4C). However, no significant differences were observed between the active and inactive compounds for nHBDon (Figure 4D). Therefore, it is difficult to predict the activity of inhibitors using simple molecular descriptors.

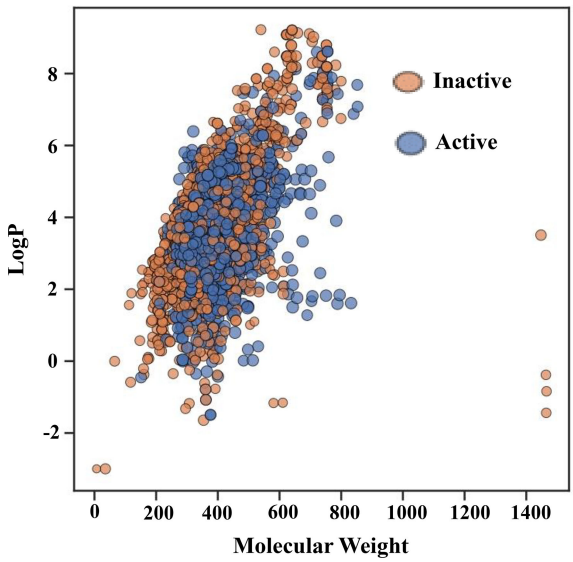


FIGURE 3
Chemical space of the training set. The molecular weight (MW) on the X-axis and the logarithm of the octanol/water partition coefficient (LogP) on the Y-axis serve as the parameters for the chemical space. Red and green spots, respectively, represent active and inactive substances.

3.2 Model construction for the prediction of GSK3 kinase inhibitors using ensemble boosting

After the molecular descriptor calculation (Table 1), the LazyPredict package was used to acquire robust ML models, as described in Section 2.5. The best-performing models, HGBM and LGBM, with an R-squared value of 0.53 and 0.52 were used for model construction to better target the GSK3 kinase protein.The performance of the model was evaluated using the R-squared and RMSE metrics, as shown in Supplementary Table S3. Boosting algorithms (HGBM and LGBM) are a type of ensemble learning technique that gradually add tree models to fix the prediction error that already exists in the ensemble (59). To evaluate the performance of our ML models, an external test set using 20 known GSK3 inhibitors was used (45–47, 60). Because these external test set compounds were not considered when creating the models, the resulting performance showed the ability of the ML models to precisely predict the inhibitory activity of already known GSK-3 kinase inhibitors, as shown in Supplementary Tables S1 and S2. The performance of the predictive ML model can also be seen in the correlation plot between predicted values vs. experimental values, as shown in Figure 5. Additionally, the overlapping GSK3 α, and β compounds in the benchmarking dataset of already known inhibitors are shown on Figure 6. The model evaluation metrics, MAPE was used to validate the prediction accuracy in benchmarking datasets (Figure 7). Both the HGBM and LGBM showed relatively good mean absolute percentage error scores of 19.1% and 22.6%, respectively, which represents the percentage difference between the predicted and experimental values. In addition to it, we compared the predictive performance of our QSAR model with another dataset (Supplementary Table S4). The data shows that compounds with higher activity in our model was comparable to the reported model (60).The results in Table 2 indicate that both models exhibited good overall prediction accuracy; however, HGBM was the more accurate model.

This QSAR-based model was then used to predict the inhibitory activity of FDA-approved drugs against GSK3 kinase. Each compound was assigned a pIC_{50} value; higher values indicate that the drug is effective at lower concentrations against GSK3 kinase and therefore will show lower systemic toxicity (61).

The top 10 FDA-approved drugs based on pIC_{50} values from both the HGBM and LGBM models are presented in Tables 3, 4. The drugs with the best-predicted pIC_{50} values produced by our model were selinexor and ruboxistaurin hydrochloride.

Selinexor is an FDA-approved drug for the treatment of multiple myeloma that binds to and inhibits exportin-1 (XPO) and is being evaluated against SARS-CoV-2 in a phase-2 clinical trial (NCT04349098) (62). XPO-1 protein plays an important role in the export of RNA transcripts and nuclear proteins having leucine-rich nuclear export signals (NES) (47, 60–62). However, blocking XPO-1 with its selective inhibitors causes the nuclear accumulation of transcription factor EB (TFEB) and results in autophagy enhancement in human cells and model organisms (33). Because of this, it has been demonstrated that selinexor inhibits the spread of SARS-CoV-2 by preventing the movement of nuclear proteins into the cytoplasm (63, 64). Similar phenomena have been observed when GSK3 is inhibited; this causes translocation of TFEB into the nucleus, where it controls the transcription of around 400 genes involved in autophagy, which eliminates the invading viruses like SARS-CoV-2 (65–67). The current study indicates additional, mechanisms through which selinexor could prevent SARS-CoV-2 replication, such as by preventing the phosphorylation of the virally encoded N protein. In addition to the drug's effect on autophagy (16, 20).

10.3389/fendo.2023.1084327

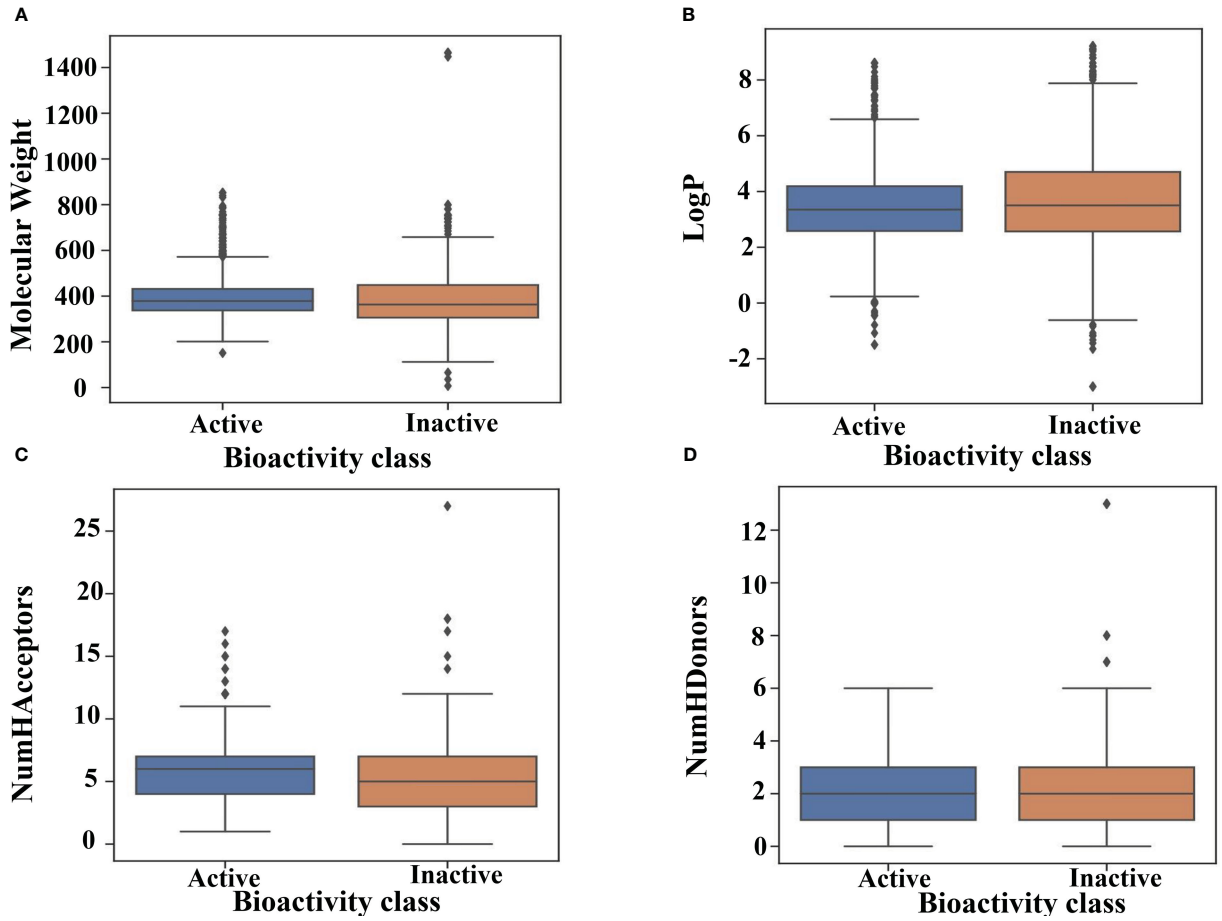


FIGURE 4
Drug-likeness evaluation. Box plot of GSK3 inhibitors using Lipinski's rule of five descriptors. **(A)** molecular weight **(B)** logP **(C)** hydrogen bond acceptors and **(D)** hydrogen bond donors.

Ruboxistaurin is an investigational drug that targets protein kinase C beta for the treatment of diabetic retinopathy. Protein kinase C (PKC) is a serine/threonine protein kinase that has been identified to modulate autophagy (68). Autophagy is an innate immune response to kill and degrade invading viruses (69). In this context, one such mechanism is the activation of PKC-α/δ induced by GSK3β inhibition which leads to the phosphorylation repression of TFEB and its nuclear localization and activation of autophagy pathways (30, 70). The nuclear localization of TFEB induced by PKC-α/δ occurs *via* GSK3β in mTORC1-independent manner (30). Additionally, an orally active PKC inhibitor ruboxistaurin proves to be active against SARS-CoV-2 as it inhibits NETosis, and has passed phase 3 for other indications (71). Previous studies have also shown that ruboxistaurin was active against GSK3α and GSK3β with IC_{50} values of 695.9 nM and 97.3 nM, respectively (60). This data indicates that ruboxistaurin inhibits both PKC and GSK3 which supports our QSAR model prediction.

In conclusion, the top predicted medications (selinexor, ruboxistaurin) by our model also correlate with the data that revealed their effectiveness against SARS-CoV-2 as shown above, and GSK3 inhibition seems to play a significant part in the activity of these drugs. Furthermore, the drugs with the highest predicted IC50 including selinexor and ruboxistaurin indicate that the molecular fingerprints of these compounds exist in the compounds used to develop the training set of GSK3 inhibitors. In addition, we validate the performance of our ML model from a structural viewpoint, these two drugs were subjected to MD simulation.

3.3 Atomic-level characterization and binding free energy calculations

Characterization of the protein-ligand complex is essential for predicting selective GSK3β inhibitors. The top two FDA-approved drugs (selinexor and ruboxistaurin) were selected for molecular dynamics (MD) simulation based on the highest predicted activity [according to the negative log of the half-maximal inhibitory concentration (pIC_{50}) value] to further investigate the structural stability of the protein-ligand complexes. The three-dimensional structure (PDB ID: 5K5N) of GSK3β was retrieved from the Protein DataBank and used to generate multiple docking poses to select the best conformer for the MD simulation.

All GSK3-apo, PF-04802367 and ligand-bound complexes were docked and subjected to MD simulations in an aqueous

10.3389/fendo.2023.1084327

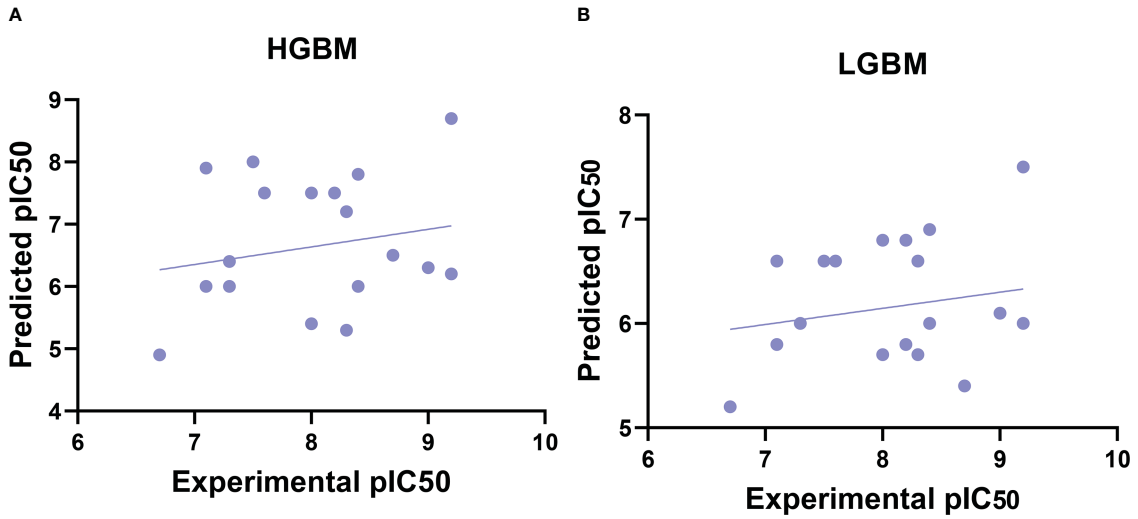


FIGURE 5
Correlation plots of experimental vs. predicted pIC_{50} values for GSK3 inhibition to the benchmarking dataset. **(A)** Histogram-based gradient boosting (HGBM) model. **(B)** Light gradient boosted machine (LGBM) model.

environment for 200 nano second (ns) to discern conformational variations and dynamic stability. The dynamic nature of the GSK3β protein and the test drugs (selinexor, ruboxistaurin, and PF-04802367) was explored separately using MD simulation trajectories. To explore the average displacement of the atoms, the root mean squared deviation (RMSD) of the complexes was evaluated and contrasted with the RMSD of the GSK3-apo and PF-04802367 (control) bound structures, as shown in Figure 8A. The GSK3β–ruboxistaurin complex showed a stable trajectory with slight variations throughout the simulation, as the RMSD value ranged from 0 to 2.4 Å. However, from 100 ns to 150 ns, the RMSD showed an incremental deviation from 2.5 to 2.7 Å before attaining the final trajectory of 2.4 Å. In the GSK3–selinexor complex, a similar RMSD trajectory profile was observed from 0 to 2.1, and from 50 to 150 ns, with an incremental deviation (2 to 2.4 Å) before reaching 2.1. The RMSD plots for GSK3–PF-04802367 (control drug) showed a steady incremental deviation with the trajectory originating from 0 to 2.8 Å, along with some acceptable variations during different time intervals. The RMSD profile of the apo structure showed a pattern similar to that of GSK3β in complexes with PF-04802367 and ruboxistaurin. However, acceptable deviations were observed in complexes with selinexor compared with those of the apo-form. These findings suggest that after a few conformational changes, the protein-ligand complex achieved a stable conformation during the simulations.

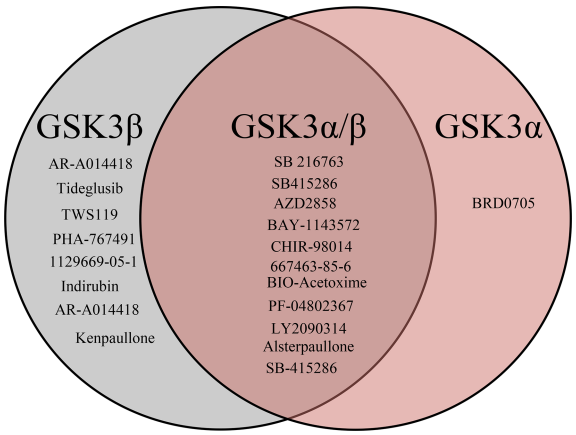


FIGURE 6
Venn diagrams visualizing the overlap of known GSK3 (α, β) inhibitors retrieved from the ChEMBL database. In total benchmarking dataset has 20 compounds out of which 8 are specific to GSK3β, 11 are to both GSK3 (α,/β) and 1 is to GSKα specific.

Additionally, to analyze the protein flexibility in the ligand-bound complexes, root mean square fluctuation (RMSF) values were calculated, as shown in Figure 8B. The RMSF of the residues measures the residue-level structural fluctuation of a specific segment of the protein that deviates from its mean structure, which often occurs upon ligand binding. The variations observed for each residue represent the degree of flexibility they attained. The ATP-binding site of GSK3β is present at the interface of the N and C termini and consists of Pro136, Leu132, Asp133, Tyr134, and Val135; however, the hydrophobic side chain of Arg141 forms another segment of the pocket (72). In the case of the GSK3β–apo structure, no significant fluctuations occurred in the binding site residues such as Pro136 (0.733 Å), Asp133 (0.633 Å), Tyr134 (0.637 Å), Leu132 (0.806 Å), and Val135 (0.87 Å). In the GSK3–PF04802367 complex, the RMSF values of residues Pro136 (0.614 Å), Leu132 (0.681 Å), Tyr134 (0.584 Å), Val135 (0.547 Å), and Arg141 (0.77 Å) were lower than those found for the apo structure. Similarly, in the GSK3β–selinexor complex, all active-site residues fluctuated less than in the apo structure, particularly residues Pro136 (0.64 Å), Leu132 (0.715 Å), Asp133 (0.593 Å), Tyr134 (0.589 Å), Val135 (0.698 Å), and Arg141 (0.797 Å). The fluctuation patterns of residues Asp133 (0.609 Å) and Tyr134 (0.622 Å) in the GSK3–ruboxistaurin complex were similar to those of the apo structure; however, slightly higher fluctuations were observed for

10.3389/fendo.2023.1084327

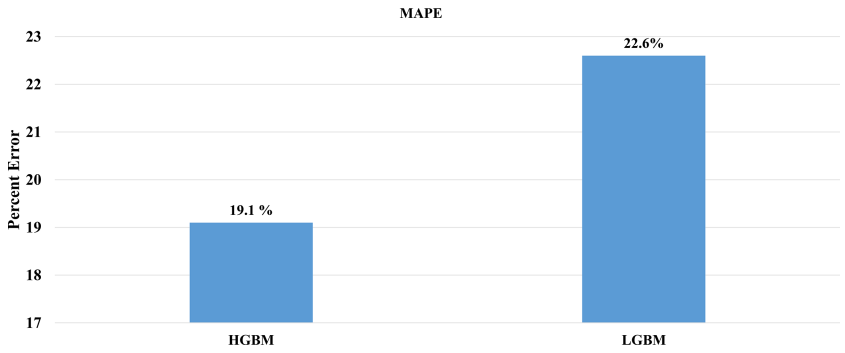


FIGURE 7
Comparison of mean absolute percentage error of the external test set. **(A)** HGBM and **(B)** LGBM in the context of prediction accuracy.

TABLE 3 GSK3 inhibition activity prediction by HGBM.

	Investigational and FDA Drugs	PubChem ID	pIC_{50}
1	Selinexor	71481097	9.4125746
2	Raltegravir potassium	23668479	9.175947616
3	Dasabuvir	56640146	9.001651838
4	Kuvan (sapropterin)	135409471	8.848548463
5	Deferiprone	2972	8.171706964
6	Propylthiouracil	657298	8.399683851
7	Trelagliptin	15983988	8.202770399
8	Urapidil	5639	7.831437785
9	Ruboxistaurin	9870785	7.419813853
10	Methylcobalamin	10898559	6.098851787

pIC_{50} is the negative log of the IC_{50}.
Histogram-based gradient boosting machine learning (HGBM).

TABLE 4 GSK3 inhibition activity prediction by LGBM.

	Investigational and FDA Drugs	PubChem ID	pIC_{50}
1	Ruboxistaurin	9870785	7.448867965
2	Methylcobalamin	10898559	7.229432941
3	Cefpirome	5479539	7.050002608
4	Allopurinol	135401907	7.030920707
5	Simeprevir	24873435	7.007010834
6	Neratinib	9915743	7.000923602
7	Selinexor	71481097	6.96201482
8	Lafutidine	5282136	6.882663164
9	Enoxacin	3229	6.831823482
10	Cefodizime	5361871	6.828905555

pIC_{50} is the negative log of the IC_{50}.
Light gradient boosting machine learning algorithm (LGBM).

residues Pro136 (0.938 Å) and Arg141 (0.959 Å). The GSK3–ligand interaction reduced the overall fluctuations of the protein; the RMSF value of the selected FDA drugs (selinexor and ruboxistaurin) was similar to that of the control drug PF-04802367.

To further characterize the compactness of the protein in the binding of the ligand at the ATP site of GSK3, the radius of gyration (Rg) was determined, as shown in Figure 8C. The impact of ligand binding on the Rg of the GSK3β protein was calculated and compared with that of the apo-GSK3β protein structure. The Rg of the apo and ligand-bound complexes remained between 21.5 and 21.7 Å, which indicated their compactness and sustained stability.

To analyze the formation of hydrogen bonds throughout the MD simulation, the gmx hbond program from the GROMACS package was used. As shown in Figure 8D, the average hydrogen bond for the control (PF-04802367) was about 1.5; however, an increase in the number of hydrogen bonds from 1.5 to 3.9 was observed. On the other hand, the average hydrogen bond for selinexor and ruboxistaurin was about 1.3 and 1.6. However, during the early steps (0 to 66 ns) of MD simulation, a low number of hydrogen bonds was observed in selinexor, which was about 0.6. Overall, the hydrogen bond formation remains intact throughout the simulations, indicating that the ligands were present in the binding pocket throughout the process.

The MM-PBSA technique was used to calculate binding free energy to measure the strength of receptor-ligand binding. The change in binding free energy upon binding of ruboxistaurin, selinexor, and PF-04802367 (control) with GSK3 receptors is shown in Figure 9. Here, GSK3-drugs (ruboxistaurin, selinexor) displayed acceptable binding free energy, which is comparable with a positive control (PF-04802367). The cumulative binding energy (Gbinding) of selinexor and ruboxistaurin is -8509.006 56.335 kcal/mol and -8491.473 57.039 kcal/mol, whereas PF-04802367 is -8681.320 58.8627 kcal/mol. However, the total binding energy of PF-04802367 is slightly higher but comparable with that of selinexor and ruboxistaurin, which clearly reflects the robustness of our ML model in predicting the drug candidates that could bind tightly to GSK3.

4 Conclusion

The COVID-19 pandemic caused by the new coronavirus, SARS-CoV-2, poses a serious threat to the global health system. We employed ML-based predictive modeling to identify FDA-approved and clinical candidate drugs inhibiting GSK3, as this kinase plays a critical role in the phosphorylation of the SARS-CoV-2 N protein that is required for viral replication (16, 20). Furthermore, among the FDA-approved compound libraries, leads with good pIC_{50} values were subjected to MD simulations to investigate protein–drug interactions in a dynamic environment.

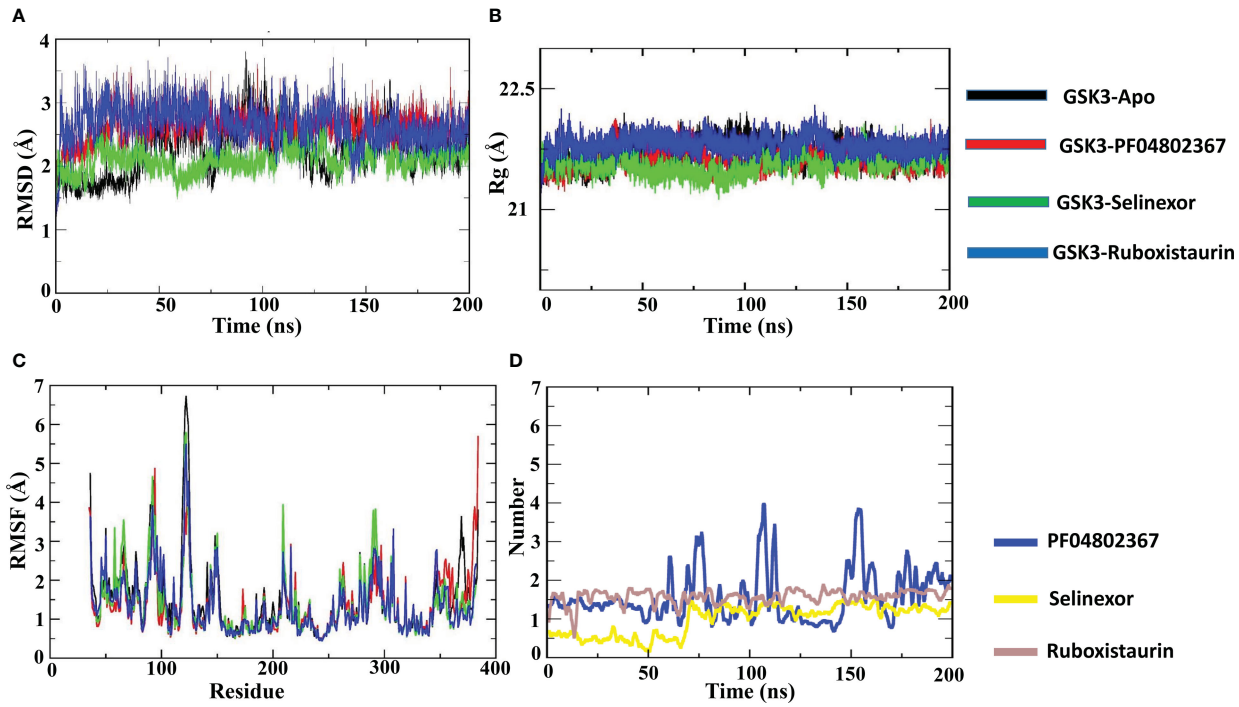


FIGURE 8
Atomic level characterization. Molecular docking simulation results on apo-GSK3β and complexes of GSK3β with PF04802367, selinexor, and ruboxistaurin. **(A)** Root mean square deviation (RMSD) of the apo-form of GSK3β and the complexes; **(B)** root mean square fluctuations (RMSFs) of the apo-form of GSK3β and the complexes; **(C)** the radius of gyration (Rg) of the apo-form of GSK3β and the complexes; **(D)** Number of hydrogen bonds in the three complexes; PF-04802367, selinexor and ruboxistaurin. (ns:nanoseconds).

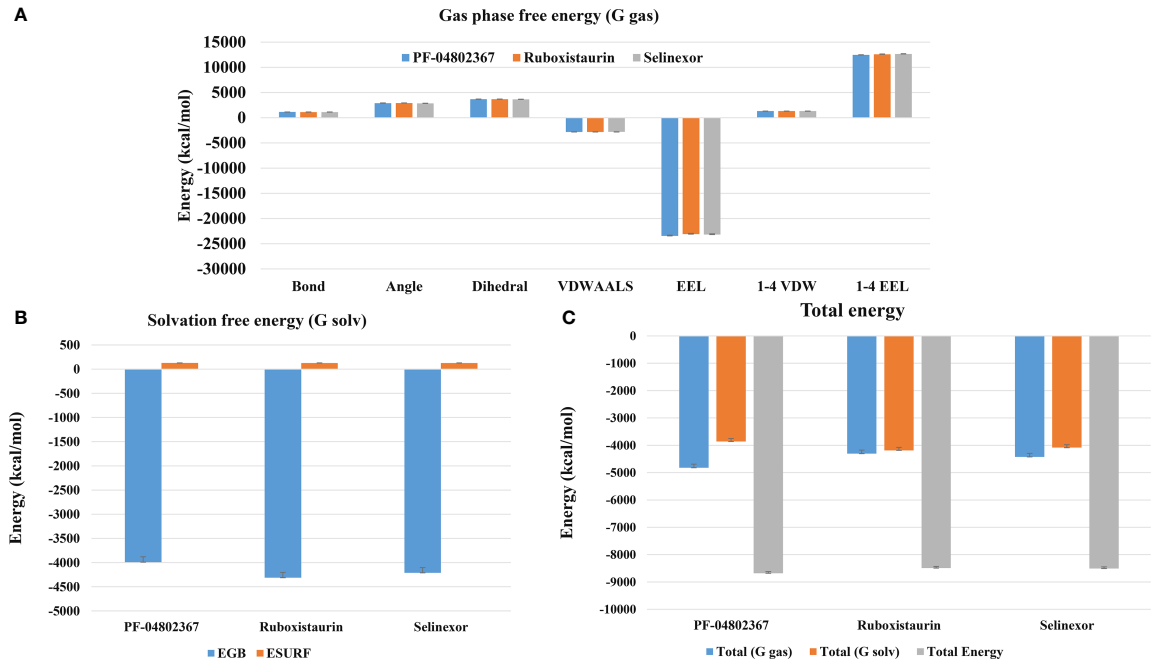


FIGURE 9
MM-PBSA binding free energy calculation. **(A)** Representative contributions of each Gas phase free energy (G-gas) components for GSK3 with the selected drugs (PF-04802367, ruboxistaurin, and selinexor); **(B)** The solvation free energy for GSK3 in complex with the selected drugs; **(C)** Total of gas phase and solvation free energies along with cumulative total binding free energy for PF-04802367, ruboxistaurin, and selinexor.

Recently, one of the identified drugs, selinexor, was also found to be effective against SARS-CoV-2. It is currently in the clinical trial recruitment phase (NCT04534725) for COVID-19 treatment.

We anticipate that the current research, which combines data curation from relevant databases with ML-based predictive algorithms to identify possible therapeutic candidates for COVID-19, could complement ongoing antiviral research efforts. These artificial intelligence–based pipelines may help in the design of preclinical laboratory studies, future clinical trials, and drug discovery. These approaches may also help to improve our understanding of other diseases and related biological phenomena. COVID-19 and potential future outbreaks of coronaviruses may be treatable because of the interference with the conserved dependence of the N protein on GSK3 and its potential role in the regulation of autophagy.

Data availability statement

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

Author contributions

Conceptualization, RP and SC. methodology, RP and BA. Writing—original draft preparation, RP. Writing—review and editing, RP, BA, NQ, and SC. Visualization, BA, and NQ. Supervision, SC. Funding acquisition, SC. All authors contributed to the article and approved the submitted version.

Funding

The Korea Drug Development Fund funded by the Ministry of Science and ICT; the Ministry of Trade, Industry, and Energy; and the Ministry of Health and Welfare (HN21C1058) supported this work. The National Research Foundation of Korea (NRF-2022M3A9G1014520, 2019M3D1A1078940, and 2019R1A6A1A11051471) also supported thisstudy.

Conflict of interest

Authors RHP and SC were employed by S&K Therapeutics.

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/fendo.2023.1084327/full#supplementary-material

References

1. V'kovski P, Kratzel A, Steiner S, Stalder H, Thiel V. Coronavirus biology and replication: Implications for SARS-CoV-2. *Nat Rev Microbiol* (2021) 19:155–70. doi: 10.1038/S41579-020-00468-6

2. Fehr AR, Perlman S. Coronaviruses: An overview of their replication and pathogenesis. *Methods Mol Biol* (2015) 1282:1–23. doi: 10.1007/978-1-4939-2438-7_1

3. Bouhaddou M, Memon D, Meyer B, White KM, Rezelj VV, Correa Marrero M, et al. The global phosphorylation landscape of SARS-CoV-2 infection. *Cell* (2020) 182:685–712.e19. doi: 10.1016/J.CELL.2020.06.034

4. Riva L, Yuan S, Yin X, Martin-Sancho L, Matsunaga N, Pache L, et al. Discovery of SARS-CoV-2 antiviral drugs through large-scale compound repurposing. *Nature* (2020) 586:113–9. doi: 10.1038/S41586-020-2577-1

5. Garcia G, Sharma A, Ramaiah A, Sen C, Purkayastha A, Kohn DB, et al. Antiviral drug screen identifies DNA-damage response inhibitor as potent blocker of SARS-CoV-2 replication. *Cell Rep* (2021) 35. doi: 10.1016/J.CELREP.2021.108940

6. Dittmar M, Lee JS, Whig K, Segrist E, Li M, Kamalia B, et al. Drug repurposing screens reveal cell-type-specific entry pathways and FDA-approved drugs active against SARS-Cov-2. *Cell Rep* (2021) 35. doi: 10.1016/J.CELREP.2021.108959

7. Cubuk J, Alston JJ, Incicco JJ, Singh S, Stuchell-Brereton MD, Ward MD, et al. The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nat Commun* (2021) 12. doi: 10.1038/S41467-021-21953-3

8. Chang CK, Hou MH, Chang CF, Hsiao CD, Huang TH. The SARS coronavirus nucleocapsid protein–forms and functions. *Antiviral Res* (2014) 103:39–50. doi: 10.1016/J.ANTIVIRAL.2013.12.009

9. Surjit M, Lal SK. The SARS-CoV nucleocapsid protein: A protein with multifarious activities. *Infect Genet Evol* (2008) 8:397–405. doi: 10.1016/J.MEEGID.2007.07.004

10. Wu CH, Chen PJ, Yeh SH. Nucleocapsid phosphorylation and RNA helicase DDX1 recruitment enables coronavirus transition from discontinuous to continuous transcription. *Cell Host Microbe* (2014) 16:462–72. doi: 10.1016/J.CHOM.2014.09.009

11. Wu C, Liu Y, Yang Y, Zhang P, Zhong W, Wang Y, et al. Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharm Sin B* (2020) 10:766–88. doi: 10.1016/J.APSB.2020.02.008

12. Eldar-Finkelman H, Eisenstein M. Peptide inhibitors targeting protein kinases. *Curr Pharm Des* (2009) 15:2463–70. doi: 10.2174/138161209788682253

13. Doble BW, Woodgett JR. GSK-3: Tricks of the trade for a multi-tasking kinase. *J Cell Sci* (2003) 116:1175–86. doi: 10.1242/JCS.00384

14. Gould TD, Zarate CA, Manji HK. Glycogen synthase kinase-3: A target for novel bipolar disorder treatments. *J Clin Psychiatry* (2004) 65:7849. doi: 10.4088/JCP.v65n0103

15. Hur EM, Zhou FQ. GSK3 signalling in neural development. *Nat Rev Neurosci* (2010) 11:539–51. doi: 10.1038/NRN2870

16. Liu X, Verma A, Garcia G, Ramage H, Lucas A, Myers RL, et al. Targeting the coronavirus nucleocapsid protein through GSK-3 inhibition. *Proc Natl Acad Sci U.S.A.* (2021) 118. doi: 10.1073/PNAS.2113401118/-/DCSUPPLEMENTAL

17. Kaidanovich-Beilin O, Woodgett JR. GSK-3: Functional insights from cell biology and animal models. *Front Mol Neurosci* (2011) 4:40. doi: 10.3389/FNMOL.2011.00040

18. Liang MH, Chuang DM. Differential roles of glycogen synthase kinase-3 isoforms in the regulation of transcriptional activation. *J Biol Chem* (2006) 281:30479–84. doi: 10.1074/JBC.M607468200

19. Rao R, Patel S, Hao C, Woodgett J, Harris R. GSK3beta mediates renal response to vasopressin by modulating adenylate cyclase activity. *J Am Soc Nephrol* (2010) 21:428–37. doi: 10.1681/ASN.2009060672

20. Rudd CE. GSK-3 inhibition as a therapeutic approach against SARs CoV2: Dual benefit of inhibiting viral replication while potentiating the immune response. *Front Immunol* (2020) 11:1638. doi: 10.3389/FIMMU.2020.01638

21. White TC, Yi Z, Hogue BG. Identification of mouse hepatitis coronavirus A59 nucleocapsid protein phosphorylation sites. *Virus Res* (2007) 126:139–48. doi: 10.1016/J.VIRUSRES.2007.02.008

22. Calvo E, Escors D, López JA, González JM, Álvarez A, Arza E, et al. Phosphorylation and subcellular localization of transmissible gastroenteritis virus nucleocapsid protein in infected cells. *J Gen Virol* (2005) 86:2255–67. doi: 10.1099/VIR.0.80975-0

23. Wu CH, Yeh SH, Tsay YG, Shieh YH, Kao CL, Chen YS, et al. Glycogen synthase kinase-3 regulates the phosphorylation of severe acute respiratory syndrome coronavirus nucleocapsid protein and viral replication. *J Biol Chem* (2009) 284:5229–39. doi: 10.1074/JBC.M805747200

24. Emmott E, Munday D, Bickerton E, Britton P, Rodgers MA, Whitehouse A, et al. The cellular interactome of the coronavirus infectious bronchitis virus nucleocapsid protein and functional implications for virus biology. *J Virol* (2013) 87:9486–500. doi: 10.1128/JVI.00321-13

25. He L, Fei DL, Nagiec MJ, Mutvei AP, Lamprakis A, Kim BY, et al. Regulation of GSK3 cellular location by FRAT modulates mTORC1-dependent cell growth and sensitivity to rapamycin. *Proc Natl Acad Sci U.S.A.* (2019) 116:19523–9. doi: 10.1073/PNAS.1902397116

26. Bautista SJ, Boras I, Vissa A, Mecica N, Yip CM, Kim PK, et al. mTOR complex 1 controls the nuclear localization and function of glycogen synthase kinase 3β. *J Biol Chem* (2018) 293:14723–39. doi: 10.1074/JBC.RA118.002800

27. Qin X, Jiang B, Zhang Y. 4E-BP1, a multifactor regulated multifunctional protein. *Cell Cycle* (2016) 15:781–6. doi: 10.1080/15384101.2016.1151581

28. Sridharan S, Basu A. Distinct roles of mTOR targets S6K1 and S6K2 in breast cancer. *Int J Mol Sci* (2020) 21. doi: 10.3390/IJMS21041199

29. Settembre C, Fraldi A, Medina DL, Ballabio A. Signals from the lysosome: a control centre for cellular clearance and energy metabolism. *Nat Rev Mol Cell Biol* (2013) 14:283–96. doi: 10.1038/NRM3565

30. Li Y, Xu M, Ding X, Yan C, Song Z, Chen L, et al. Protein kinase c controls lysosome biogenesis independently of mTORC1. *Nat Cell Biol* (2016) 18:1065–77. doi: 10.1038/NCB3407

31. García-Pérez BE, González-Rojas JA, Salazar MI, Torres-Torres C, Castrejón-Jiménez NS. Taming the autophagy as a strategy for treating COVID-19. *Cells* (2020) 9. doi: 10.3390/CELLS9122679

32. Lakshmana MK. SARS-CoV-2-induced autophagy dysregulation may cause neuronal dysfunction in COVID-19. *Neural Regener Res* (2022) 17:1255–6. doi: 10.4103/1673-5374.327333

33. Kumar AV, Thakurta TG, Silvestrini MJ, Johnson JR, Reenan RA, Lapierre LR. Give me a SINE: How selective inhibitors of nuclear export modulate autophagy and aging. *Mol Cell Oncol* (2018) 5. doi: 10.1080/23723556.2018.1502511

34. Sedykh A, Zhu H, Tang H, Zhang L, Richard A, Rusyn I, et al. Use of *in vitro* HTS-derived concentration-response data as biological descriptors improves the accuracy of QSAR models of *in vivo* toxicity. *Environ Health Perspect* (2011) 119:364–70. doi: 10.1289/EHP.1002476

35. Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin II, Cronin M, et al. QSAR modeling: Where have you been? where are you going to? *J Med Chem* (2014) 57:4977–5010. doi: 10.1021/JM4004285

36. Neves BJ, Braga RC, Melo-Filho CC, Moreira-Filho JT, Muratov EN, Andrade CH. QSAR-based virtual screening: Advances and applications in drug discovery. *Front Pharmacol* (2018) 9:1275. doi: 10.3389/FPHAR.2018.01275

37. Yan D, Lee S, Thakkar VD, Luo M, Moore ML, Plemper RK. Cross-resistance mechanism of respiratory syncytial virus against structurally diverse entry inhibitors. *Proc Natl Acad Sci U.S.A.* (2014) 111. doi: 10.1073/PNAS.1405198111

38. Silva RCMC, Ribeiro JS, da Silva GPD, da Costa LJ, Travassos LH. Autophagy modulators in coronavirus diseases: A double strike in viral burden and inflammation. *Front Cell Infect Microbiol* (2022) 12:845368. doi: 10.3389/FCIMB.2022.845368

39. Jang WD, Jeon S, Kim S, Lee SY. Drugs repurposed for COVID-19 by virtual screening of 6,218 drugs and cell-based assay. *Proc Natl Acad Sci U.S.A.* (2021) 118. doi: 10.1073/PNAS.2024302118

40. Mendez D, Gaulton A, Bento AP, Chambers J, de VM, Félix E, et al. ChEMBL: Towards direct deposition of bioassay data. *Nucleic Acids Res* (2019) 47:D930–40. doi: 10.1093/NAR/GKY1075

41. Fernández-Torras A, Comajuncosa-Creus A, Duran-Frigola M, Aloy P. Connecting chemistry and biology through molecular descriptors. *Curr Opin Chem Biol* (2022) 66. doi: 10.1016/J.CBPA.2021.09.001

42. Shameera Ahamed TK, Rajan VK, Sabira K, Muraleedharan K. QSAR classification-based virtual screening followed by molecular docking studies for identification of potential inhibitors of 5-lipoxygenase. *Comput Biol Chem* (2018) 77:154–66. doi: 10.1016/J.COMPBIOLCHEM.2018.10.002

43. Yap CW. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J Comput Chem* (2011) 32:1466–74. doi: 10.1002/JCC.21707

44. Jović A, Brkić K, Bogunović N. A review of feature selection methods with applications. In: *2015 38th international convention on information and communication technology, electronics and microelectronics, MIPRO 2015 - proceedings* (2015). Opatija, Croatia: IEEE, p. 1200–5. doi: 10.1109/MIPRO.2015.7160458

45. Ye Q, Li M, Zhou Y, Pang T, Xu L, Cao J, et al. Synthesis and biological evaluation of 3-benzisoxazolyl-4-indolylmaleimides as potent, selective inhibitors of glycogen synthase kinase-3β. *Molecules* (2013) 18:5498–516. doi: 10.3390/MOLECULES18055498

46. Heider F, Ansideri F, Tesch R, Pantsar T, Haun U, Döring E, et al. Pyridinylimidazoles as dual glycogen synthase kinase 3β/p38α mitogen-activated protein kinase inhibitors. *Eur J Med Chem* (2019) 175:309–29. doi: 10.1016/J.EJMECH.2019.04.035

47. Zhang HC, White KB, Ye H, McComsey DF, Derian CK, Addo MF, et al. Macrocyclic bisindolylmaleimides as inhibitors of protein kinase c and glycogen synthase kinase-3. *Bioorg Med Chem Lett* (2003) 13:3049–53. doi: 10.1016/S0960-894X(03)00644-9

48. Lavecchia A. Machine-learning approaches in drug discovery: methods and applications. *Drug Discovery Today* (2015) 20:318–31. doi: 10.1016/J.DRUDIS.2014.10.012

49. *GitHub - shankarpandala/lazypredict at master* . Available at: https://github.com/shankarpandala/lazypredict/tree/master (Accessed 9, 2023).

50. Pirzada RH, Haseeb M, Batool M, Kim M, Choi S. Remdesivir and ledipasvir among the FDA-approved antiviral drugs have potential to inhibit SARS-CoV-2 replication. *Cells* (2021) 10. doi: 10.3390/CELLS10051052

51. Valdés-Tresanco MS, Valdés-Tresanco ME, Valiente PA, Moreno E. gmx_MMPBSA: A new tool to perform end-state free energy calculations with GROMACS. *J Chem Theory Comput* (2021) 17:6281–91. doi: 10.1021/ACS.JCTC.1C00645

52. Ahmad B, Choi S. Unraveling the tomaralimab epitope on the toll-like receptor 2 *via* molecular dynamics and deep learning. *ACS Omega* (2022) 7:28226–37. doi: 10.1021/ACSOMEGA.2C02559

53. Vempati UD, Przydzial MJ, Chung C, Abeyruwan S, Mir A, Sakurai K, et al. Formalization, annotation and analysis of diverse drug and probe screening assay datasets using the BioAssay ontology (BAO). *PloS One* (2012) 7. doi: 10.1371/JOURNAL.PONE.0049198

54. Menichetti R, Kanekal KH, Bereau T. Drug-membrane permeability across chemical space. *ACS Cent Sci* (2019) 5:290–8. doi: 10.1021/ACSCENTSCI.8B00718/SUPPL_FILE/OC8B00718_SI_003.TXT

55. Di L, Kerns EH. Lipophilicity. In: *Drug-like properties* (2016). Boston: Academic Press, p. 39–50. doi: 10.1016/B978-0-12-801076-1.00005-8

56. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Delivery Rev* (2001) 46:3–26. doi: 10.1016/S0169-409X(00)00129-0

57. Gleeson MP. Generation of a set of simple, interpretable ADMET rules of thumb. *J Med Chem* (2008) 51:817–34. doi: 10.1021/JM701122Q

58. Lipinski CA. Lead- and drug-like compounds: The rule-of-five revolution. *Drug Discovery Today Technol* (2004) 1:337–41. doi: 10.1016/J.DDTEC.2004.11.007

59. Piryonesi SM, El-Diraby TE. Using machine learning to examine impact of type of performance indicator on flexible pavement deterioration modeling. *J Infrastructure Syst* (2021) 27:4021005. doi: 10.1061/(ASCE)IS.1943-555X.0000602

60. Vignaux PA, Minerali E, Foil DH, Puhl AC, Ekins S. Machine learning for discovery of GSK3β inhibitors. *ACS Omega* (2020) 5:26551–61. doi: 10.1021/ACSOMEGA.0C03302/ASSET/IMAGES/LARGE/AO0C03302_0005.JPEG

61. Berrouet C, Dorilas N, Rejniak KA, Tuncer N. Comparison of drug inhibitory effects (IC 50) in monolayer and spheroid cultures. *Bull Math Biol* (2020) 82:1–23. doi: 10.1007/S11538-020-00746-7/TABLES/3

62. Gandhi UH, Senapedis W, Baloglu E, Unger TJ, Chari A, Vogl D, et al. Clinical implications of targeting XPO1-mediated nuclear export in multiple myeloma. *Clin Lymphoma Myeloma Leuk* (2018) 18:335–45. doi: 10.1016/J.CLML.2018.03.003

63. Mostafa-Hedeab G, Al-kuraishy HM, Al-Gareeb AI, Welson NN, Batiha GE-S, Conte-Junior CA. Selinexor and COVID-19: The neglected warden. *Front Pharmacol* (2022) 13:884228. doi: 10.3389/FPHAR.2022.884228

64. Kashyap T, Murray J, Walker CJ, Chang H, Tamir S, Hou B, et al. Selinexor, a novel selective inhibitor of nuclear export, reduces SARS-CoV-2 infection and protects the respiratory system *in vivo*. *Antiviral Res* (2021) 192. doi: 10.1016/J.ANTIVIRAL.2021.105115

65. Lear T, Larsen M, Lin B, Cao Q, Alfaras I, Kennerdell J, et al. Modulation of lysosomal function as a therapeutic approach for coronaviral infections. *Res Sq* (2021) 36. doi: 10.21203/RS.3.RS-419305/V1

66. Choi Y, Bowman JW, Jung JU. Autophagy during viral infection - a double-edged sword. *Nat Rev Microbiol* (2018) 16:341–54. doi: 10.1038/S41579-018-0003-6

67. Deretic V, Saitoh T, Akira S. Autophagy in infection, inflammation and immunity. *Nat Rev Immunol* (2013) 13:722–37. doi: 10.1038/NRI3532

68. Wang C, Wang H, Zhang D, Luo W, Liu R, Xu D, et al. Phosphorylation of ULK1 affects autophagosome fusion and links chaperone-mediated autophagy to macroautophagy. *Nat Commun* (2018) 9. doi: 10.1038/S41467-018-05449-1

69. Mao J, Lin E, He L, Yu J, Tan P, Zhou Y. Autophagy and viral infection. *Adv Exp Med Biol* (2019) 1209:55–78. doi: 10.1007/978-981-15-0606-2_5

70. Pan HY, Valapala M. Regulation of autophagy by the glycogen synthase kinase-3 (GSK-3) signaling pathway. *Int J Mol Sci* (2022) 23. doi: 10.3390/IJMS23031709

71. Dowey R, Cole J, Roger Thompson AA, Hull RC, Huang C, Whatmore J, et al. Enhanced neutrophil extracellular trap formation in COVID-19 is inhibited by the protein kinase c inhibitor ruboxistaurin. *ERJ Open Res* (2022) 8. doi: 10.1183/23120541.00596-2021

72. Liang SH, Chen JM, Normandin MD, Chang JS, Chang GC, Taylor CK, et al. Discovery of a highly selective glycogen synthase kinase-3 inhibitor (PF-04802367) that modulates tau phosphorylation in the brain: Translation for PET neuroimaging. *Angew Chem Int Ed Engl* (2016) 55:9601–5. doi: 10.1002/ANIE.201603797

frontiers | Frontiers in Endocrinology

TYPE Original Research
PUBLISHED 21 April 2023
DOI 10.3389/fendo.2023.1163682

OPEN ACCESS

EDITED BY
Prasoon Agarwal,
National Bioinformatics Infrastructure Sweden, Sweden

REVIEWED BY
Yan Wang,
University of Pittsburgh, United States
Thirupugal Govindarajan,
Cincinnati Children's Hospital Medical Center, United States
Andi Alfian Zainuddin,
Hasanuddin University, Indonesia

*CORRESPONDENCE
Bo Qu
qubo_1970@163.com
Chaoqi Yan
yanchaoqi2002@163.com

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

SPECIALTY SECTION
This article was submitted to Obesity, a section of the journal Frontiers in Endocrinology

RECEIVED 11 February 2023
ACCEPTED 06 April 2023
PUBLISHED 21 April 2023

CITATION
Wang H, Zhang Y, Liu Y, Li H, Xu R, Fu H, Yan C and Qu B (2023) Comparison between traditional and new obesity measurement index for screening metabolic associated fatty liver disease. *Front. Endocrinol.* 14:1163682. doi: 10.3389/fendo.2023.1163682

Comparison between traditional and new obesity measurement index for screening metabolic associated fatty liver disease

Hongyan Wang[1†], Yuxue Zhang[2†], Yupeng Liu[3], Hui Li[4], Ruiling Xu[4], Hongmei Fu[1], Chaoqi Yan[1*] and Bo Qu[4*]

[1]Department of International Physical Examination and Health Center, Second Affiliated Hospital of Harbin Medical University, Harbin, China, [2]Department of Hygiene Microbiology, School of Public Health, Harbin Medical University, Harbin, China, [3]Department of Epidemiology and Health Statistics, Institute for Public Health and Management of Wenzhou Medical University, Wenzhou, China, [4]Department of Gastroenterology, Second Affiliated Hospital of Harbin Medical University, Harbin, China

Objectives: Obesity measurement indexes have certain screening value for metabolic diseases. To investigate associations between metabolic associated fatty liver disease (MAFLD) and obesity measurement indexes, including traditional indexes (BMI, WC, WHtR) and new indexes (ABSI, BRI, VAI, LAP), and assess their screening ability.

Methods: 12,658 subjects aged 18-75 at the Health Center of a Class III Grade A Hospital were included, who were divided into MAFLD and non-MAFLD groups. Spearman's rank correlation was used to study the correlation between MAFLD and obesity measurement indexes. Receiver operating characteristic (ROC) curves were used to calculate the area under the curve (AUC) to evaluate their screening accuracy.

Results: MAFLD had strong correlation with traditional BMI and new index LAP. ROC analysis showed that BMI had the highest AUC (0.89), followed by LAP (0.87). Stratification by BMI, LAP had the highest AUC (0.90) for MAFLD in population without obesity (BMI< 23kg/m^2), and its optimal cutoff value was 20.75, with a sensitivity and specificity of 85.9% and 79.0%, respectively.

Conclusions: We proposed a two-step screening strategy for MAFLD, combining BMI and LAP, and defined a high-risk population for MAFLD as follows: 1) BMI ≥ 23 kg/m^2; and 2) BMI< 23 kg/m^2 and LAP ≥ 20.75.

KEYWORDS

MAFLD, obesity measurement index, body mass index, lipid accumulation product, screening ability, screening strategy

1 Introduction

Nonalcoholic fatty liver disease (NAFLD) is the most common liver disease, with a worldwide prevalence of 25%. NAFLD is estimated to affect approximately 173-310 million (12.5-22.4%) people in China. The prevalence of NAFLD in China has increased from 17% (2003) to 22.4% (2012), which is comparable to that in the US (24.13%), Europe (23.71%) and Japan (25%) according to the latest global burden of liver disease data in 2019. In general, NAFLD has become the primary cause of liver disease burden in many countries and regions (1). NAFLD is a term used to describe metabolic dysfunction associated with liver disease and is closely related to genetic susceptibility, insulin resistance, type 2 diabetes (T2DM), metabolic syndrome (MetS), cardiovascular disease and so on. NAFLD can lead to death from not only liver diseases such as liver cirrhosis and hepatic carcinoma but also cardiovascular disease and extrahepatic malignant carcinoma, which seriously threatens human health and places a very large economic burden on society (2). Given that metabolic dysfunction better represents the heterogeneity of NAFLD. In early 2020, a panel of international experts from many countries and regions and the Asian-Pacific Association for the study of the Liver (APASL) proposed renaming NAFLD to metabolic associated fatty liver disease (MAFLD) and developed new diagnostic criteria for MAFLD. The criteria are based on evidence of hepatic steatosis, in addition to one of the following three criteria, namely, overweight/obesity, presence of T2DM, or evidence of metabolic dysregulation (3–5).

Human morphologic or obesity measurement indexes can reflect the degree of obesity and have certain screening value for metabolic diseases (6–9). In addition to body mass index (BMI), there are many other obesity measurement indexes, including traditional indexes such as waist circumference (WC), waist-hip ratio (WHR) and waist-to-height ratio (WHtR), as well as new indexes that have emerged in recent years, such as a body shape index (ABSI), body roundness index (BRI), visceral fat index (VAI) and lipid accumulation product (LAP) (10–13). In recent years, a small number of studies have explored the correlation between obesity measurement indexes and NAFLD, demonstrating the screening value of the new obesity measurement indexes for NAFLD (14, 15). However, since the concept of MAFLD was proposed in 2020, there have been few studies on the correlation between the obesity measurement index and MAFLD. The screening value of these indexes for MAFLD is unclear and deserves further exploration. Therefore, this study mainly utilized adult health examination big data from the Physical Examination and Health Center of Class III Grade A Hospital in a city in northeast China to explore the correlation between the above obesity measurement indexes and MAFLD and to compare the screening abilities of these indexes for MAFLD.

Abbreviations: MAFLD, metabolic associated fatty liver disease; NAFLD, non-alcoholic fatty liver disease; T2DM, type 2 diabetes mellitus; BMI, body mass index; WC, waist circumference; WHtR, waist-to-height ratio; ABSI, a body shape index; BRI, body roundness index; VAI, visceral fat index; LAP, lipid accumulation product; SBP, systolic blood pressure; DBP, diastolic blood pressure; tCHO, total cholesterol; LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol; TGs, triglycerides; FPG, fasting plasma glucose; HOMA-IR, homeostasis model assessment of insulin resistance index; hs-CRP, high-sensitivity C-reactive protein; ROC, receiver operating characteristics curve; AUC, area under the ROC.

2 Methods

2.1 Study design and subjects

This study was a retrospective study. Physical examination data derived from the International Physical Examination and Health Center of a Class III Grade A Hospital (Class III Grade A hospitals represent the highest level of classification in mainland China. They are capable of providing high-level medical and health services, and they also undertake higher education and scientific research tasks within their local region and surrounding areas) in a city in northeast China from January to December 2021 were collected. The inclusion criteria were as follows: 1) adults aged 18-75 years and 2) subjects whose physical examination included general physical examination, laboratory tests and abdominal ultrasound. General physical examination included height, weight, waist circumference (WC), systolic blood pressure (SBP) and diastolic blood pressure (DBP). Laboratory tests included total cholesterol (tCHO), low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), triglycerides (TGs) and fasting plasma glucose (FPG). The exclusion criteria were as follows: 1) patients with a history of liver surgery, liver cirrhosis, liver congestion, liver parasitosis, polycystic liver, portal system diseases, unclear nature of occupying lesions of the liver and other serious diseases of gallbladder, bile duct and pancreas; 2) subjects whose liver ultrasound diagnosis conclusion was not clear; 3) subjects who had incomplete physical examination data or error data; and 4) subjects who could not be accurately diagnosed with MAFLD according to existing physical examination data due to lack of homeostasis model assessment of insulin resistance index (HOMA-IR) and high-sensitivity C-reactive protein (hs-CRP), which were not routinely carried out in the Health Center. According to the inclusion and exclusion criteria listed above, a total of 12,658 subjects were included in this study (Supplementary Material_Flow Chart). The subjects were divided into a group with MAFLD and a group without MAFLD. MAFLD was diagnosed according to APASL clinical practice guidelines for the diagnosis and management of MAFLD as described as below (5). Written consent from subjects was waived because their data were retrospectively and anonymously extracted from the electronic information system of the hospital. This study was approved by the ethics committee of the Second Affiliated Hospital of Harbin Medical University (KY2022-058).

2.2 MAFLD diagnostic criteria

The diagnosis of MAFLD was based on the detection of liver steatosis by ultrasound together with the presence of at least one of three criteria that includes overweight/obesity, T2DM and clinical

evidence of metabolic dysfunction (3–5). Overweight/obesity is defined as BMI ≥23 kg/m^2 by Asian standards; T2DM is defined according to widely accepted international criteria; evidence of metabolic dysfunction includes the presence of at least two metabolic risk abnormalities: 1) WC ≥90/80 cm in Asian males and females; 2) SBP/DBP ≥130/85 mmHg or specific drug treatment; 3) TGs≥1.7 mmol/L or specific drug treatment; 4) HDL-C<1.0 mmol/L for males and<1.3 mmol/L for females or specific drug treatment; 5) prediabetes (i.e., fasting glucose level 5.6-6.9 mmol/L, or 2-hour postload glucose levels 7.8-11.0 mmol/L or HbA1c 5.6-6.4%; 6) HOMA-IR ≥2.5; 7) hs-CRP >2 mg/L. Since HOMA-IR and hs-CRP were not routinely carried out in the Health Center, 543 subjects could not be accurately diagnosed with MAFLD according to existing physical examination data and were excluded from the analysis.

2.3 Physical examination data collection

The physical examination records of adults aged 18-75 years who underwent physical examination in the Health Center of a Class III Grade A Hospital in a city in northeast China from January to December 2021 were reviewed through the electronic information system of the Health Center. Basic information, general physical examination, laboratory tests and abdominal ultrasound data were collected. The basic information included sex and age; the general physical examination included height, weight, WC, SBP and DBP; and the laboratory tests included tCHO, LDL-C, HDL-C, TGs and FPG. Different obesity measurement indexes, including BMI, WHtR, ABSI, BRI, VAI and LAP, were calculated using the following formulas:

$$BMI(kg/m^2) = Weight(kg)/Height^2(m^2) \quad (1)$$

$$WHtR = WC\,(cm)/Height\,(cm) \quad (2)$$

$$ABSI = WC\,(m)/[BMI\,(kg/m^2)^{2/3} \times Height\,(m)^{1/2}] \quad (3)$$

$$BRI = 364.2 - 365.5 \times [1 - (WC\,(m)/2\pi)^2/(0.5 \times Height\,(m))^2]^{1/2} \quad (4)$$

$$LAP\,(male) = [WC\,(cm) - 65] \times TG\,(mmol/L), LAP\,(female) = [WC\,(cm) - 58] \times TG\,(mmol/L) \quad (5)$$

$$VAI\,(male) = [WC\,(cm)/[39.68 + 1.88 \times BMI\,(kg/m^2)]] \times [TG\,(mmol/L)/1.03] \times [1.31/HDL - C\,(mmol/L)],$$
$$VAI\,(female) = [WC\,(cm)/[36.58+1.89 \times \mathrm{BMI}\,(kg/m^2)]] \times [TG\,(mmol/L)/0.81] \times [1.52/HDL - C\,(mmol/L)] \quad (6)$$

2.4 General physical examination

Weight and height were measured by an ultrasonic height and weight measuring instrument (SG-1000SC, Beijing Chioy Medical Technology Co., LTD) with the examinee wearing light clothing and no shoes. WC was measured around the abdomen with a soft tape parallel to the floor halfway between the lower rib and the iliac crest when the examinee relaxed and exhaled normally. Blood pressure was measured using an automatic medical electronic sphygmomanometer (ABP-1000, Beijing Chioy Medical Technology Co., LTD) after the examinee rested quietly for at least 5 minutes, and the values of SBP/DBP were recorded.

2.5 Laboratory tests

All physical examinees fasted at least 8 hours, and blood samples were collected by the registered nurse in the Health Center. All blood samples were tested by the department of clinical laboratory in the hospital. tCHO, LDL-C, HDL-C, TGs and FPG were detected by an auto biochemical analyzer (Roche MODULAR ISE900, Switzerland).

2.6 Ultrasound assessment of fatty liver

All physical examinees fasted at least 8 hours. Ultrasound examination was performed using an ultrasonography instrument (Siemens 2000, Germany) equipped with a curved array transducer probe (4-8 MHz). The ultrasound measurements were conducted by an accredited sonographer. Fatty liver was assessed according to the echogenicity of the liver parenchyma, the visibility of the vascular structure and the clarification of the diaphragm.

2.7 Statistical analysis

In this study, we included all individuals who met the inclusion and exclusion criteria during 2021. We used PASS (version 11.0.7) to calculate the statistical power of our analyses. The sample size of 6911 MAFLD subjects and 5747 non-MAFLD subjects in this study achieve 100% statistical power to detect AUCs between 0.70 and 0.90 using a two-sided z-test at the significance level of 0.05.

Statistical analyses were conducted using SPSS (version 19.0). Measurement data were analyzed by Kolmogorov–Smirnov normality tests; data that presented a nonnormal distribution are expressed as the median (lower quartile to upper quartile) [M (P25-P75)], and the Mann–Whitney U test was used for comparisons between groups. Count data are expressed as the frequency (rate) [n (%)], and the χ^2 test was used for comparisons between groups. After sex stratification, Spearman's rank correlation was used to assess the correlation between MAFLD and traditional and new obesity measurement indexes separately. Correlation coefficients (r) of 0.8-1.0, 0.6-0.8, 0.4-0.6, 0.2-0.4 and<0.2 were defined as very strong, strong, moderate, weak, and very weak correlation or no correlation, respectively. Receiver operating characteristic (ROC) curves were used to calculate the area under the curve (AUC) and 95% confidence interval (95% CI). The Z test was used to test the statistical significance of the AUCs for each index, and the optimal cutoff value and its corresponding sensitivity and specificity were

determined according to Youden's index. All reported *p* values were two-tailed, and the level of statistical significance was set at 0.05.

3 Results

3.1 Characteristics of study subjects and prevalence of MAFLD

A total of 12,658 subjects, including 5952 males and 6706 females, were included in this study. Overall, the prevalence of MAFLD was 54.48% (6911/12658). The prevalence of MAFLD in males was significantly higher than that in females (71.93% vs. 39.22%) ($p<0.001$). The prevalence of MAFLD in females gradually increased with age, but the prevalence of MAFLD in males was the highest among individuals between 50 and 59 years old. Although the prevalence of MAFLD in males declined slightly after 60 years of age, the overall prevalence of MAFLD was 70% or higher in both males and females after 60 years of age (as shown in Figure 1). The clinical characteristics of the subjects are shown in Table 1. The clinical values (SBP, DBP, tCHO, LDL-C, TGs and FPG) of the subjects with MAFLD were higher than those in individuals without MAFLD, but HDL-C in the MAFLD group was lower than that in the non-MAFLD group ($p<0.001$). Obesity measurement indexes (WC, WHtR, BMI, BRI, VAI, LAP) in the MAFLD group were higher than those in the non-MAFLD group ($p<0.001$), but ABSI showed no significant difference between the two groups ($p=0.40$).

3.2 The prevalence of MAFLD according to quartiles of different obesity measurement indexes

Seven obesity measurement indexes, including BMI, WC, WHtR, ABSI, BRI, VAI and LAP, were grouped according to their quartiles, and the prevalence of MAFLD in the quartiles of these indexes is shown in Table 2. The prevalence of MAFLD both in males and females increased linearly with quartile increases in BMI, WC, WHtR, BRI, VAI and LAP. However, the prevalence of MAFLD in males decreased linearly from Q1 to Q3 and leveled off from Q3 to Q4 as the ABSI quartile area increased, and the prevalence of MAFLD in females had no linear increasing or decreasing trend as the ABSI quartile increased (as shown in Figure 2).

3.3 Correlation analysis between MAFLD and different obesity measurement indexes

Spearman's rank correlation analysis showed that there was a very negative weak correlation in males (r=-0.17, $p<0.001$) and no correlation in females ($p>0.05$) between ABSI and MAFLD. There was a positive correlation between the other six indexes and MAFLD. Overall, MAFLD had a strong correlation with the traditional index BMI and the new index LAP (r>0.6) and had a moderate correlation with the traditional index WC, WHtR, the new index VAI and BRI (0.4<r<0.6). After stratifying by sex, the correlation coefficients between the other six indexes and MAFLD were as follows, ranked from high to low: BMI (0.54) >LAP (0.48) > WHtR (0.43) >WC (0.42) >VAI (0.39) >BRI (0.36) in males and BMI (0.66) >LAP (0.65) >WHtR (0.60) > WC (0.59) >BRI (0.57) >VAI (0.54) in females. In summary, in males, there was a moderate correlation between BMI, WHtR, WC, LAP and MAFLD (0.4<r<0.6) and a weak correlation between VAI, BRI and MAFLD (r<0.4). In females, there was a strong correlation between BMI and LAP and MAFLD (r>0.6) and a moderate correlation between WHtR, WC, VAI, and BRI and MAFLD (0.4<r<0.6) (Supplementary Material).

3.4 Comparison of the screening ability of different obesity measurement indexes for MAFLD

ROC curves of seven obesity measurement indexes to distinguish MAFLD were drawn (as shown in Figure 3), and their

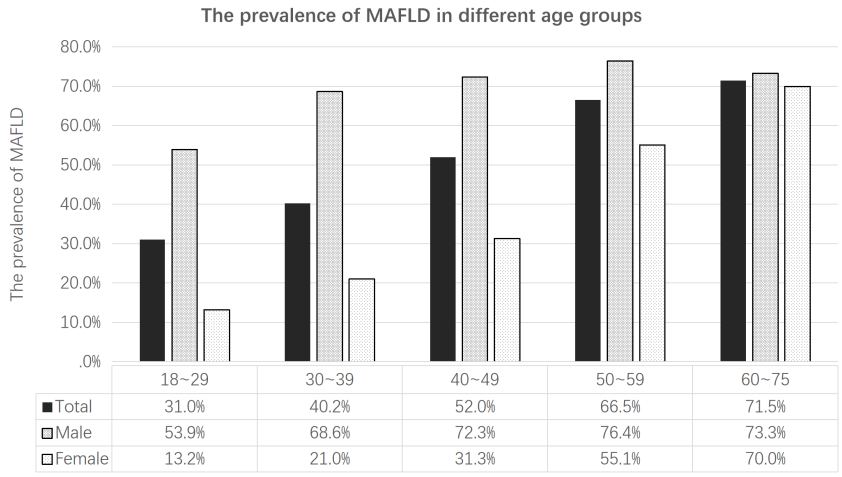


FIGURE 1
The prevalence of MAFLD in different sex and age groups.

TABLE 1 Basic characteristics of study subjects.

Variables	Total (n=12658)	non-MAFLD (n=5747)	MAFLD (n=6911)	Z/χ2	p
Demographic data characteristics					
Male (%)	5952 (47.02%)	1671 (28.07%)	4281 (71.93%)	1360.772	<0.001
Female (%)	6706 (52.98%)	4076 (60.78%)	2630 (39.22%)		
Age (years)	46 (37-57)	41 (34-51)	50 (40-59)	-30.26	<0.001
Clinical data characteristics					
SBP (mmHg)	123 (112-137)	116 (107-127)	130 (119-143)	-44.42	<0.001
DBP (mmHg)	77 (69-85)	72 (65-79)	81 (74-89)	-45.27	<0.001
t-CHO (mmol/L)	4.88 (4.29-5.53)	4.70 (4.18-5.32)	5.05 (4.43-5.70)	-18.937	<0.001
LDL-C (mmol/L)	2.97 (2.47-3.52)	2.83 (2.35-3.34)	3.11 (2.58-3.65)	-18.361	<0.001
HDL-C (mmol/L)	1.26 (1.07-1.50)	1.43 (1.22-1.65)	1.15 (1.00-1.33)	-49.016	<0.001
TGs (mmol/L)	1.32 (0.90-1.99)	0.96 (0.72-1.32)	1.74 (1.24-2.44)	-57.835	<0.001
FPG (mmol/L)	5.18 (4.84-5.64)	4.99 (4.72-5.31)	5.37 (5.00-5.97)	-38.93	<0.001
Obesity measurement indexes					
BMI (kg/m2)	24.39 (21.94-26.84)	21.80 (20.09-23.66)	26.28 (24.54-28.40)	-75.363	<0.001
WC (cm)	83 (77-90)	77 (70-83)	87 (83-94)	-65.882	<0.001
WHtR	0.491 (0.459-0.523)	0.458 (0.429-0.486)	0.512 (0.490-0.546)	-67.277	<0.001
ABSI	0.076 (0.072-0.079)	0.075 (0.073-0.079)	0.076 (0.720-0.080)	-0.841	0.40
BRI	4.12 (3.74-4.53)	3.80 (3.49-4.17)	4.36 (4.04-4.77)	-54.564	<0.001
VAI	1.61 (1.00-2.61)	1.10 (0.77-1.63)	2.19 (1.47-3.39)	-56.138	<0.001
LAP	28.65 (15.66-49.59)	15.64 (9.74-25.08)	43.56 (28.92-66.24)	-71.529	<0.001

MAFLD, metabolic associated fatty liver disease; non-MAFLD, without metabolic associated fatty liver disease; SBP, systolic blood pressure; DBP, diastolic blood pressure; tCHO, total cholesterol; LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol; TGs, triglycerides; FPG, fasting plasma glucose; BMI, body mass index; WC, waist circumference; WHtR, waist-height-ratio; ABSI, a body shape index; BRI, body roundness index; VAI, visceral adiposity indicators; LAP, lipid accumulation product. Data were expressed in form of Median (upper and lower quartile) [M(P25~P75)] or Frequency (rate) [n (%)]. n refers to the total number of each group.

TABLE 2 Prevalence of MAFLD according to quartiles of seven obesity measurement indexes.

Group	Quartile	Traditional Indexes			New Indexes			
		BMI	WC	WHtR	ABSI	BRI	VAI	LAP
Total	Q1	6.33%	9.68%	8.86%	58.11%	17.77%	20.21%	9.11%
(n=12658)	Q2	43.67%	46.11%	48.51%	51.46%	49.10%	44.45%	44.55%
	Q3	76.16%	72.35%	72.76%	53.23%	68.17%	67.93%	73.71%
	Q4	92.20%	89.61%	86.35%	57.38%	82.05%	85.07%	91.02%
Male	Q1	32.24%	38.31%	45.46%	85.76%	47.17%	45.50%	37.86%
(n=5952)	Q2	72.75%	62.21%	74.34%	72.55%	71.16%	68.36%	70.39%
	Q3	86.31%	79.23%	83.12%	65.92%	80.37%	82.10%	85.45%
	Q4	95.92%	93.07%	92.26%	65.49%	88.82%	91.25%	93.95%
Female	Q1	2.59%	1.22%	2.81%	49.14%	4.49%	7.65%	2.03%
(n=6706)	Q2	15.55%	13.08%	24.45%	34.26%	24.99%	24.26%	19.20%
	Q3	53.95%	50.31%	50.61%	34.69%	51.24%	49.56%	53.07%
	Q4	84.49%	76.93%	77.28%	44.02%	76.13%	75.10%	82.58%

BMI, body mass index; WC, waist circumference; WHtR, waist-height-ratio; ABSI, a body shape index; BRI, body roundness index; VAI, visceral adiposity indicators; LAP, lipid accumulation product. Quartiles, Q1, p0-p25; Q2, p25-p50; Q3, p50-p75; Q4, p75-p100. n refers to the total number of each group.

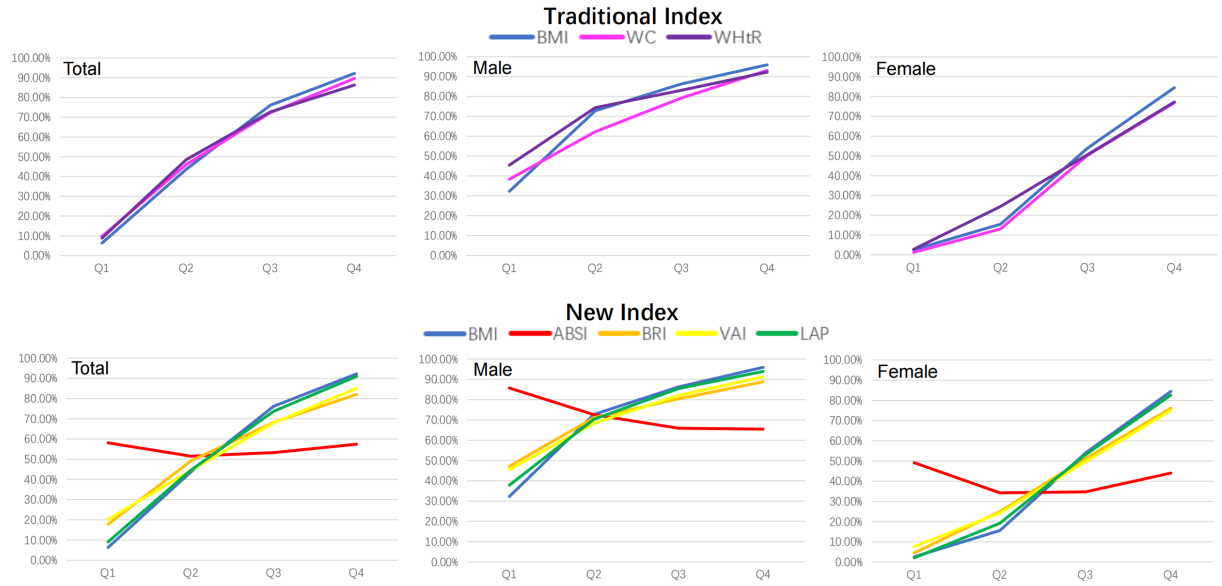


FIGURE 2
The prevalence trends of MAFLD according to quartiles of seven obesity measurement indexes. BMI was used as a reference standard. BMI, body mass index; WC, waist circumference; WHtR, waist-height-ratio; ABSI, a body shape index; BRI, body roundness index; VAI, visceral adiposity indicators; LAP, lipid accumulation product.

AUCs were calculated (as shown in Table 3). After stratification by sex, the AUCs of the seven obesity measurement indexes were ranked from high to low as follows: BMI (0.84) >LAP (0.81) >WHtR (0.78) >WC (0.77) >VAI (0.75) >BRI (0.73) >ABSI (0.59) in males and BMI (0.89) >LAP (0.88) >WHtR (0.86) >WC (0.84) >BRI (0.83) >VAI (0.82) >ABSI (0.50) in females. Among the seven indexes, ABSI had the lowest accuracy for MAFLD (AUC 0.59 for males and 0.50 for females). The AUCs of the other six indexes were all higher than 0.7, which had certain accuracy and certain predictive or screening value for MAFLD. Overall, except for ABSI, the accuracy of the other six indexes for screening MAFLD in females was higher than that in males. The traditional BMI had the best screening ability for MAFLD, with the highest accuracy (AUC 0.84 for males and 0.89 for females), and the optimal cutoff value was 24.74 kg/m^2 in males and 23.04 kg/m^2 in females. The new index LAP had better accuracy (AUC 0.81 for males and 0.88 for females), with an optimal cutoff value of 30.85 in males and 20.79 in females. Compared with BMI, the AUC of the other six indexes was statistically significant (p<0.001). However, there was no statistically significant difference between LAP and BMI in females (p=0.14) (Supplementary Material).

3.5 Comparison of the screening ability of different obesity measurement indexes for MAFLD in populations with or without obesity

To compare the screening ability of different obesity measurement indexes for MAFLD in populations with or without obesity, ROC analysis was stratified by BMI. BMI was grouped according to ≥ 23 kg/m^2 and< 23 kg/m^2, defined as obesity and non-obesity, respectively, by Asian standards. The proportion of subjects without obesity in this study was 34% (4304/12658). The prevalence of MAFLD in subjects without obesity was 10.87%. ROC curve analysis (Figure 4) showed that the AUCs of the seven obesity measurement indexes for MAFLD in subjects without obesity were ranked from high to low as follows: LAP (0.90) >VAI (0.88) >WHtR (0.80) >BRI (0.764) >BMI (0.761) >WC (0.76) >ABSI (0.64) (as shown in Table 4). In addition to ABSI, the AUCs of the other six indexes were all higher than 0.7, which had certain accuracy and certain predictive or screening value for MAFLD in populations without obesity. However, the new index LAP had the highest AUC of 0.90 (0.886-0.905); the optimal cutoff value was 20.75, and the sensitivity and specificity were 85.9% and 79.0%, respectively. The AUC of LAP was significantly better than those of the other indexes in predicting MAFLD in populations without obesity (p<0.01).

4 Discussion

At present, large-scale epidemiological studies on MAFLD are lacking, but NAFLD can be referred. A recent study based on a comprehensive search of the literature from 1999 to 2018 revealed an alarming national prevalence of NAFLD in China (29.6%). NAFLD prevalence is parallel with urbanization and industrialization. During the past two decades, the burden of NAFLD has increased substantially with the rapid development of economy and radical modifications in lifestyle in China. According to the statistics from the World Bank, the relative increases of national gross domestic product (GDP) per capita were in line with the pooled annual prevalence of NAFLD. This suggested that the epidemic of NAFLD is associated with its economic growth (16). But it's worth noting that NAFLD or MAFLD is a heterogeneous

10.3389/fendo.2023.1163682

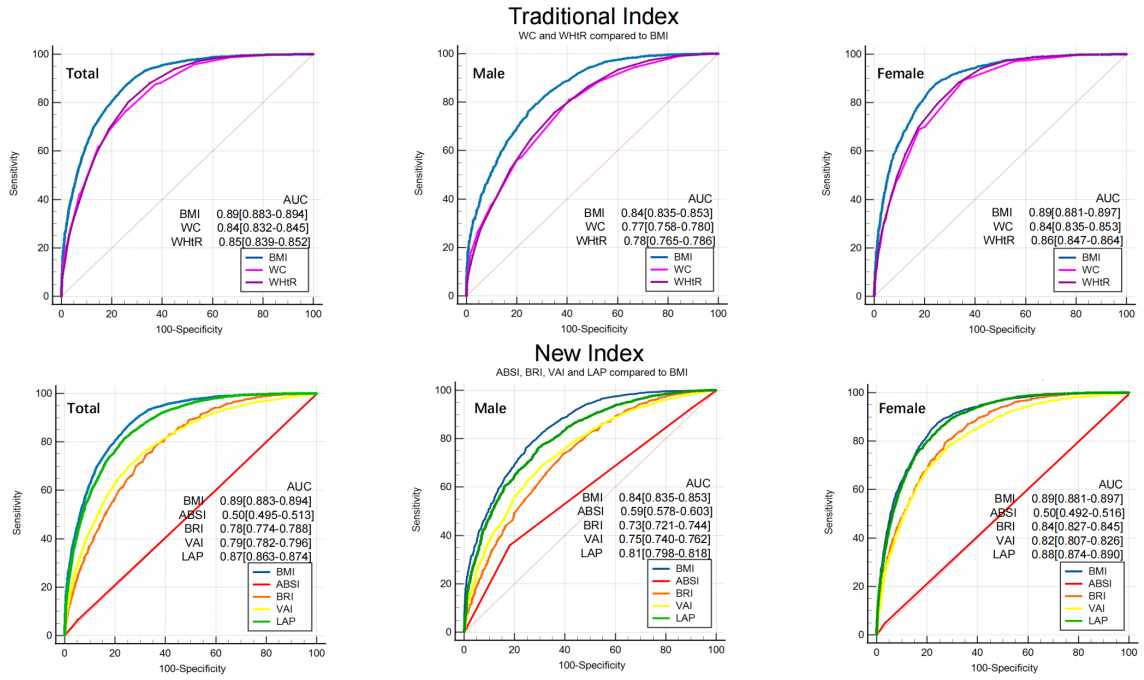


FIGURE 3
ROC of seven obesity measurement indexes for screening MAFLD. BMI was used as a reference standard. BMI, body mass index; WC, waist circumference; WHtR, waist-height-ratio; ABSI, a body shape index; BRI, body roundness index; VAI, visceral adiposity indicators; LAP, lipid accumulation product. AUC, area under the curve.

disease, its epidemiology also relates to sex, age, race, genetic variations, mild to moderate alcohol consumption, obesity, metabolism, lifestyle and educational level (17). With uneven economic development, different regional cultures and diverse lifestyles among the different provinces in China, the epidemiology of NAFLD or MAFLD has shown remarkable regional differences. A recent meta-analysis showed that the incidence of NAFLD is higher in northern China (35.78%) and lower in southeastern China (21.52%). Among provinces in northern China, Heilongjiang has the highest incidence, with up to 50.48% (18). This study was carried out in Harbin of Heilongjiang Province. The prevalence of MAFLD in the healthy physical examination population was 54.48% in this study, which was consistent with previous literature reports. The prevalence of MAFLD in Heilongjiang was significantly higher than the average prevalence in China. On the one hand, northeast China is located in a cold region with a long winter, and a cold climate has a great impact on lifestyle; thus, local residents tend to eat a high-calorie diet and exercise less. On the other hand, the subjects included in this study were a healthy physical examination population who generally have good economic conditions. The high prevalence of MAFLD should be taken seriously. It is important to have an accurate, effective, convenient and low-cost method for screening MAFLD. Therefore, this study explored the correlations between seven obesity measurement indexes and MAFLD and compared their screening accuracy for MAFLD. The seven obesity measurement indexes used in this study included three traditional indexes (BMI, WC, WHtR) and four new indexes (ABSI, BRI, VAI, LAP).

Both ABSI and BRI are new indexes to describe human body shape, which are calculated based on traditional indexes such as WC and BMI. Since body shape seems to be an important risk factor for premature death in the general population, Krakauer first proposed and established ABSI based on BMI and WC, which was initially used to predict the risk of death (19). A study published in 2016 showed that ABSI had a stronger association with total, cardiovascular and cancer mortality (20). In recent years, some researchers have also studied the correlation between ABSI and metabolic diseases. A recent study in 2021 showed that ABSI was significantly associated with cardio-ankle vascular index (CAVI) and the presence of MetS in the middle-aged population and helped to identify individuals with MetS and increased CAVI. ABSI could serve to identify individuals with MetS and increased arterial stiffness (21). After the concept of MAFLD was proposed in 2020, a research team from Sun Yat-sen University in China was the first to study the screening ability of anthropometric indexes for MAFLD, including BMI, WC, WHtR, ABSI, and BRI (22). Their study showed that the AUC of different indexes screening for MAFLD in males were as follows in descending order: BMI (0.81) >WC (0.79) >WHtR (0.77) =BRI (0.77) >ABSI (0.55). The AUCs of the different indexes above for female MAFLD patients was generally consistent with those of male MAFLD patients, but the AUC values were slightly lower than those of male MAFLD patients. Our study also included the above five indexes, and the AUC of each index indicating its screening ability for MAFLD was as follows in descending order: BMI (0.89) >WHtR (0.85) >WC (0.84) >BRI (0.78) >ABSI (0.50). Our study was generally consistent with theirs. Both our study and their study showed that the

10.3389/fendo.2023.1163682

TABLE 3 ROC analysis of seven obesity measurement indexes for screening MAFLD.

Group	AUC (95% CI)	Optimal cutoff points	Sensitivity%	Specificity%	Youden's index
Total (n=12658)					
BMI	0.89 (0.883-0.894)	23.62	86.8	74.8	0.62
WC	0.84 (0.832-0.845)	82	76.6	74.4	0.51
WHtR	0.85 (0.839-0.852)	0.48	80.2	73.5	0.54
ABSI	0.50 (0.495-0.513)	0.08	6.4	94.8	0.01
BRI	0.78 (0.774-0.788)	4.04	74.9	67.5	0.42
VAI	0.79 (0.782-0.796)	1.46	75.4	69.0	0.44
LAP	0.87 (0.863-0.874)	25.75	81.3	76.4	0.58
Male (n=5952)					
BMI	0.84 (0.835-0.853)	24.74	79.0	72.7	0.52
WC	0.77 (0.758-0.780)	86	81.0	59.2	0.40
WHtR	0.78 (0.765-0.786)	0.49	75.7	65.1	0.41
ABSI	0.59 (0.578-0.603)	0.07	36.0	81.9	0.18
BRI	0.73 (0.721-0.744)	3.99	73.8	60.6	0.34
VAI	0.75 (0.740-0.762)	1.59	69.5	68.5	0.38
LAP	0.81 (0.798-0.818)	30.85	76.1	70.6	0.47
Female (n=6706)					
BMI	0.89 (0.881-0.897)	23.04	87.5	75.9	0.63
WC	0.84 (0.835-0.853)	76	89.2	64.7	0.54
WHtR	0.86 (0.847-0.864)	0.46	88.3	66.4	0.55
ABSI	0.50 (0.492-0.516)	0.08	4.9	96.7	0.02
BRI	0.83 (0.827-0.845)	4.11	80.3	71.8	0.52
VAI	0.82 (0.807-0.826)	1.46	77.9	71.6	0.49
LAP	0.88 (0.874-0.890)	20.79	86.1	74.2	0.60

BMI, body mass index; WC, waist circumference; WHtR, waist-height-ratio; ABSI, a body shape index; BRI, body roundness index; VAI, visceral adiposity indicators; LAP, lipid accumulation product. AUC, area under the curve; 95% CI, 95% confidence interval; Youden's index=Sensitivity+Specificity-1. n refers to the total number of each group.

traditional indexes BMI, WC and WHtR had a certain screening ability for MAFLD, and BMI exhibited the best ability. The new index BRI also had a certain accuracy for screening MAFLD, with an AUC higher than 0.7, but ABSI had poor screening ability for MAFLD. Neither ABSI nor BRI exceeded the ability of the traditional BMI.

The new index ABSI and BRI based on morphologic measurements did not exceed the traditional index for screening MAFLD. Therefore, in addition to the above indexes, our study added two new indexes, LAP and VAI, which take into account both the external morphologic index and internal lipid metabolism-related index in the calculation. Theoretically, LAP and VAI could better reflect the degree of accumulation of body or visceral fat. Several researchers have studied the correlation between VAI and LAP and metabolic diseases, such as MetS, prediabetes and diabetes, metabolic-related cardiovascular disease, and polycystic ovary syndrome (23–28). As mentioned above, ABSI will not be described again due to its poor accuracy for screening MAFLD. Therefore, this section will mainly discuss the correlation between MAFLD and the other six indexes and compare their screening performance for MAFLD. Overall, our study showed that MAFLD had a strong correlation with the traditional index BMI and the new index LAP ($r>0.6$) and had a moderate correlation with the traditional indexes WC and WHtR and the new indexes VAI and BRI ($0.4<r<0.6$). ROC curve analysis showed that compared with BMI, the AUC of the other six indexes was statistically significant ($p<0.001$), but there was no statistically significant difference between LAP and BMI in females (p=0.14). ROC curve analysis showed that BMI had the highest AUC for MAFLD (0.84 in males and 0.89 in females), and LAP had a better AUC for MAFLD (0.81 in males and 0.88 in females) than the other indexes. The latest study in 2021 analyzed the correlation between NAFLD and various indexes, including lipid metabolism-related index, LAP, BMI, etc., according to the old diagnostic criteria of NAFLD and found that

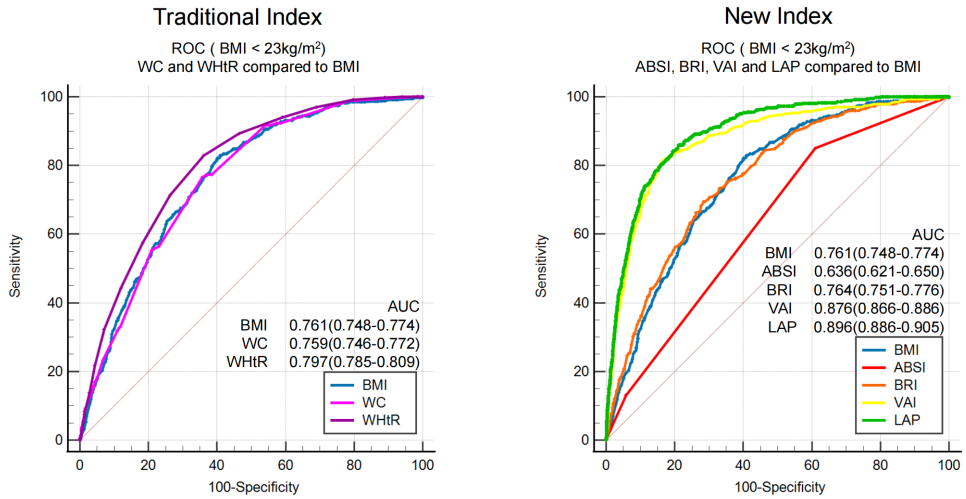


FIGURE 4
ROC analysis of seven obesity measurement indexes for screening MAFLD in populations with or without obesity.

lipid metabolism-related index and LAP had better screening ability for NAFLD than BMI; the AUCs of LAP and BMI were 0.8659 and 0.8577, respectively (p<0.05) (15). In contrast to the study above, our study found that the accuracy of LAP for screening MAFLD did not exceed that of BMI according to the new diagnostic criteria for MAFLD. Stratified by sex, the accuracy of LAP and BMI for screening MALFD in females was similar, but LAP did not exceed BMI. In our study, we found that the accuracy of BMI and LAP for screening MAFLD was higher in females than in males. We believe that this fundamental difference may be attributed to the more complex risk factors for MAFLD present in males. One factor that should not be overlooked is that men typically have a higher alcohol consumption than women. This difference in lifestyle habits may play a crucial role in the observed discrepancy between the sexes regarding the effectiveness of BMI and LAP for MAFLD screening.

TABLE 4 ROC analysis of seven obesity measurement indexes for screening MAFLD in populations with or without obesity.

Group	AUC (95% CI)	Optimal cutoff points	Sensitivity%	Specificity%	Youden's index
BMI≥23 (n=8354)					
BMI	0.75 (0.738-0.757)	25.39	95.7	15.4	0.37
WC	0.69 (0.689-0.709)	86	65.0	64.4	0.29
WHtR	0.70 (0.693-0.712)	0.49	71.5	57.7	0.29
ABSI	0.53 (0.514-0.536)	0.07	52.5	51.9	0.04
BRI	0.64 (0.628-0.649)	4.11	71.4	48.3	0.20
VAI	0.72 (0.705-0.725)	1.75	63.6	68.7	0.32
LAP	0.76 (0.753-0.772)	30.75	72.2	66.8	0.39
BMI<23 (n=4304)					
BMI	0.76 (0.748-0.774)	21.13	82.1	59.9	0.42
WC	0.76 (0.746-0.772)	76	76.5	64.3	0.41
WHtR	0.80 (0.785-0.809)	0.45	82.3	64.8	0.47
ABSI	0.64 (0.625-0.654)	0.07	81.2	43.9	0.25
BRI	0.76 (0.751-0.776)	3.90	69.4	71.6	0.41
VAI	0.88 (0.866-0.886)	1.74	81.4	83.6	0.65
LAP	0.90 (0.886-0.904)	20.75	85.9	79.0	0.65

BMI, body mass index; WC, waist circumference; WHtR, waist-height-ratio; ABSI, a body shape index; BRI, body roundness index; VAI, visceral adiposity indicators; LAP, lipid accumulation product; AUC, area under the curve, 95% CI, 95% confidence interval; Youden's index=Sensitivity+Specificity-1; n refers to the total number of each group.

It is worth noting that most MAFLD patients have co-existing overweight/obesity status; however, MAFLD is common in populations without obesity, especially in Asia. Due to the different criteria used to define obesity in different countries and regions, according to different literature reports, the prevalence of MAFLD in populations without obesity is as high as 5-45% (29–31). This study showed that the prevalence of MAFLD in population without obesity was 10.87%. At meantime, compared with BMI (AUC 0.76), LAP and VAI had excellent screening accuracy for MAFLD in populations without obesity, LAP had a slightly higher AUC than VAI (0.90 vs. 0.88, $p< 0.05$).

This study showed that among the entire population, traditional BMI had the highest screening accuracy for MAFLD with an AUC of 0.89, which was better than LAP (0.87) and VAI (0.79). In fact, stratified by BMI, LAP (AUC 0.76) and VAI (AUC 0.72) did not demonstrate significantly better screening accuracy for overweight or obese populations (BMI ≥ 23 kg/m^2) compared to BMI (AUC 0.75). Considering the simplicity and convenience of BMI measurement, which does not require additional laboratory tests, BMI should be preferentially considered for the general population. Strictly speaking, LAP and VAI can certainly be used as well. Additionally, in the non-obese population (BMI< 23 kg/m^2), both LAP (AUC 0.90) and VAI (AUC 0.88) showed significantly improved screening ability compared to BMI (AUC 0.76). Therefore, in the general population, the first step is to use BMI for screening, as its effectiveness is superior to LAP and VAI. In the overweight or obese population, further screening using LAP or VAI may yield better results. While LAP and VAI had excellent screening accuracy for MAFLD in non-obese populations, LAP had a slightly higher AUC than VAI (0.90 vs. 0.88, $p<0.05$) and required fewer variables for calculation (WC and TG) compared to VAI (WC, TGs, height, weight, and HDL-C), making LAP a more convenient and cost-effective option. Therefore, LAP could serve as an accurate, efficient, convenient, and low-cost screening index for MAFLD in non-obese populations. To sum up, for the general population, the first step is to calculate BMI. Individuals with overweight or obesity (BMI ≥23 kg/m^2) should be directly classified as high-risk for MAFLD. The second step is to calculate LAP for those without obesity (BMI< 23 kg/m^2), and those with LAP ≥ 20.75 should also be considered high-risk for MAFLD.

Currently, there is no consensus on universal screening methods for MAFLD, despite the significant health burden it poses. European guidelines support screening for MAFLD in high-risk patients with obesity or metabolic syndrome, while the American Association for the Study of Liver Disease (AASLD) questions the utility of routine screening for MAFLD in these high-risk individuals due to the lack of cost-effective tests and established effective pharmacologic treatments (32). However, new guidelines from the Asian-Pacific Association for the Study of the Liver (APASL) have been developed for the diagnosis and management of MAFLD, as mentioned earlier in our manuscript (5). The guideline base the diagnosis of MAFLD on the detection of fatty liver in conjunction with at least one of three criteria: overweight/obesity, T2DM, or clinical evidence of metabolic dysfunction (including waist circumference, blood pressure, blood lipids, blood glucose, HOMA-IR, hS-CRP, etc.). Fatty liver can be confirmed through various methods, such as ultrasound, transient elastography, regular MRI scanners applying magnetic resonance proton density fat fraction (MR-PDFF), magnetic resonance spectroscopy (MRS), liver biopsy, or serum biomarkers. However, this screening strategy for MAFLD is both time-consuming and costly, making it unsuitable for widespread implementation in the general population.

This study offers a non-invasive, cost-effective, and easily accessible method for screening high-risk populations for MAFLD with good sensitivity and specificity. This approach provides both patients with MAFLD and healthcare professionals with a convenient and persuasive tool for predicting MAFLD risk, and has significant implications for promoting individual stratification management in the realm of health economics. In conclusion, our research contributes to the existing knowledge on obesity measurement indexes and provides guidance for MAFLD screening. By proposing a method that is accurate, efficient, and cost-effective, we aim to support healthcare professionals and patients in identifying high-risk individuals for MAFLD more effectively. We believe our study has the potential to improve both individualized management and overall public health outcomes in the context of MAFLD.

The advantage of this study is that it is the first study to evaluate the correlation between obesity measurement indexes and MAFLD and compare their screening ability for MAFLD by using a healthy physical examination population in northeast China since the concept of MAFLD was proposed. As many as seven indexes, covering traditional and new indexes, were included. In addition, this study defined a population at high risk for MAFLD in a simple way and proposed a new screening strategy for MAFLD. Of course, there are also shortcomings in this study: 1) The design of this study is cross-sectional, which results in some limitations regarding causal inference; 2) The subjects of this study were mostly healthy people who underwent physical examination and could not represent the general population; 3) The determination of the cut-off value of different obesity measurement indexes may be affected by the study subjects and different regions, so it needs to be further verified in the multi-center study; 4) The research center did not carry out HOMA-IR and hs-CRP tests, and some subjects could not be diagnosed as MAFLD according to existing evidence and were excluded; 5) Whether obesity measurement indexes are related to the severity of MAFLD also deserves further investigation. The shortcomings will be further improved in future research.

5 Conclusion

In conclusion, regarding screening for MAFLD in the whole population, the traditional index BMI had the highest accuracy, followed by the new index LAP. However, when screening for MAFLD in populations without obesity (BMI< 23 kg/m^2), LAP had the highest accuracy, and the optimal cutoff value was 20.75, with a sensitivity and specificity of 85.9% and 79.0%, respectively.

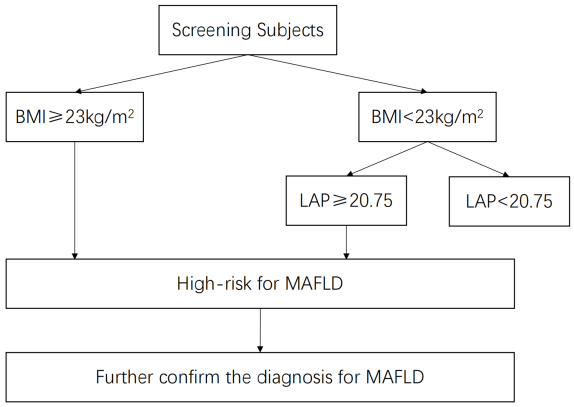


FIGURE 5
A two-step screening strategy or a high-risk population for MAFLD.

Therefore, we proposed a two-step screening strategy for MAFLD, combining BMI and LAP, and defined a high-risk population for MAFLD as follows: 1) BMI ≥ 23 kg/m^2; and 2) BMI< 23 kg/m^2 and LAP ≥ 20.75 (as shown in Figure 5).

Data availability statement

The datasets presented in this article are not readily available because of ethical and privacy restrictions. Requests to access the datasets should be directed to the corresponding authors.

Ethics statement

The studies involving human participants were reviewed and approved by The Ethics Committee of the Second Affiliated Hospital of Harbin Medical University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

HW and YZ contributed equally to this work and share first authorship. HW, BQ and CY designed the research. HW, HL, RX and HF acquired the data. HW, YZ and YL analyzed and interpreted the data. HW and YZ drafted the manuscript. HW, YL and BQ made critical revision of the manuscript for important intellectual content. All authors contributed to the article and approved the submitted version.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fendo.2023.1163682/full#supplementary-material

References

1. Xiao J, Wang F, Wong NK, He J, Zhang R, Sun R, et al. Global liver disease burdens and research trends: analysis from a chinese perspective. *J Hepatol* (2019) 71 (1):212–21. doi: 10.1016/j.jhep.2019.03.004

2. Powell EE, Wong VW, Rinella M. Non-alcoholic fatty liver disease. *Lancet* (2021) 397(10290):2212–24. doi: 10.1016/S0140-6736(20)32511-3

3. Eslam M, Newsome PN, Sarin SK, Anstee QM, Targher G, Romero-Gomez M, et al. A new definition for metabolic dysfunction-associated fatty liver disease: an international expert consensus statement. *J Hepatol* (2020) 73(1):202–9. doi: 10.1016/j.jhep.2020.03.039

4. Eslam M, Sanyal AJ, George JInternational Consensus Panel. Mafld: a consensus-driven proposed nomenclature for metabolic associated fatty liver disease. *Gastroenterology* (2020) 158(7):1999–2014.e1. doi: 10.1053/j.gastro.2019.11.312

5. Eslam M, Sarin SK, Wong VW, Fan JG, Kawaguchi T, Ahn SH, et al. The Asian pacific association for the study of the liver clinical practice guidelines for the diagnosis and management of metabolic associated fatty liver disease. *Hepatol Int* (2020) 14 (6):889–919. doi: 10.1007/s12072-020-10094-2

6. Feng J, He S, Chen X. Body adiposity index and body roundness index in identifying insulin resistance among adults without diabetes. *Am J Med Sci* (2019) 357 (2):116–23. doi: 10.1016/j.amjms.2018.11.006

7. Biyik Z, Guney I. Lipid accumulation product and visceral adiposity index: two new indices to predict metabolic syndrome in chronic kidney disease. *Eur Rev Med Pharmacol Sci* (2019) 23(5):2167–73. doi: 10.26355/eurrev_201903_17262

8. Wang H, Chen Y, Sun G, Jia P, Qian H, Sun Y. Validity of cardiometabolic index, lipid accumulation product, and body adiposity index in predicting the risk of hypertension in chinese population. *Postgrad Med* (2018) 130(3):325–33. doi: 10.1080/00325481.2018.1444901

9. Jablonowska-Lietz B, Wrzosek M, Wlodarczyk M, Nowicka G. New indexes of body fat distribution, visceral adiposity index, body adiposity index, waist-to-height ratio, and metabolic disturbances in the obese. *Kardiol Pol* (2017) 75(11):1185–91. doi: 10.5603/KP.a2017.0149

10. Banu H, Morshed MS, Sultana T, Shah S, Afrine S, Hasanat MA. Lipid accumulation product better predicts metabolic status in lean polycystic ovary syndrome than that by visceral adiposity index. *J Hum Reprod Sci* (2022) 15(1):27–33. doi: 10.4103/jhrs.jhrs_114_21

11. Ding YS, Li Y, Zhang XH, Ma RL, Guo H, Ma L, et al. The improved lipid accumulation product is an accurate index for predicting metabolic syndrome in the xinjiang population. *BioMed Environ Sci* (2021) 34(6):503–7. doi: 10.3967/bes2021.070

12. Stefanescu A, Revilla L, Lopez T, Sanchez SE, Williams MA, Gelaye B. Using a body shape index (absi) and body roundness index (bri) to predict risk of metabolic syndrome in peruvian adults. *J Int Med Res* (2020) 48(1):300060519848854. doi: 10.1177/0300060519848854

13. Wan H, Wang Y, Xiang Q, Fang S, Chen Y, Chen C, et al. Associations between abdominal obesity indices and diabetic complications: chinese visceral adiposity index and neck circumference. *Cardiovasc Diabetol* (2020) 19(1):118. doi: 10.1186/s12933-020-01095-4

14. Xie F, Pei Y, Zhou Q, Cao D, Wang Y. Comparison of obesity-related indices for identifying nonalcoholic fatty liver disease: a population-based cross-sectional study in china. *Lipids Health Dis* (2021) 20(1):132. doi: 10.1186/s12944-021-01560-3

15. Sheng G, Lu S, Xie Q, Peng N, Kuang M, Zou Y. The usefulness of obesity and lipid-related indices to predict the presence of non-alcoholic fatty liver disease. *Lipids Health Dis* (2021) 20(1):134. doi: 10.1186/s12944-021-01561-2

16. Zhou J, Zhou F, Wang W, Zhang XJ, Ji YX, Zhang P, et al. Epidemiological features of NAFLD from 1999 to 2018 in China. *Hepatology* (2020) 71(5):1851–64. doi: 10.1002/hep.31150

17. Huang TD, Behary J and Zekry A. Non-alcoholic fatty liver disease: a review of epidemiology, risk factors, diagnosis and management. *Intern Med J* (2020) 50 (9):1038–47. doi: 10.1111/imj.14709

18. Wu Y, Zheng Q, Zou B, Yeo YH, Li X, Li J, et al. The epidemiology of NAFLD in mainland China with analysis by adjusted gross regional domestic product: a meta-analysis. *Hepatol Int* (2020) 14(2):259–69. doi: 10.1007/s12072-020-10023-3

19. Krakauer NY, Krakauer JC. A new body shape index predicts mortality hazard independently of body mass index. *PloS One* (2012) 7(7):e39504. doi: 10.1371/journal.pone.0039504

20. Dhana K, Kavousi M, Ikram MA, Tiemeier HW, Hofman A, Franco OH. Body shape index in comparison with other anthropometric measures in prediction of total and cause-specific mortality. *J Epidemiol Community Health* (2016) 70(1):90–6. doi: 10.1136/jech-2014-205257

21. Sugiura T, Dohi Y, Takagi Y, Yokochi T, Yoshikane N, Suzuki K, et al. A body shape index could serve to identify individuals with metabolic syndrome and increased arterial stiffness in the middle-aged population. *Clin Nutr ESPEN* (2021) 46:251–8. doi: 10.1016/j.clnesp.2021.10.001

22. Yang JL, Luo YF, Shan RH, Zuo ZG, Yu Z, Zhu YW, et al. The ability of two new anthropometric indices to screen metabolic associated fatty liver disease among adults. *J Sun Yat-sen Univ* (2021) 42(6):854–63. doi: 10.13471/j.cnki.j.sun.yat-sen.univ (med.sci).2021.0607

23. Naghshband Z, Kumar L, Mandappa S, Niranjana MA, Malini SS. Visceral adiposity index and lipid accumulation product as diagnostic markers of metabolic syndrome in south indians with polycystic ovary syndrome. *J Hum Reprod Sci* (2021) 14(3):234–43. doi: 10.4103/jhrs.jhrs_12_21

24. Demirbas N, Kutlu R. Importance of measured body fat, visceral adiposity index, and lipid accumulation product index in predicting cardiometabolic risk factors. *Metab Syndr Relat Disord* (2021) 19(3):174–9. doi: 10.1089/met.2020.0098

25. Karatas S, Beysel S. Visceral adiposity index, triglyceride/high-density lipoprotein ratio, and lipid accumulation product index to discriminate metabolic syndrome among adult type 1 diabetes patients. *Metab Syndr Relat Disord* (2021) 19 (9):507–12. doi: 10.1089/met.2021.0047

26. Bagyura Z, Kiss L, Lux Á, Csobay-Novák C, Jermendy ÁL, Polgár L, et al. Association between coronary atherosclerosis and visceral adiposity index. *Nutr Metab Cardiovasc Dis* (2020) 30(5):796–803. doi: 10.1016/j.numecd.2020.01.013

27. Ahn N, Baumeister SE, Amann U, Rathmann W, Peters A, Huth C, et al. Visceral adiposity index (vai), lipid accumulation product (lap), and product of triglycerides and glucose (tyg) to discriminate prediabetes and diabetes. *Sci Rep* (2019) 9(1):9693. doi: 10.1038/s41598-019-46187-8

28. Amato MC, Giordano C, Galia M, Criscimanna A, Vitabile S, Midiri M, et al. Visceral adiposity index: a reliable indicator of visceral fat function associated with cardiometabolic risk. *Diabetes Care* (2010) 33(4):920–2. doi: 10.2337/dc09-1825

29. Chrysavgis L, Ztriva E, Protopapas A, Tziomalos K, Cholongitas E. Nonalcoholic fatty liver disease in lean subjects: prognosis, outcomes and management. *World J Gastroenterol* (2020) 26(42):6514–28. doi: 10.3748/wjg.v26.i42.6514

30. Chen F, Esmaili S, Rogers GB, Bugianesi E, Petta S, Marchesini G, et al. Lean nafld: a distinct entity shaped by differential metabolic adaptation. *Hepatology* (2020) 71(4):1213–27. doi: 10.1002/hep.30908

31. Wang AY, Dhaliwal J, Mouzaki M. Lean non-alcoholic fatty liver disease. *Clin Nutr* (2019) 38(3):975–81. doi: 10.1016/j.clnu.2018.08.008

32. Pal SC, Méndez-Sánchez N. Screening for MAFLD: who, when and how? *Ther Adv Endocrinol Metab* (2023) 14:20420188221145650. doi: 10.1177/20420188221145650

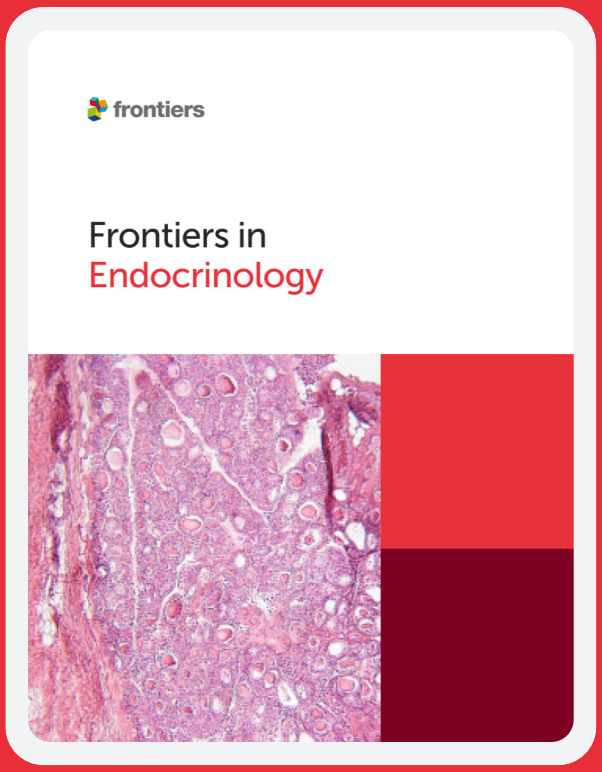
Frontiers in Endocrinology

Explores the endocrine system to find new therapies for key health issues

The second most-cited endocrinology and metabolism journal, which advances our understanding of the endocrine system. It uncovers new therapies for prevalent health issues such as obesity, diabetes, reproduction, and aging.

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