JDI Journal of Diabetes Investigation Open access
Official Journal of the Asian Association for the Study of Diabetes
ORIGINAL ARTICLE

Revisiting the causal impact of lipid traits on metabolic dysfunction-associated fatty liver disease: Insights from a multidimensional plasma lipid profile

Fang Xie[1,2], Wenkai Zheng[2], Jing Chen[1], Chuanxia Yao[2], Cong Li[2], Li Tang[3], Ping Li[2]*, Shanzhong Tan[1]*

[1]Department of Integrated TCM and Western Medicine, Nanjing Hospital Affiliated to Nanjing University of Chinese Medicine, Nanjing, Jiangsu Province, China, [2]Department of Liver Disease, Jinling Hospital Affiliated to Medical College of Nanjing University, Nanjing, Jiangsu Province, China, and [3]Department of Gastroenterology, Nanjing Hospital of Chinese Medicine Affiliated to Nanjing University of Chinese Medicine, Nanjing, Jiangsu Province, China

Keywords
Causal association, Mendelian randomization, Metabolic dysfunction-associated fatty liver disease

***Correspondence**
Shanzhong Tan
Tel.: +8613270731778
Fax: +86-025-85091790
E-mail address:
fsyy01455@njucm.edu.cn

Ping Li
Tel.: +8613813964300
Fax: +86-025-80864332
E-mail address:
leep2002@163.com

J Diabetes Investig 2025; 16: 917–928

doi: 10.1111/jdi.14413

ABSTRACT

Aims: Recent advancements in plasma lipidomes genome-wide association studies data have enhanced our understanding of lipid categories, significantly improving risk assessments for metabolic dysfunction-associated fatty liver disease (MAFLD) beyond traditional lipid biomarkers.

Materials and Methods: This study utilized Mendelian randomization (MR) to assess the causal relationships between 179 lipid species across 13 subclasses and MAFLD, primarily using the Wald ratio and IVW methods. Corrections were made using false discovery rate (FDR), supplemented by Bayesian colocalization analysis.

Results: Elevated levels of genetically predicted phosphatidylcholine (16:0_16:1) [$OR_{Wald\ ratio}$ = 2.638, 95% CI 1.557–4.469, $P = 3.11 \times 10^{-4}$], phosphatidylcholine (16:1_18:0) ($OR_{Wald\ ratio}$ = 2.644, 95% CI 1.559–4.486, $P = 3.11 \times 10^{-4}$), triacylglycerol (46:2) ($OR_{Wald\ ratio}$ = 2.515, 95% CI 1.524–4.153, $P = 3.11 \times 10^{-4}$), and triacylglycerol (48:2) (OR_{IVW} = 1.863, 95% CI 1.300–2.669, $P = 6.95 \times 10^{-4}$) were significantly associated with increased MAFLD risk, with rs1260326 within the GCKR gene playing a crucial role. Colocalization analysis indicated that in significant evidences, the posterior probability for hypothesis 4 was over 80%, identifying rs780093 as a shared causal variant. Additionally, 16 suggestive evidences were identified.

Conclusion: The study confirmed the significant role of specific lipid molecules in influencing MAFLD risk, providing new scientific bases and potential therapeutic targets for future treatment strategies.

INTRODUCTION

Non-alcoholic fatty liver disease (NAFLD) is the most prevalent chronic liver condition worldwide, characterized by excessive fat accumulation in the liver, occurring without significant alcohol consumption[1]. In recent years, the international consensus expert panel has redefined MAFLD as metabolic dysfunction-associated fatty liver disease (MAFLD) based on a deeper understanding of metabolic factors, stigmatization, and diagnostic criteria[2]. Between 1990 and 2006, the global prevalence of MAFLD was 25.3%, which increased to 38.0% between 2016 and 2019. With the increasing prevalence of obesity and aging, this rate is expected to rise further[3]. MAFLD spans a spectrum of liver pathologies from simple steatosis to steatohepatitis, fibrosis, and cirrhosis and is considered a manifestation of metabolic dysfunction in the liver[1, 4]. The complexity of its etiology, which involves an interplay of metabolic, inflammatory, genetic, and environmental factors, increases the difficulty of disease prevention and management, making early intervention a challenge[5, 6]. While liver biopsy aids in diagnosis, MAFLD is often associated with obesity, hypertension, and type 2 diabetes and can progress to more severe liver diseases such

Fang Xie and Wenkai Zheng both are co-contributing authors.
Received 23 October 2024; revised 21 December 2024; accepted 9 January 2025

as metabolic-associated steatohepatitis (NASH), fibrosis, cirrhosis, and hepatocellular carcinoma, posing significant economic burdens and severe health risks[7]. However, it is important to note that MAFLD can also occur in non-obese individuals, suggesting that genetic and metabolic factors play a crucial role in its pathogenesis[7]. Therefore, accurately defining metabolic-related risk factors is crucial for early prevention and devising targeted treatment strategies.

Lipid metabolism plays a pivotal role in the pathogenesis of MAFLD. Lipids are crucial not only as basic structural components of cellular membranes but also in regulating inflammation, oxidative balance, and insulin signaling through their roles in cell signaling and energy storage, directly impacting liver health[8]. Extensive research supports dyslipidemia as an independent risk factor for MAFLD[9, 10]. However, observational studies that typically identify correlations do not imply causation and are inherently flawed due to confounders and reverse causality. Although randomized controlled trials (RCTs) are the gold standard for establishing causation, they are often impractical due to various constraints[11]. With the advancement of genetic technologies and the proliferation of genome-wide association studies (GWAS), Mendelian randomization (MR) has emerged as a powerful tool for establishing causation, exploiting genetic variants as instrumental variables (IVs) which are randomly allocated at conception, thereby effectively circumventing confounders and reverse causality[12]. While causal associations between standard lipid profiles [low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), triglycerides (TG), and total cholesterol (TC)] and MAFLD have been preliminarily explored in MR studies[13–15], recent research has expanded our understanding of various lipid types, enhancing our knowledge of the diversity and complexity of plasma lipids[16]. Accurately identifying different lipid species is crucial for refining MAFLD risk stratification and guiding future therapeutic interventions. This study aims to transcend the limitations of conventional lipid phenotypes, re-examining the relationship between lipid traits and MAFLD using the latest plasma lipidomic GWAS data to explore the genetic susceptibility of 179 lipid species across 13 lipid subclasses within four major categories, providing a more precise and detailed causal perspective on the risk of MAFLD.

METHODS

Study design

This study utilizes GWAS data to perform univariate MR analysis, employing genetic variants as instrumental variables to explore the causal relationships between 179 lipid species and MAFLD among European ancestry. Reporting adheres to the STROBE-MR guidelines[17]. The validity of MR causal evidence is predicated on three foundational assumptions: (1) The relevance assumption, which requires that genetic variants used as instrumental variables must exhibit a strong association with the exposure. (2) The independence assumption dictates that genetic variants must be uncorrelated with any confounders that could affect the outcome. (3) The exclusion restriction assumption states that the impact of the genetic variants on the outcome must operate exclusively through the exposure, without any alternative pathways[18]. A flowchart of the study design is presented in Figure 1.

Instrumental variable selection criteria

IVs were initially selected from the exposure GWAS dataset based on genome-wide significance ($P < 5 \times 10^{-8}$) and stringent linkage disequilibrium (LD) ($r^2 < 0.001$, within a 10 MB) clumping using Phase 3 of the 1,000 Genomes Project. Single nucleotide polymorphisms (SNPs) failing to match in the outcome GWAS dataset were discarded to ensure precise causal inference, avoiding the use of proxy SNPs for imputation. Further, SNPs were harmonized between exposure and outcome datasets, explicitly excluding alleles with intermediate effect allele frequencies (EAF) > 0.42 or ambiguous allele pairings (e.g., T/A vs T/C)[19]. Additionally, SNPs where the association with the outcome exceeded that with the exposure were excluded using the Steiger test, with a rigorous threshold set at 5×10^{-520}. The *F*-statistic was calculated for each SNP, and those with $F < 10$ were deemed weak IVs and subsequently removed[21].

Summary data sources for plasma lipidome and MAFLD

The GWAS summary data for the plasma lipidome encompassed 7,174 Finnish individuals, employing high-resolution shotgun lipidomics to measure four major lipid classes, 13 lipid subclasses, and 179 lipid species[16]. Univariate and multivariate GWAS analyses identified 495 genetic trait associations, including 56 loci, of which eight were newly discovered, significantly enhancing statistical power. Additionally, detailed mapping pinpointed 26 loci with a high probability of causal variants, providing potential targets for future therapies aimed at lipid metabolism disorders.

Summary data for MAFLD were sourced from the FinnGen Consortium (R10) and a GWAS meta-analysis by Ghodsian *et al.* Within the FinnGen Consortium[22], MAFLD, defined by ICD-10-K76.0, comprised 2,568 cases (1,440 women, 1,128 men) and 409,613 controls, with 19,345,634 SNPs analyzed. The unadjusted prevalence was 0.62%, with a mean age at the first event of 53.09 years, detailed at https://r10.risteys.finngen.fi/endpoints/MAFLD. Ghodsian *et al.*'s summary data[23], derived from a meta-analysis including cohorts from electronic medical records and genomics (eMERGE), FinnGen (R4), the UK Biobank, and the Estonian Biobank, included 8,434 cases and 770,180 controls across 6,784,388 SNPs, identifying five potential susceptibility loci for MAFLD and providing insights into its genetic architecture, which could inform future predictions and treatments.

Given the comprehensive scope and the larger number of sequenced SNPs in the FinnGen R10 data, it served as the discovery dataset, ensuring no potential lipidomic phenotype was overlooked, while Ghodsian *et al.*'s data functioned as a

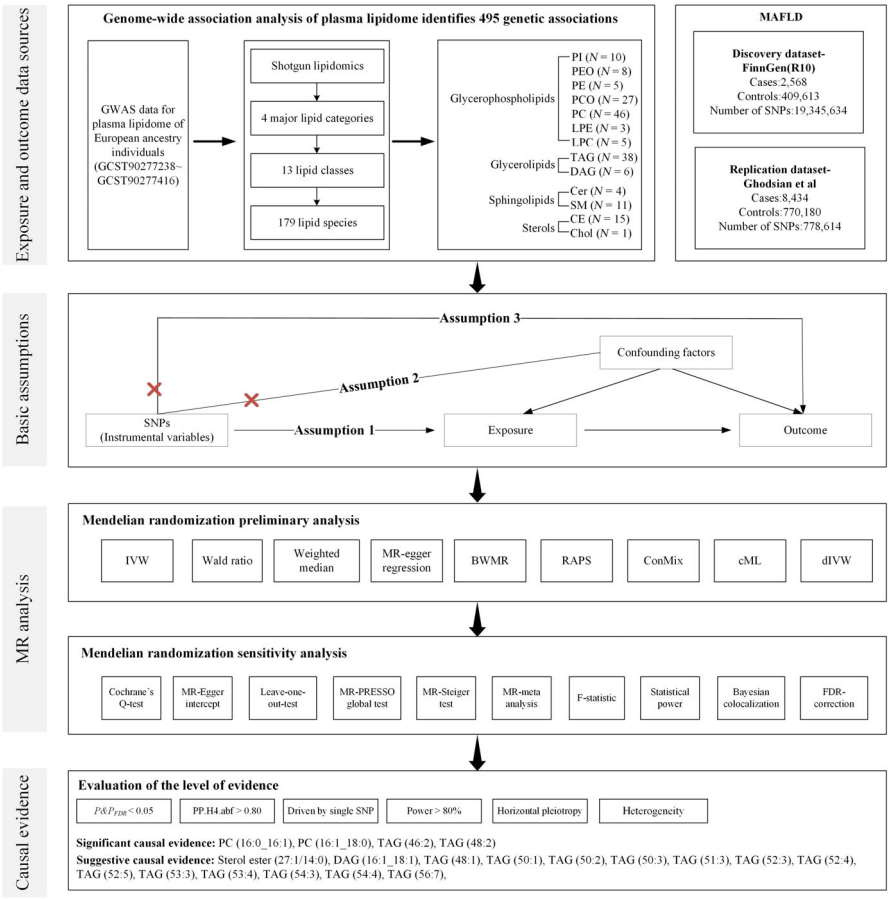

Figure 1 | Study design. BWMR, Bayesian weighted Mendelian randomization; CE, cholesteryl ester; Cer, ceramide; Chol, cholesterol; cML, constrained maximum likelihood; ConMix, contamination mixture; DAG, diacylglycerols; dIVW, debiased inverse-variance weighted; FDR, false discovery rate; GWAS, genome-wide association studies; IV, instrumental variable; IVW, inverse-variance-weighted; LPC, lysophosphatidylcholine; LPE, lysophosphatidylethanolamine; MALFD, metabolic dysfunction-associated fatty liver disease; MR, Mendelian randomization; MR-PRESSO, MR pleiotropy residual sum and outlier; PC, phosphatidylcholine; PCO, phosphatidylcholine-ether; PE, phosphatidylethanolamine; PEO, phosphatidylethanolamine-ether; PI, phosphatidylinositol; PPH4, posterior probability for hypothesis 4; RAPS, robust adjusted profile score; SM, sphingomyelin; SNP, single nucleotide polymorphism; TAG, triacylglycerols.

replication set. Causal effects from both datasets were synthesized in a meta-analysis, selecting effect models based on heterogeneity, with non-significant causal evidence further excluded. The minimal overlap in samples between exposure and outcome (<2%) precluded the winner's curse, thus avoiding an increase in Type I error rates. All utilized summary data underwent rigorous quality control, adjusted by age, gender, and up to 20 principal genetic components. The study leveraged publicly available GWAS summary data for secondary analysis, thereby obviating the need for ethical approval or clinical registration, with detailed approval documentation such as patient consent available in the cited original research.

Statistical analyses

MR analysis and Bayesian colocalization analysis

When only a single IV is available for analysis, the primary method employed is the Wald ratio. This ratio, calculated by dividing the SNP-exposure association (β_X) by the SNP-outcome association (β_Y), estimates the causal effect for each genetic variant, assuming that the IV significantly influences exposure, is independent of confounding factors, and affects the outcome solely through the exposure[24]. With two or more IVs, the primary method typically utilizes the inverse-variance weighted (IVW) method, which estimates causal effects by averaging the Wald ratios of each SNP, weighted accordingly[24]. This method presumes that all selected SNPs are valid instruments. For analyses involving three or more IVs, additional methods such as the Weighted Median[25] and MR-Egger[26] are incorporated. The former estimates the median, assuming at least 50% of the IVs are valid, whereas the latter adjusts for potential pleiotropy by regressing the Wald ratios, thus proving especially useful when all SNPs may be affected by pleiotropy to some extent.

Due to a very stringent IV selection process, the number of SNPs ultimately used for analysis is limited. Using online tools such as LDtrait and Phenoscanner to identify and exclude SNPs based on their traits could lead to "blind denoising," reducing statistical power and potentially leaving no SNPs for analysis. Therefore, this study did not proactively exclude SNPs. Instead, it supplemented the analysis with advanced methods such as constrained maximum likelihood (cML)[27], robust adjusted profile score (RAPS)[28], debiased inverse-variance weighted (dIVW)[29], Bayesian weighted Mendelian randomization (BWMR)[30], and contamination mixture (ConMix)[31] to correct for potential heterogeneity and pleiotropy-induced bias, providing multidimensional validation. Additionally, Bayesian colocalization analysis was conducted to ensure that the exposure and outcome shared a common causal variant at the same genomic location, thus mitigating bias from unremoved confounding SNPs[32]. This approach uses five hypothesis models (H0–H4) to explore the causal relationship between two traits. H0 assumes no genetic correlation between the traits; H1 indicates a genetic variant related only to the exposure trait; H2 relates a variant only to the outcome trait; H3 suggests genetic correlation between traits with different causal variants; and H4 posits a common causal variant affecting both traits. When the posterior probability for hypothesis 4 (PPH4) exceeds 0.8, H4 is considered significant, indicating a shared causal relationship.

Sensitivity analysis

In the sensitivity analysis, Cochran's Q test is employed to detect heterogeneity among IVs, with a P-value less than 0.05 indicating significant heterogeneity[33]. MR-Egger regression is used to assess horizontal pleiotropy, where the statistical significance of its intercept ($P < 0.05$) suggests notable pleiotropy[34]. The MR Pleiotropy Residual Sum and Outlier (MR-PRESSO) test enhances the reliability of results by identifying and excluding outlier SNPs that may bias causal estimates, with a global P-value below 0.05 indicating potential horizontal pleiotropy[35]. Additionally, leave-one-out analysis assesses the robustness of results by sequentially excluding each SNP and recalculating the causal estimate, helping to identify and mitigate the undue influence of dominant SNPs[36]. Considering that IVs explain only a small fraction of genetic variance, statistical power becomes a significant challenge in MR analysis. This study uses an online tool (https://shiny.cnsgenomics.com/mRnd/) to calculate the statistical power for binary outcomes of positive findings[37], requiring a statistical power of ≥80% to minimize the risk of type II errors—failing to detect an existing effect. Further, considering multiple testing, significance is adjusted using the false discovery rate (FDR), with $P_{FDR} < 0.05$ & $P < 0.05$ considered as potential significant causal evidence, and $P_{FDR} > 0.05$ & $P < 0.05$ as suggestive causal evidence.

Based on these analytical approaches, the study evaluates the levels of evidence for positive results. Findings that surpass the significance threshold, remain significant after meta-analysis, exhibit a PPH4 >80%, achieve a statistical power >80%, are not driven by a single SNP, and show no signs of horizontal pleiotropy or heterogeneity are considered significant causal evidence. Any result that does not meet one of these criteria is considered suggestive causal evidence.

Data analysis software and packages

All statistical analyses were performed using R software (version 4.2.2), TwoSampleMR (version 0.5.6), coloc (version 5.2.2), locuscomparer (version 1.0.0), MR-PRESSO (version 1.0), mr.raps (version 0.2), MRcML (version 0.9), meta (version 7.0) were analyzed.

RESULTS

Selection of genetic instruments and summary of results

Within a stringent IV selection framework, the number of SNPs used in the analysis varied from one to ten, each with an F-statistic greater than 10, effectively reducing bias from weak instrumental variables. The Steiger test verified the correct directionality as "TRUE" for all SNPs. This study ensured that all SNPs met the three fundamental assumptions of MR, thus maximizing the validity of the derived causal effects. Detailed information on each SNP within both the discovery and replication datasets is available in Tables S1 and S2. Initially, 27 causal associations were identified in the discovery dataset ($P < 0.05$). Upon merging with the replication dataset and conducting a meta-analysis, seven phenotypes were found to lose significance ($P > 0.05$) (Table S3). Further adjustments using the FDR identified seven potentially significant causal associations out of the remaining 20 results (Tables S4 and S5). After extensive filtering for PPH4, statistical power, pleiotropy, and heterogeneity, four significant and 16 suggestive causal associations were determined (Table 1). Comprehensive results of all MR analyses are documented in Tables S6 and S7, all sensitivity analyses are summarized in Table S8, colocalization analysis results are compiled in Table S9, and graphical summaries are presented in Figure S1.

Causal evidence from mendelian randomization and Bayesian colocalization analysis

This study rigorously assessed the obtained causal evidence, ultimately identifying four strong causal associations (Table 2). Specifically, genetically predicted increases in the levels of Phosphatidylcholine (PC) (16:0_16:1) [Odds ratio $(OR)_{Wald\ ratio} = 2.638$, 95% confidence interval (CI) 1.557–4.469, $P = 3.11 \times 10^{-4}$, $P_{FDR} = 0.002$], PC (16:1_18:0) ($OR_{Wald\ ratio} = 2.644$, 95% CI 1.559–4.486, $P = 3.11 \times 10^{-4}$, $P_{FDR} = 0.002$), triacylglycerol (TAG) (46:2) ($OR_{Wald\ ratio} = 2.515$, 95% CI 1.524–4.153, $P = 3.11 \times 10^{-4}$, $P_{FDR} = 0.002$), and TAG (48:2) ($OR_{IVW} = 1.863$, 95% CI 1.300–2.669, $P = 6.95 \times 10^{-4}$, $P_{FDR} = 0.046$) were significantly associated with an increased risk of developing MAFLD. Each analysis was corroborated by supplementary methods—cML, RAPS, dIVW, and BWMR—providing consistent causal evidence with the primary method (Figure 2). Table 3 provides detailed information on each SNP

Table 1 | Summary of causal evidence in Mendelian randomization analyses for discovery datasets with final screening completed

Plasma lipidome levels	SNPs	F_{mean} [†]	R^2 [‡] (%)	Power (%)	Method	OR (95% CI)	*P* value	P_{FDR}	$P_{pleiotropy}$	$P_{heterogeneity}$	$P_{MR\text{-}PRESSO}$	PPH3 (%)	PPH4 (%)
Sterol ester (27:1/14:0)	1	44	0.62	72	Wald ratio	0.359 (0.214, 0.603)	1.09E-04	0.002	NA	NA	NA	0.49	95.66
DAG (16:1_18:1)	2	39	1.26	100	IVW	1.847 (1.021, 3.344)	0.043	0.244	NA	0.088	NA	2.59	3.84
PC (16:0_16:1)	1	38	0.53	100	Wald ratio	2.638 (1.557, 4.469)	3.11E-04	0.002	NA	NA	NA	1.16	87.64
PC (16:1_18:0)	1	38	0.53	100	Wald ratio	2.644 (1.559, 4.486)	3.11E-04	0.002	NA	NA	NA	1.13	87.95
TAG (46:2)	1	32	0.59	100	Wald ratio	2.515 (1.524, 4.153)	3.11E-04	0.002	NA	NA	NA	1.13	87.91
TAG (48:1)	2	38	1.07	100	IVW	2.022 (1.295, 3.157)	1.94E-03	0.069	NA	0.236	NA	1.04	88.84
TAG (48:2)	2	49	1.39	100	IVW	1.863 (1.300, 2.669)	6.95E-04	0.046	NA	0.276	NA	1.07	88.60
TAG (50:1)	3	41	1.70	99	IVW	1.685 (1.101, 2.581)	0.016	0.157	0.812	0.131	NA	2.57	4.42
TAG (50:2)	2	63	1.75	100	IVW	1.704 (1.087, 2.671)	0.020	0.156	NA	0.127	NA	1.09	88.42
TAG (50:3)	2	84	2.31	100	IVW	1.587 (1.068, 2.359)	0.022	0.163	NA	0.122	NA	2.59	4.43
TAG (51:3)	5	55	3.83	100	IVW	1.470 (1.150, 1.880)	0.002	0.069	0.334	0.193	0.282	2.58	4.43
TAG (52:2)	4	56	3.11	100	IVW	1.487 (1.070, 2.066)	0.018	0.157	0.288	0.085	0.227	2.58	4.43
TAG (52:3)	6	64	5.30	98	IVW	1.356 (1.088, 1.689)	0.007	0.099	0.480	0.135	0.214	2.58	4.43
TAG (52:4)	6	60	4.92	100	IVW	1.416 (1.158, 1.730)	6.87E-04	0.046	0.334	0.259	0.298	2.58	4.43
TAG (52:5)	3	77	3.20	100	IVW	1.547 (1.140, 2.099)	0.005	0.097	0.382	0.142	NA	2.58	4.43
TAG (53:3)	4	63	3.51	72	IVW	1.269 (1.029, 1.563)	0.026	0.167	0.551	0.604	0.688	2.58	4.43
TAG (53:4)	4	62	3.76	100	IVW	1.477 (1.119, 1.950)	0.006	0.097	0.275	0.129	0.294	2.58	4.43
TAG (54:3)	3	84	3.45	76	IVW	1.285 (1.040, 1.587)	0.020	0.156	0.543	0.437	NA	2.57	4.43
TAG (54:4)	4	69	3.80	82	IVW	1.292 (1.056, 1.580)	0.013	0.155	0.557	0.620	0.624	2.58	4.43
TAG (56:7)	2	65	1.79	83	IVW	1.430 (1.065, 1.919)	0.017	0.157	0.277	0.321	NA	1.87	1.54

The bolded type represents the four significant causal evidences obtained. CI, confidence interval; DAG, diacylglycerol; FDR, false discovery rate; IVW, inverse-variance weighted; NA, not available; OR, odds ratio; PC, phosphatidylcholine; PPH3, posterior probability for hypothesis 3; PPH4, posterior probability for hypothesis 4; SNP, single nucleotide polymorphism; TAG, triacylglycerol. [†] $F = ((n - k - 1)/k)(R^2/(1 - R^2))$, where R^2 represents the proportion of exposure variance explained by the SNPs, *N* denotes the GWAS sample size, and $k = 1$ reflecting the individual SNP analysis. [‡] R^2 (the proportion of explained variance) was calculated employing the formula 2 × MAF × (1-MAF) × beta^2, wherein MAF represents the minor allele frequency of each specified SNP.

involved in the significant causal evidence. Additionally, the study achieved 100% statistical power to detect these associations, with no detected heterogeneity, pleiotropy, or single SNP-driven results. Furthermore, colocalization analysis indicated that PPH4 was greater than 80%, reinforcing the causal evidence in MR analysis. Notably, all four causal evidences suggested that rs780093 is a shared causal variant between the exposure and outcome (Figure 3).

The study identified 16 suggestive evidences (Table 1), specifically indicating that genetically predicted increases in Sterol ester (27:1/14:0) ($OR_{Wald\ ratio} = 0.359$, 95% CI 0.214–0.603, $P = 1.09 \times 10^{-4}$, $P_{FDR} = 0.002$) levels are associated with a reduced risk of MAFLD, while increases in Diacylglycerol (DAG) (16:1_18:1) ($OR_{IVW} = 1.847$, 95% CI 1.021–3.344, $P = 0.043$, $P_{FDR} = 0.244$) and various types of TAG (OR >1) —TAG (48:1), TAG (50:1), TAG (50:2), TAG (50:3), TAG (51:3), TAG (52:2), TAG (52:3), TAG (52:4), TAG (52:5), TAG (53:3), TAG (53:4), TAG (54:3), TAG (54:4), and TAG (56:7) —suggest an increased risk. It is important to note that although both Sterol ester (27:1/14:0) and TAG (52:4) passed the FDR correction, the former only showed a statistical power of 72%, and the latter a PPH4 of just 4.43%, warranting their classification as suggestive evidences. Additionally, the leave-one-out analysis revealed that [TAG (50:1)-rs1061808], [TAG (52:2)-rs15285], [TAG (52:5)-rs964184], and [TAG (54:4)-rs10105606] were all driven by a single SNP (Figure S2). This finding, likely due to a limited number of SNPs used for analysis, did not exhibit heterogeneity or horizontal pleiotropy, thus also classified as suggestive evidence.

Further mendelian randomization analysis

This study conducted a reverse MR analysis based on significant causal evidence to investigate plasma lipid levels in the MAFLD population. The findings revealed no causal relationship between genetic susceptibility to MAFLD and the levels of PC (16:0_16:1) ($Beta_{IVW} = -0.002$, 95% CI −0.079 to 0.074,

Table 2 | Summary of causal evidence ratings entries

Exposure	*P* & $P_{FDR} < 0.05$	PPH4 > 80%	Leave-one-out analysis	Power > 80%	No horizontal pleiotropy	No heterogeneity	Levels of evidence
Sterol ester (27:1/14:0) levels	√	√	√	N	√	√	Suggestive
Diacylglycerol (16:1_18:1) levels	×	×	√	√	√	√	Suggestive
Phosphatidylcholine (16:0_16:1) levels	√	√	√	√	√	√	Significant
Phosphatidylcholine (16:1_18:0) levels	√	√	√	√	√	√	Significant
Triacylglycerol (46:2) levels	√	√	√	√	√	√	Significant
Triacylglycerol (48:1) levels	×	√	√	√	√	√	Suggestive
Triacylglycerol (48:2) levels	√	√	√	√	√	√	Significant
Triacylglycerol (50:1) levels	×	×	×	√	√	√	Suggestive
Triacylglycerol (50:2) levels	×	√	√	√	√	√	Suggestive
Triacylglycerol (50:3) levels	√	×	√	√	√	√	Suggestive
Triacylglycerol (51:3) levels	×	×	√	√	√	√	Suggestive
Triacylglycerol (52:2) levels	×	×	×	√	√	√	Suggestive
Triacylglycerol (52:3) levels	×	×	√	√	√	√	Suggestive
Triacylglycerol (52:4) levels	√	×	√	√	√	√	Suggestive
Triacylglycerol (52:5) levels	×	×	×	√	√	√	Suggestive
Triacylglycerol (53:3) levels	×	×	×	×	√	√	Suggestive
Triacylglycerol (53:4) levels	×	×	√	√	√	√	Suggestive
Triacylglycerol (54:3) levels	×	×	×	×	√	√	Suggestive
Triacylglycerol (54:4) levels	×	×	×	√	√	√	Suggestive
Triacylglycerol (56:7) levels	×	×	×	√	√	√	Suggestive

Evidence of significant causation is considered significant only when all entries are √. FDR, false discovery rate; PPH4, posterior probability for hypothesis 4.

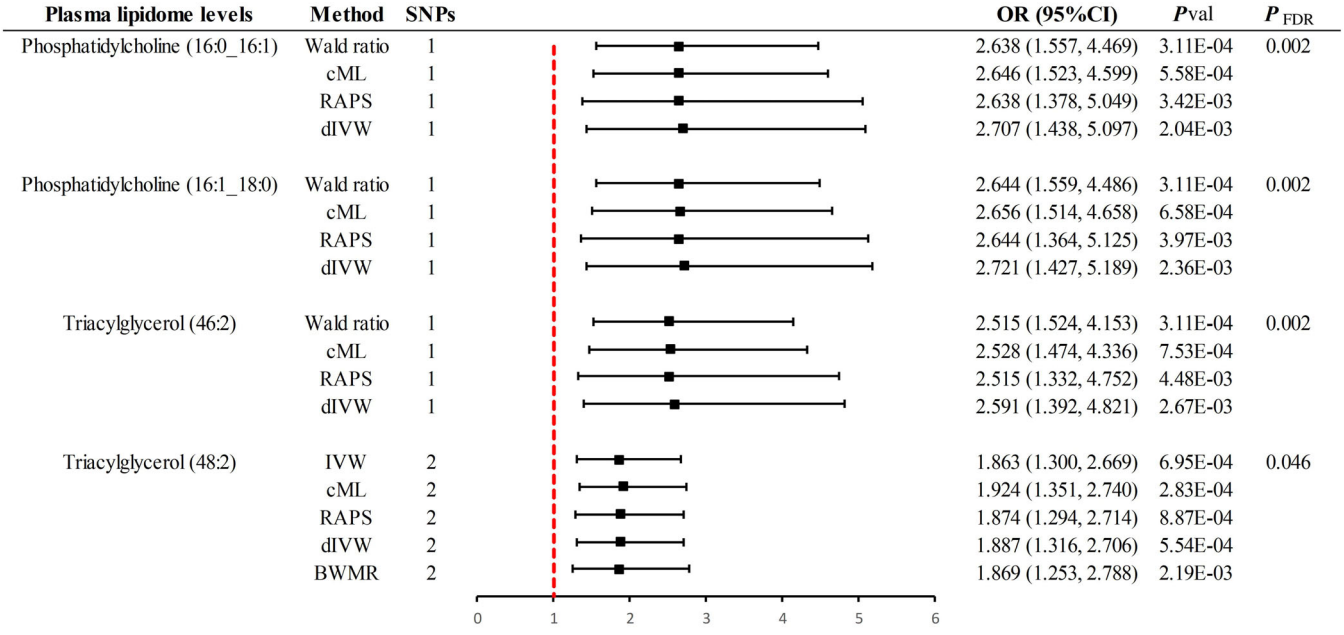

Figure 2 | Summary of results of significant causal evidence obtained from Mendelian randomization analysis. BWMR, Bayesian weighted Mendelian randomization; CI, confidence interval; cML, constrained maximum likelihood; dIVW, debiased inverse-variance weighted; FDR, false discovery rate; IVW, inverse-variance-weighted; OR, odds ratio; RAPS, robust adjusted profile score; SNP, single nucleotide polymorphism.

Table 3 | Summary of instrumental variables used for all significant causal association evidence

SNP	Effect Allele	Other Allele	Gene	Chr	Pos	EAF	*F*	R^2	Association with exposure		
									Beta	SE	*P* value
Phosphatidylcholine (16:0_16:1) levels											
rs1260326	C	T	GCKR	2	27,508,073	0.651	38	0.53%	−0.108	0.017	6.12E-10
Phosphatidylcholine (16:1_18:0) levels											
rs1260326	C	T	GCKR	2	27,508,073	0.651	38	0.53%	−0.108	0.018	3.75E-09
Triacylglycerol (46:2) levels											
rs1260326	C	T	GCKR	2	27,508,073	0.651	32	0.59%	−0.114	0.020	1.35E-08
Triacylglycerol (48:2) levels											
rs1260326	C	T	GCKR	2	27,508,073	0.651	62	0.86%	−0.138	0.018	4.46E-15
rs964184	C	G	ZPR1	11	116,778,201	0.849	37	0.52%	−0.143	0.023	9.60E-10

Chr, chromosome; Pos, position; SNP, single nucleotide polymorphisms.

P = 0.957), PC (16:1_18:0) ($Beta_{Wald\ ratio}$ = 0.025, 95% CI −0.063 to 0.114, *P* = 0.572), TAG (46:2) ($Beta_{IVW}$ = −0.066, 95% CI −0.157 to 0.026, *P* = 0.159), and TAG (48:2) ($Beta_{Wald\ ratio}$ = −0.075, 95% CI −0.160 to 0.010, *P* = 0.081) (Table S10).

We conducted an additional multivariable Mendelian randomization (MVMR) analysis to investigate the direct causal associations of four significant causal evidences on MAFLD after adjusting for risk factors. For this analysis, obesity traits such as body mass index (BMI) and waist-to-hip ratio adjusted for BMI (WHRadjBMI) were selected from the Genetic Investigation of ANthropometric Traits (GIANT) consortium[38], and type 2 diabetes (T2D) was adjusted for using data from the Diabetes Genetics Replication And Meta-analysis (DIAGRAM) consortium[39]. It is important to emphasize that based on prior evidence, conventional lipid phenotypes were not included in this analysis[13]. The results indicate that after adjusting for T2D, BMI, and WHRadjBMI, the four significant causal associations lost their significance with respect to MAFLD (Table 4).

DISCUSSION

This MR study goes beyond the scope of traditional lipid profiling by utilizing the latest plasma lipidome GWAS data to explore the causal relationship between hundreds of lipid species and MAFLD, ultimately identifying four significant causal associations. These associations include genetically predicted higher levels of PC (16:0_16:1), PC (16:1_18:0), TAG (46:2), and TAG (48:2) with an increased risk of MAFLD. Additionally, 16 suggestive evidences were found, indicating that higher levels of DAG (16:1_18:1), TAG (48:1, 50:1, 50:2, 50:3, 51:3, 52:2, 52:3, 52:4, 52:5, 53:3, 53:4, 54:3, 54:4, 56:7) increase the risk of MAFLD, while Sterol ester (27:1/14:0) appears to have a protective effect. These findings not only confirm the direct link between plasma lipid levels and the progression of MAFLD but also provide a scientific basis for further exploration of how to prevent or treat MAFLD by regulating these lipid species. After adjusting for confounding factors such as BMI, T2D, and WHRadjBMI, the significant causal associations previously observed were attenuated, suggesting that these factors may mediate the observed relationships between specific lipid profiles and MAFLD risk. Our will elaborate on the findings from multiple perspectives.

All significant causal evidence analyses involved rs1260326 (Table 3), a polymorphism in the glucokinase regulator protein kinase gene (GCKR), located on chromosome 2. This gene, consisting of 19 exons and 18 introns, encodes a protein that indirectly influences hepatic lipid accumulation by regulating key pathways in carbohydrate metabolism and fatty acid synthesis[40]. A meta-analysis incorporating 25 studies, including 6,598 cases and 19,954 controls, indicated that the polymorphic form rs1260326 of GCKR is associated with an increased risk of fatty liver disease. This association may augment hepatic lipid synthesis and storage by affecting insulin signaling and glucose metabolism within the liver[41]. Subsequent GWAS in our study's replication cohort further confirmed GCKR as a potential susceptibility gene for MAFLD[23]. Although the primary focus was on European populations, research by Wu *et al.*[42] demonstrated a significant association between rs1260326 variation and lean MAFLD in an elderly Han Chinese cohort, mediated by changes in waist circumference, suggesting these findings might also be applicable to East Asian populations. Future studies could expand the association between lipidomes and MAFLD to a multi-ancestry analysis. Additionally, research by Yuan *et al.* underscored the significant role of the GCKR gene in the development of MAFLD, highlighting the mediating role of TG in this relationship[43]. In the significant causal evidence of TAG (48:2), we also identified the importance of rs964184[ZPR1] (Table 3). Previous GWAS have emphasized that rs964184 [ZPR1] is associated with higher risks of elevated fasting TG[44] and postprandial plasma TG levels[45]. Esteve-Luque *et al.*[46] further confirmed that rs964184 [ZPR1] is independently associated with moderate/severe MAFLD. By proxying lipidomic phenotypes through these related genetic variants, this study offers new insights into the genetic foundations of MAFLD and underscores the

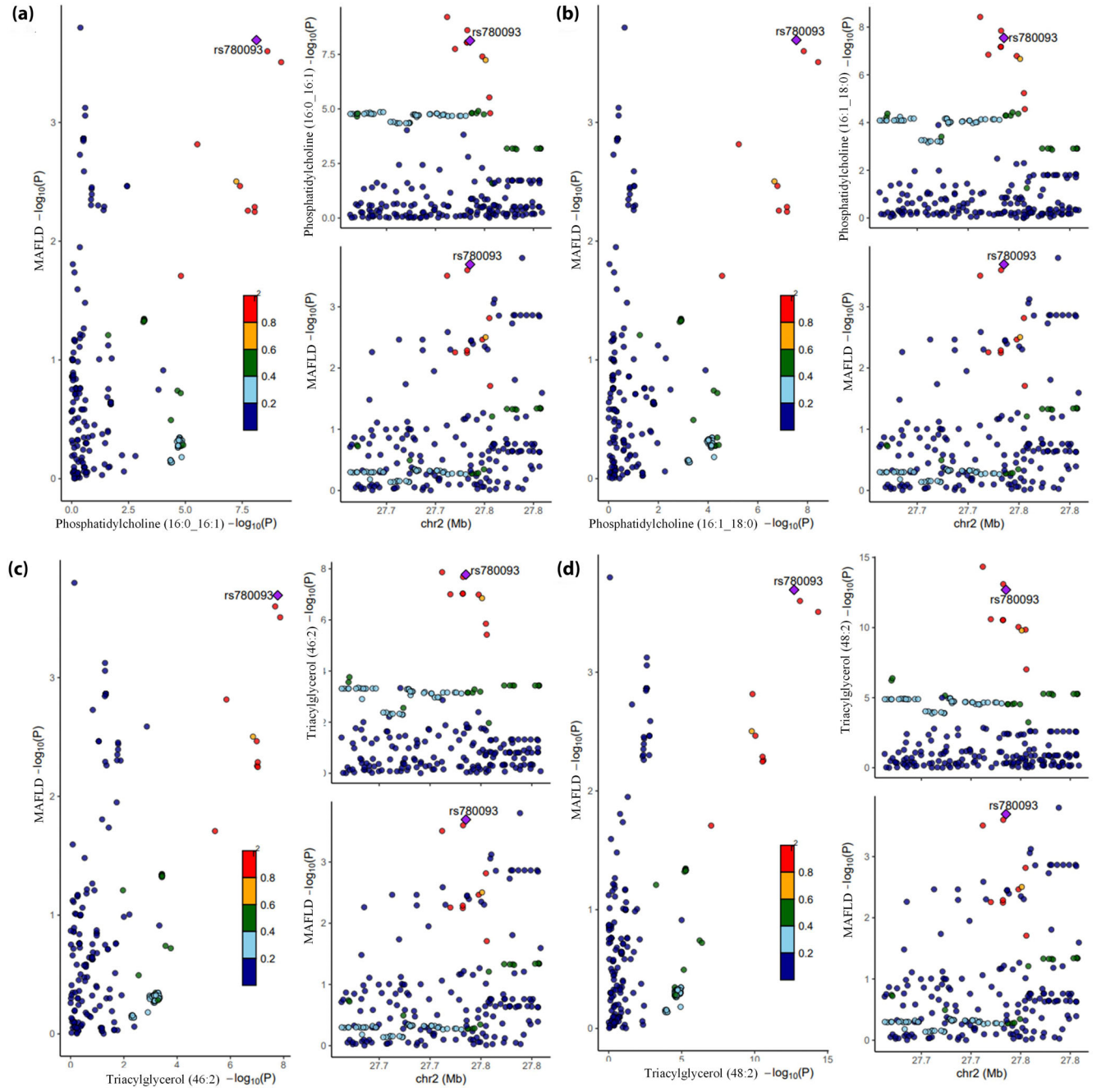

Figure 3 | Colocalization analysis of genetically proxied plasma lipidome and MAFLD (significant causal evidence). Points were color-coded according to the LD (r^2) of each variant relative to the variant with the highest posterior probability of colocalization within the gene region. In the left panel, $-\log_{10}$ *P* values for associations with plasma lipidome are on the *x*-axes, and $-\log_{10}$ *P* values for associations with the MAFLD on the *y*-axes. In the right panels, genomic positions are on the *x*-axes, and the *y*-axes show $-\log_{10}$ *P* values for plasma lipidome on the upper panel and $-\log_{10}$ *P* values with the MAFLD on the lower panel for the corresponding region. The genetic variants represented by the purple diamond-shaped squares in the figure are causal variants shared by the exposure and the ending. (a) Phosphatidylcholine (16:0_16:1) on MAFLD. (b) Phosphatidylcholine (16:1_18:0) on MAFLD. (c) Triacylglycerol (46:2) on MAFLD. (d) Triacylglycerol (48:2) on MAFLD.

importance of considering these genetic markers in clinical practice for more accurate MAFLD risk assessment and management.

This study revealed through colocalization analysis the pivotal role of the rs780093 variant in MAFLD. Like rs1260326, rs780093 is located on the GCKR gene, yet research on

Table 4 | Summary of analytical results for MVMR

MVMR models	Phenotypes	SNP	Beta	SE	*P* value	OR (95% CI)
Model 1	PC (16:0_16:1)	435	0.103	0.094	0.273	1.108 (0.922, 1.332)
	PC (16:1_18:0)	435	0.137	0.089	0.125	1.147 (0.963, 1.366)
	TAG (46:2)	434	0.151	0.078	0.053	1.163 (0.998, 1.355)
	TAG (48:2)	434	0.159	0.085	0.062	1.172 (0.992, 1.385)
Model 2	PC (16:0_16:1)	310	−0.059	0.098	0.546	0.942 (0.778, 1.142)
	PC (16:1_18:0)	310	−0.067	0.089	0.452	0.936 (0.786, 1.113)
	TAG (46:2)	309	−0.013	0.087	0.881	0.987 (0.832, 1.171)
	TAG (48:2)	310	0.053	0.090	0.554	1.055 (0.884, 1.259)
Model 3	PC (16:0_16:1)	505	0.079	0.071	0.264	1.082 (0.942, 1.243)
	PC (16:1_18:0)	503	0.028	0.065	0.663	1.029 (0.906, 1.168)
	TAG (46:2)	505	0.028	0.062	0.645	1.029 (0.912, 1.161)
	TAG (48:2)	505	0.083	0.066	0.208	1.086 (0.955, 1.236)

Model 1, direct causal effect of liposomes on MAFLD after adjusting for T2D; Model 2, direct causal effect of liposomes on MAFLD after adjusting for WHRadjBMI; Model 2, direct causal effect of liposomes on MAFLD after adjusting for BMI. CI, confidence interval; MVMR, multivariable Mendelian randomization; OR, odds ratio; PC, phosphatidylcholine; SNP, single nucleotide polymorphism; TAG, triacylglycerol.

rs780093 remains comparatively limited. Gu *et al.*[47] found a significant association between rs780093 and the spleen yang deficiency pattern (SYDP) in MAFLD through integrated traditional Chinese and Western medicine genetic research. Additionally, Li *et al.*[48] observed that the C allele frequency of rs780093 was significantly higher in obese patients with MAFLD than in lean patients, with both studies focusing on East Asian populations. In this MR study involving European ancestry, four significant causal evidences underscored the importance of rs780093 as a shared causal variant (Figure 3), consistent findings were also observed in suggestive associations with TAG (48:1) and TAG (50:2) (Figure S2). Despite not being selected as an instrumental variant in the two-sample MR analysis due to stringent selection criteria, the colocalization analysis results highlighted the significance of this locus, demonstrating the advantage of this method in identifying shared causal variants. Future research should explore the role of rs780093 across different ethnic and geographical backgrounds, as well as its interactions with other known risk factors, which will aid in developing targeted therapeutic strategies to improve management and treatment outcomes for patients with MAFLD.

It should be noted that the results of published MR studies exploring the causal links between conventional lipid phenotypes and MAFLD risk remain inconsistent and confusing. For example, Yuan *et al.*'s[14] findings suggested that the association between TG and MAFLD could be attributed to the "winner's curse" due to high sample overlap, while Xie *et al.*'s[15] investigation confirmed the association only within the FinnGen consortium, losing statistical significance when merging three datasets. Similarly, recent research by Li *et al.*[13] did not support a causal relationship between TG and MAFLD. These discrepancies may stem from limitations inherent in the summarized GWAS data relied upon, or be restricted to European ancestries. Although these data are extensive, they may lack detailed information at the specific lipid molecule level, failing to capture the nuanced relationships between specific lipid types and MAFLD. However, this study, by independently analyzing specific TAG molecules such as TAG (46:2) and TAG (48:2) rather than TG as a whole, successfully demonstrated significant causal relationships between these particular lipid types and MAFLD risk. This finding underscores the importance of exploring specific lipid molecules in research on metabolic-related diseases, aiding in the revelation of more accurate pathological mechanisms and potentially guiding the development of targeted therapeutic strategies. Moreover, this approach reduces the signal confounding and statistical noise common in traditional studies, rendering the results more stable and reliable. Future research should focus more on the role of specific lipid molecules, such as by adjusting dietary intake of certain types of fatty acids or developing medications targeting specific TAG metabolic pathways, which could help reduce hepatic TAG accumulation, thereby slowing or reversing the progression of MAFLD and further understanding their specific roles in the development of MAFLD.

To our knowledge, this is the first MR study to explore the causal relationship between multidimensional plasma lipidomes and the risk of MAFLD. The study boasts several advantages: it employs a rigorous analytical framework, supplemented with advanced analytical methods and stringent evidence evaluation, yielding robust causal evidence. Additionally, the inclusion of data from two sources in a meta-analysis reduces biases associated with reliance on a single sample cohort. However, the study also has limitations. As it is a secondary analysis of public data, the scope of analysis is restricted, and the population is confined to European ancestries, limiting the generalizability of the results. Furthermore, while the study relies on summarized GWAS data, which precludes subgroup analysis for variables such as sex and age, the data used have been adjusted for these variables through quality control, ensuring that the causal

evidence remains unaffected. Moreover, the use of summarized GWAS data limits the ability to incorporate longitudinal analyses to track lipidomic changes over time and their impact on MAFLD progression. This limitation underscores the necessity for future cohort-based MR studies with access to individual-level longitudinal data to further elucidate the temporal dynamics and causal pathways of lipid alterations in MAFLD development.

CONCLUSION

To encapsulate, this MR analysis, leveraging the latest plasma lipidomes GWAS data, explores the causal relationships between multidimensional lipids and MAFLD. It reveals significant causal associations between specific lipid molecules, such as PCs and various TAGs, and an increased risk of MAFLD. Additionally, it highlights the roles of rs1260326 and rs780093 within the GCKR gene in regulating these lipid molecules, providing new insights for future research into pathogenic mechanisms and the development of therapeutic strategies.

ETHICS

The study was a secondary analysis of publicly available data and therefore did not require ethical approval.

ACKNOWLEDGMENTS

We thank all GWAS participants and investigators for publicly making the summary statistics data available.

FUNDING

This study was supported by the Leading Talent Project of Jiangsu. Province Traditional Chinese Medicine No. SLJ0216, the Science and Technology Innovation Research Program of Jinling Hospital No. 2023JCYJZD078, and Nanjing Health Science and Technology Development Project No. ZKX21058.

DISCLOSURE

There are no competing interests in this study.
Approval of the research protocol: N/A.
Informed consent: N/A.
Registry and the registration no. of the study/trial: N/A.
Animal studies: N/A.

DATA AVAILABILITY STATEMENT

GWAS summary data for plasma lipidomes are available for download at https://www.ebi.ac.uk/gwas/publications/37907536 and MAFLD data from Ghodsian *et al.* are available at https://www.ebi.ac.uk/gwas/publications/ 34841290 online, FinnGen Consortium MAFLD data available for download at https://www.finngen.fi/en.

REFERENCES

1. Powell EE, Wong VW-S, Rinella M. Non-alcoholic fatty liver disease. *Lancet* 2021; 397: 2212–2224.
2. Eslam M, Newsome PN, Sarin SK, *et al*. A new definition for metabolic dysfunction-associated fatty liver disease: an international expert consensus statement. *J Hepatol* 2020; 73: 202–209.
3. Riazi K, Azhari H, Charette JH, *et al*. The prevalence and incidence of NAFLD worldwide: a systematic review and meta-analysis. *Lancet Gastroenterol Hepatol* 2022; 7: 851–861.
4. Ramai D, Facciorusso A, Vigandt E, *et al*. Progressive liver fibrosis in non-alcoholic fatty liver disease. *Cells* 2021; 10: 3401.
5. Lindén D, Romeo S. Therapeutic opportunities for the treatment of NASH with genetically validated targets. *J Hepatol* 2023; 79: 1056–1064.
6. Loomba R, Friedman SL, Shulman GI. Mechanisms and disease consequences of nonalcoholic fatty liver disease. *Cell* 2021; 184: 2537–2564.
7. Younossi Z, Anstee QM, Marietti M, *et al*. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. *Nat Rev Gastroenterol Hepatol* 2018; 15: 11–20.
8. Malhotra P, Gill RK, Saksena S, *et al*. Disturbances in cholesterol homeostasis and non-alcoholic fatty liver diseases. *Front Med (Lausanne)* 2020; 7: 467.
9. Jarvis H, Craig D, Barker R, *et al*. Metabolic risk factors and incident advanced liver disease in non-alcoholic fatty liver disease (NAFLD): a systematic review and meta-analysis of population-based observational studies. *PLoS Med* 2020; 17: e1003100.
10. Deprince A, Haas JT, Staels B. Dysregulated lipid metabolism links NAFLD to cardiovascular disease. *Mol Metab* 2020; 42: 101092.
11. Murad MH, Asi N, Alsawas M, *et al*. New evidence pyramid. *Evid Based Med* 2016; 21: 125–127.
12. Tin A, Kottgen A. Mendelian randomization analysis as a tool to gain insights into causes of diseases: a primer. *J Am Soc Nephrol* 2021; 32: 2400–2407.
13. Li Z, Zhang B, Liu Q, *et al*. Genetic association of lipids and lipid-lowering drug target genes with non-alcoholic fatty liver disease. *EBioMedicine* 2023; 90: 104543.
14. Yuan S, Chen J, Li X, *et al*. Lifestyle and metabolic factors for nonalcoholic fatty liver disease: Mendelian randomization study. *Eur J Epidemiol* 2022; 37: 723–733.
15. Xie J, Huang H, Liu Z, *et al*. The associations between modifiable risk factors and nonalcoholic fatty liver disease: a comprehensive mendelian randomization study. *Hepatology* 2023; 77: 949–964.
16. Ottensmann L, Tabassum R, Ruotsalainen SE, *et al*. Genome-wide association analysis of plasma lipidome identifies 495 genetic associations. *Nat Commun* 2023; 14: 6934.
17. Skrivankova VW, Richmond RC, Woolf BAR, *et al*. Strengthening the reporting of observational studies in epidemiology using mendelian randomisation (STROBE-MR): explanation and elaboration. *BMJ* 2021; 375: n2233.

18. Lawlor DA, Harbord RM, Sterne JAC, *et al*. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stat Med* 2008; 27: 1133–1163.
19. Cai J, Li X, Wu S, *et al*. Assessing the causal association between human blood metabolites and the risk of epilepsy. *J Transl Med* 2022; 20: 437.
20. Hemani G, Tilling K, Davey SG. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet* 2017; 13: e1007081.
21. Burgess S, Thompson SG, CRP CHD Genetics Collaboration. Avoiding bias from weak instruments in mendelian randomization studies. *Int J Epidemiol* 2011; 40: 755–764.
22. Kurki MI, Karjalainen J, Palta P, *et al*. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* 2023; 613: 508–518.
23. Ghodsian N, Abner E, Emdin CA, *et al*. Electronic health record-based genome-wide meta-analysis provides insights on the genetic architecture of non-alcoholic fatty liver disease. *Cell Rep Med* 2021; 2: 100437.
24. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* 2013; 37: 658–665.
25. Bowden J, Davey Smith G, Haycock PC, *et al*. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol* 2016; 40: 304–314.
26. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015; 44: 512–525.
27. Xue H, Shen X, Pan W. Constrained maximum likelihood-based mendelian randomization robust to both correlated and uncorrelated pleiotropic effects. *Am J Hum Genet* 2021; 108: 1251–1269.
28. Zhao Q, Wang J, Hemani G, *et al*. Statistical inference in two-sample summary-data mendelian randomization using robust adjusted profile score. *Ann Stat* 2020; 48: 1742–1769.
29. Ye T, Shao J, Kang H. Debiased inverse-variance weighted estimator in two-sample summary-data mendelian randomization. *Ann Stat* 2021; 49: 2079–2100.
30. Zhao J, Ming J, Hu X, *et al*. Bayesian weighted mendelian randomization for causal inference based on summary statistics. *Bioinformatics* 2020; 36: 1501–1508.
31. Burgess S, Foley CN, Allara E, *et al*. A robust and efficient method for mendelian randomization with hundreds of genetic variants. *Nat Commun* 2020; 11: 376.
32. Giambartolomei C, Vukcevic D, Schadt EE, *et al*. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* 2014; 10: e1004383.
33. Kulinskaya E, Dollinger MB, Bjørkestøl K. On the moments of Cochran's Q statistic under the null hypothesis, with application to the meta-analysis of risk difference. *Res Synth Methods* 2020; 11: 920.
34. Burgess S, Thompson SG. Interpreting findings from mendelian randomization using the MR-Egger method. *Eur J Epidemiol* 2017; 32: 377–389.
35. Verbanck M, Chen C-Y, Neale B, *et al*. Detection of widespread horizontal pleiotropy in causal relationships inferred from mendelian randomization between complex traits and diseases. *Nat Genet* 2018; 50: 693–698.
36. Cheng H, Garrick DJ, Fernando RL. Efficient strategies for leave-one-out cross validation for genomic best linear unbiased prediction. *J Anim Sci Biotechnol* 2017; 8: 38.
37. Brion M-JA, Shakhbazov K, Visscher PM. Calculating statistical power in mendelian randomization studies. *Int J Epidemiol* 2013; 42: 1497–1501.
38. Pulit SL, Stoneman C, Morris AP, *et al*. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum Mol Genet* 2019; 28: 166–174.
39. Suzuki K, Hatzikotoulas K, Southam L, *et al*. Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. *Nature* 2024; 627: 347–357.
40. Zhang Z, Ji G, Li M. Glucokinase regulatory protein: a balancing act between glucose and lipid metabolism in NAFLD. *Front Endocrinol* 2023; 14: 1247611.
41. Li J, Zhao Y, Zhang H, *et al*. Contribution of Rs780094 and Rs1260326 polymorphisms in GCKR gene to non-alcoholic fatty liver disease: a meta-analysis involving 26,552 participants. *Endocr Metab Immune Disord Drug Targets* 2021; 21: 1696–1708.
42. Wu N, Li J, Zhang J, *et al*. Waist circumference mediates the association between rs1260326 in GCKR gene and the odds of lean NAFLD. *Sci Rep* 2023; 13: 6488.
43. Yuan F, Gu Z, Bi Y, *et al*. The association between rs1260326 with the risk of NAFLD and the mediation effect of triglyceride on NAFLD in the elderly Chinese Han population. *Aging (Albany NY)* 2022; 14: 2736–2747.
44. Teslovich TM, Musunuru K, Smith AV, *et al*. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010; 466: 707–713.
45. Wojczynski MK, Parnell LD, Pollin TI, *et al*. Genome-wide association study of triglyceride response to a high-fat meal among participants of the NHLBI genetics of lipid lowering drugs and diet network (GOLDN). *Metabolism* 2015; 64: 1359–1371.
46. Esteve-Luque V, Padró-Miquel A, Fanlo-Maresma M, *et al*. Implication between genetic variants from APOA5 and ZPR1 and NAFLD severity in patients with hypertriglyceridemia. *Nutrients* 2021; 13: 552.
47. Gu Z, Wang Q, He H-Y, *et al*. Genetic variations associated with spleen-yang deficiency pattern of non-alcoholic fatty liver disease: a candidate gene study. *Eur J Integr Med* 2020; 33: 101044.
48. Li J, Wu N, Yang Y, *et al*. Unique genetic variants of lean nonalcoholic fatty liver disease: a retrospective cohort study. *BMC Endocr Disord* 2023; 23: 11.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1 | Colocalization analysis of genetically proxied plasma lipidome and NAFLD (positive results).

Figure S2 | Summary of leave-one-out analysis among positive results (for IV ≥3).

Table S1 | Summary of SNP-specific information for all SNPs in the discovery dataset.

Table S2 | Summary of SNP-specific information for all SNPs in the replication dataset.

Table S3 | Summary of meta-analysis results for replication and discovery datasets.

Table S4 | Focuses on the FDR correction results among the 27 positive results obtained from the discovery dataset.

Table S5 | Summary of all results in the discovery dataset after correction for the primary analytical method using FDR.

Table S6 | Summary of results of Mendelian randomization analysis (discovery dataset).

Table S7 | Summary of results of Mendelian randomization analysis (replication dataset).

Table S8 | Sensitivity analysis summary in the discovery dataset.

Table S9 | Summary of Bayesian colocalization results.

Table S10 | Reverse Mendelian randomization analysis.

Appendix S1 | STROBE-MR checklist of recommended items to address in reports of Mendelian randomization studies.