Lipids in Health and Disease

RESEARCH

Open Access

Druggable genome-wide Mendelian randomization identifies therapeutic targets for metabolic dysfunction-associated steatotic liver disease

Xiaohui Ma[1,4†], Li Ding[1†], Shuo Li[1†], Yu Fan[1], Xin Wang[1], Yitong Han[2], Hengjie Yuan[3*], Longhao Sun[2*], Qing He[1*] and Ming Liu[1]

Abstract

Background Metabolic dysfunction-associated steatotic liver disease (MASLD) affects > 25% of the global population, potentially leading to severe hepatic and extrahepatic complications, including metabolic dysfunction-associated steatohepatitis. Given that the pathophysiology of MASLD is incompletely understood, identifying therapeutic targets and optimizing treatment strategies are crucial for addressing this severe condition.

Methods Mendelian randomization (MR) analysis was conducted using two genome-wide association study datasets: a European meta-analysis (8,434 cases; 770,180 controls) and an additional study (3,954 cases; 355,942 controls), identifying therapeutic targets for MASLD. Of 4302 drug-target genes, 2,664 genetic instrument variables were derived from cis-expression quantitative trait loci (cis-eQTLs). Colocalization analyses assessed shared causal variants between MASLD-associated single nucleotide polymorphisms and eQTLs. Using the drug target gene cis-eQTL of liver tissue from the genotype-tissue expression project, we performed MR and summary MR to validate the significance of the gene results of the blood eQTL MR. RNA-sequencing data from liver biopsies were validated using immunohistochemistry and quantitative polymerase chain reaction (qPCR) tests to confirm gene expression findings.

Result MR analysis across both datasets identified significant MR associations between MASLD and two drug targets—milk fat globule-EGF factor 8 (*MFGE8*) (odds ratio [OR] 0.89, 95% confidence interval [CI] 0.85–0.94; $P = 2.15 \times 10^{-6}$) and cluster of differentiation 33 (*CD33*) (OR 1.17, 95% CI 1.10–1.25; $P = 1.39 \times 10^{-6}$). Both targets exhibited strong colocalization with MASLD. Genetic manipulation indicating *MFGE8* activation and *CD33* inhibition

†Xiaohui Ma, Li Ding and Shuo Li contributed equally to this work and share first authorship.

*Correspondence:
Hengjie Yuan
hengjieyuan@163.com
Longhao Sun
sun_longhao@163.com
Qing He
heqing202108@163.com

Full list of author information is available at the end of the article

did not increase the risk for other metabolic disorders. RNA-sequencing, qPCR, and immunohistochemistry validation demonstrated consistent differential expressions of *MFGE8* and *CD33* in MASLD.

Conclusion *CD33* inhibition can reduce MASLD risk, while *MFGE8* activation may offer therapeutic benefits for MASLD treatment.

Keywords *CD33*, Mendelian randomization analysis, MFGE8, Non-alcoholic fatty liver disease, Nonalcoholic steatohepatitis

Introduction

With the increasing prevalence of obesity, metabolic dysfunction-associated steatotic liver disease (MASLD) has become a major global health concern, affecting approximately 25% of adults [1]. MASLD is characterized by excessive triglyceride accumulation in the liver alongside at least one cardiometabolic risk factor. Some studies suggest that evidence from non-alcoholic fatty liver disease (NAFLD) research is applicable to MASLD, and the terms NAFLD and MASLD can be used interchangeably [2–6]. Therefore, this study adhered to these classification principles in data extraction and concept descriptions. MASLD involves a range of histologically classified pathologies, including simple steatosis, metabolic dysfunction-associated steatohepatitis (MASH), and cirrhosis [7]. MASH, a late-stage form of MASLD, is expected to become a predominant indication for liver transplantation. Despite its severity, no approved drug therapies for MASLD exist, and the few investigated therapies have shown limited efficacy [8, 9]. Advancing research on the cellular and molecular pathogenesis of MASLD is essential for identifying therapeutic targets and developing targeted treatments.

Genetic variants associated with MASLD and MASH have been identified through genome-wide association studies (GWAS) and candidate gene approaches [10]. These findings further highlight the important role of genetic factors in the risk of developing MASLD and in disease progression [11, 12]. With the increasing popularity of GWAS, Mendelian randomization (MR) has been shown to be effective in investigating the etiology of this disease and in predicting the efficacy of clinical medications through simulating randomized controlled trials [13]. In MR analyses of drug target genes, the use of mRNA, proteins, or other downstream biomarkers as exposure factors is a robust analytical approach that is more closely aligned with the causal pathway than other exposure phenotypes. Commonly, cis-expression quantitative trait loci (cis-eQTL) within the genomic regions of drug target genes are employed as proxies, which are regulators that influence gene expression [14].

This study clarified mechanisms underlying MASLD and identified potential therapeutic targets for drug development by examining the association between protein-coding genes and MASLD. Using two independent MASLD GWAS datasets, the eQTL identified in blood were combined. MR analyses were conducted to identify potential drug targets that might slow MASLD progression. The associations between representative druggable genes and MASLD risk were examined, along with the associations between these genes and 16 metabolism-related disease risk factors and seven common disease characteristics. Although several similar studies have been conducted, this study offers advantages owing to its large genome-wide association study (GWAS) sample size, broad ethnic representation, and validation of population samples through RNA sequencing [15–17].

Methods

This study was a secondary analysis of publicly available data. Informed consent was obtained from all participants following the original GWAS protocol, and all ethical approvals for GWAS were obtained from the original study authors. The experimental verification part of our study was approved by the Institutional Review Board of Tianjin Medical University General Hospital (Tianjin, China) and conducted following the Declaration of Helsinki norms.

Identification of cis-eQTL data associated with druggable genes (cis-eQTL of blood and liver tissue from genotype-tissue expression [GTEx])

A total of 4,302 druggable genes, annotated using the HUGO Gene Nomenclature Committee (HGNC) database, were identified on autosomal chromosomes [18]. These included 1,375 protein therapy targets in clinical development, 2,281 proteins related to members of key drug target families, and 646 proteins associated with drug–target interactions. Given that cis-eQTLs are more closely associated with genes relevant to drug discovery, statistically significant cis-eQTLs were obtained from the eQTLGen Consortium to minimize bias (false discovery rate < 0.05, 1 Mb per probe). Consequently, 31,684 human peripheral blood eQTLs were analyzed [19]. A total of 4,302 genetic tools for druggable targets were generated through selecting cis-eQTLs within 100 kb of the genomic location of each gene, ultimately identifying eQTLs for 2,664 druggable genes. Using the same method, the drug target gene cis-eQTLs of liver tissue were also obtained from GTEx v10.0 (https://gtexportal.

org/home/releaseInfoPage) to validate the significance of the gene results after the blood eQTL MR analysis.

Outcome data

A genome-wide meta-analysis was conducted on four cohorts (United Kingdom [UK] Biobank, Estonian Biobank, Electronic Medical Records and Genomics (eMERGE), and (Finnish Genetic) FinnGen [20]) of participants of European ancestry with electronic health records for MASLD. The analysis included the eMERGE and FinnGen cohorts, an updated MASLD GWAS from the UK Biobank (2558 cases; 395,241 controls), and a newly performed GWAS from the Estonia Biobank (4119 cases; 190120 controls). In total, 8,434 MASLD cases and 770,180 controls were included in these cohorts [20].

A separate sample cohort was used for external validation [21] (Table S1). The study was a genomics-specific case-control study in the UK Biobank [21] with a definitive diagnosis of MASLD based on the diagnostic criteria recommended in recent consensus guidelines [22]. GWAS analyses were performed on 4,761 MASLD cases and 373,227 healthy controls without MASLD who were subjected to sensitivity analyses, excluding other secondary liver pathologies that coexisted, and adjusting for body mass index (BMI) and alcohol intake.

Metabolism-related diseases

Sixteen metabolism-related disease risk factors were selected, including lipid profiles (total cholesterol [TC], triglyceride [TG]), high-density lipoproteins [HDL-C], low-density lipoproteins [LDL-C], apolipoproteins A1 and B (ApoA1, ApoB), and lipoprotein a (Lp(a)) [23]. Additionally, blood pressure metrics, including systolic blood pressure, diastolic blood pressure, were included [24]. Three blood glucose indicators were considered: fasting glucose, fasting insulin, and glycated hemoglobin [25]. Four anthropometric characteristics were also analyzed (BMI, waist circumference, hip circumference, and waist-to-hip ratio) [26, 27]. Furthermore, seven metabolic syndrome-related conditions, such as essential hypertension [28], obesity [26], hyper-triglyceridemia [29], type 2 diabetes mellitus [30], myocardial infarction [28], and ischemic stroke, were included in the analysis [30]. The GWAS database, operated by the Medical Research Council–Integrative Epidemiology Unit (MRC-IEU) consortium, was used as the source of initial exposure data for the Atherosclerotic Heart Disease Study (Table S2) [31].

MR and colocalization (cis-eQTL of blood and liver tissue from GTEx)

A two-sample MR R package was employed to analyze the MR data. A series of criteria were applied to exclude low-quality instrumental variables [32]. First, weak single nucleotide polymorphisms (SNPs) were excluded by calculating the F-statistic (F-statistic < 10). Subsequently, SNPs that did not exhibit linkage disequilibrium with independent conditions ($r^2 < 0.1$, based on the 1000 Genomes Europe reference panel) were selected as instrumental variables. Additionally, genes with greater MASLD trait exposure were excluded through Steiger filtering, a heterogeneity filtering method to filter SNPs whose coefficient of determination (R^2) (R-squared, a statistical measure that represents the proportion of the variance for a dependent variable that is explained by an independent variable) with the exposure is not substantially greater than with the outcome. In the primary analyses, MR estimates for each SNP were calculated using the Wald ratio method. The SNP estimates were meta-analyzed using inverse variance weighting (IVW), MR-Egger, and weighted median models [33]. MR-Egger regression introduces a nuisance parameter to account for directional pleiotropy, assesses whether genetic variants exhibit pleiotropic effects on the outcome, and compares the observed distance of all the variants to the regression line with the expected distance under the null hypothesis of no horizontal pleiotropy [34]. Bonferroni correction was used to assess significance thresholds for multiple test exposures. A P-value $< 1.90 \times 10^5$ ($P = 0.05/2644$) was defined as a statistically significant difference, and validation of these significantly different targets was subsequently repeated in another independent cohort. Using the same MR process, drug target gene cis-eQTLs of liver tissue from GTEx were also used to validate the significance of the gene results after the blood eQTLs MR analysis.

Significant MR results common to both independent cohorts with MASLD risk were identified using a coloc R package. For the eQTL dataset, a priori probabilities of 1×10^4 for cis-eQTL (H1) and MASLD association were assumed, along with 1×10^5 for a single variant affecting both traits (H4) [35]. Genes with significant co-localization (*a posteriori* probability PH4 > 0.80) were considered as potential drug therapy targets. Subsequently, the association between these targets and 23 metabolic syndrome-related traits was investigated using MR analysis to assess potential safety concerns and alternative indications.

Summary data-based MR (SMR)

The primary analysis applied involved MR analysis of blood eQTL and diseases. Given that MASLD is liver-specific, prior validation was conducted using liver eQTLs from GTEx in addition to an MR analysis of the disease. To ensure the robustness of study results, a dual SMR analysis was conducted with liver eQTLs from two versions of the GTEx database (v10.0 and v8.0) in conjunction with disease outcomes.

Patients

From July 2023 to March 2024, patients with obesity who underwent sleeve gastrectomy at the Tianjin Medical University General Hospital received liver biopsies to assess steatosis, inflammation, hepatocyte ballooning, and fibrosis, using a semi-quantitative method of the nonalcoholic steatohepatitis activity score (NAS) [36]. Patients were categorized as having simple steatosis (NAS 1–2, without ballooning or fibrosis) or MASH (NAS ≥ 5 or NAS 3–4 with fibrosis). The prevalence rate for MASH within the cohort under investigation was 42.12%. Liver and blood samples were collected for RNA sequencing and stored at − 80 °C for subsequent analysis.

RNA sequencing and data analysis

Overall, 41 patients with obesity who underwent metabolic surgery with intraoperative liver biopsies were selected for RNA sequencing by Shanghai OE Biotech. RNA extraction and library preparation methods are detailed in Supplementary File S1. Libraries were sequenced on the BGI sequencing platform (BGI T7), generating 150 bp paired-end reads. Trimmomatic, a multithread command tool, removed adapters and low-quality reads, with clean reads mapped to genes using the program HISAT2 [37, 38]. Read counts per gene were obtained with HTSEQCOUNT [39], with fragments per kilobase per million mapped fragments (FPKM) values calculated using CUFFLINKS [40]. Differentially expressed genes (DEGs) between groups were determined using truncated P-values < 0.05 and fold change > 1.5 or < 0.67 using DESeq2 with P-values corrected for multiple testing using the corresponding R package function for Benjamini–Hochberg method. Gene expression patterns were analyzed using hierarchical clustering.

Histological analysis

Liver tissues were fixed with 4% paraformaldehyde, paraffin-embedded, and sectioned at 5 μm. Tissue sections were then stained with hematoxylin and eosin (HE) and Masson's stain and evaluated using an orthogonal light microscope (Nikon, NIKON ECLIPSE E100). Immunohistochemical staining was performed on at least six liver samples from each group to detect protein expression of *CD33* (Proteintech Group, 17425-1-AP) and *MFGE8* (Proteintech Group, 25951-1-AP). Results were interpreted under a white light microscope (Nikon Instruments Ltd. E100). Using the AIPATHWELL analysis software, the same tan nuclei were uniformly selected to identify positive cells in all sections, while blue nuclei were selected as other cells. Each section was analyzed to obtain the number of positive and total cells. The percentage of positive cells (number of positive cells/total number of cells * 100) was calculated as the positivity rate (%).

Quantitative polymerase chain reaction

Total RNA was isolated from liver using a mirVana RNA Isolation Kit (AM1561), and its yield and integrity were assessed using a NanoDrop 2000 spectrophotometer and ethidium bromide-stained agarose gel electrophoresis, respectively. RNA was converted to cDNA with the TransScript All-in-One First-Strand cDNA Synthesis SuperMIX for qPCR (quantitative polymerase chain reaction). qPCR testing was conducted on a LightCycler® 480 II using PerfectStart Green qPCR SuperMix, with amplification conditions consisting of 94 °C for 30 s, followed by 45 cycles of 94 °C for 5 s and 60 °C for 30 s. Product specificity was confirmed by melting curve analysis from 60 °C to 97 °C. mRNA expression levels were normalized to GAPDH and calculated with the $2^{-\Delta\Delta Ct}$ method. Primers were designed and synthesized by Shanghai Ouyi Biomedical Technology Co. and Beijing Kengke Xinye Biotechnology Co, respectively (Table S3).

Statistical analysis

Statistical data are represented as mean and standard deviation values. Group differences were assessed using one-way analysis of variance (ANOVA) followed by Tukey's post-hoc test for pairwise comparisons. When data did not meet the assumptions of parametric tests, a Kruskal–Wallis analysis test was employed to assess differences across more than two groups. P-values < 0.05 were considered statistically significant.

Results

Study design

Figure 1 shows the study process, while Table S2 outlines the data sources. Initially, we screened 4,302 human protein-coding genes to evaluate their potential as drug targets. Independent cis-eQTL variants of human blood were then selected conditionally, and two-sample MR was performed to assess the association between drug-target mRNA expression and MASLD risk. MR results were simultaneously adjusted multiple times and validated in a second cohort. Colocalization analyses were conducted to test whether the MR results were affected by different causal variants owing to chain imbalance. Additionally, the drug target gene cis-eQTL of liver tissue from GTEx was used to conduct MR and SMR analysis to validate the significance of gene results of the blood eQTL MR analysis. Safety and alternative indications for the final targets were also evaluated via MR. Finally, RNA-seq was performed on liver biopsy samples, with immunohistochemistry and qPCR experimental methods used to validate the results of RNA-seq and colocalization.

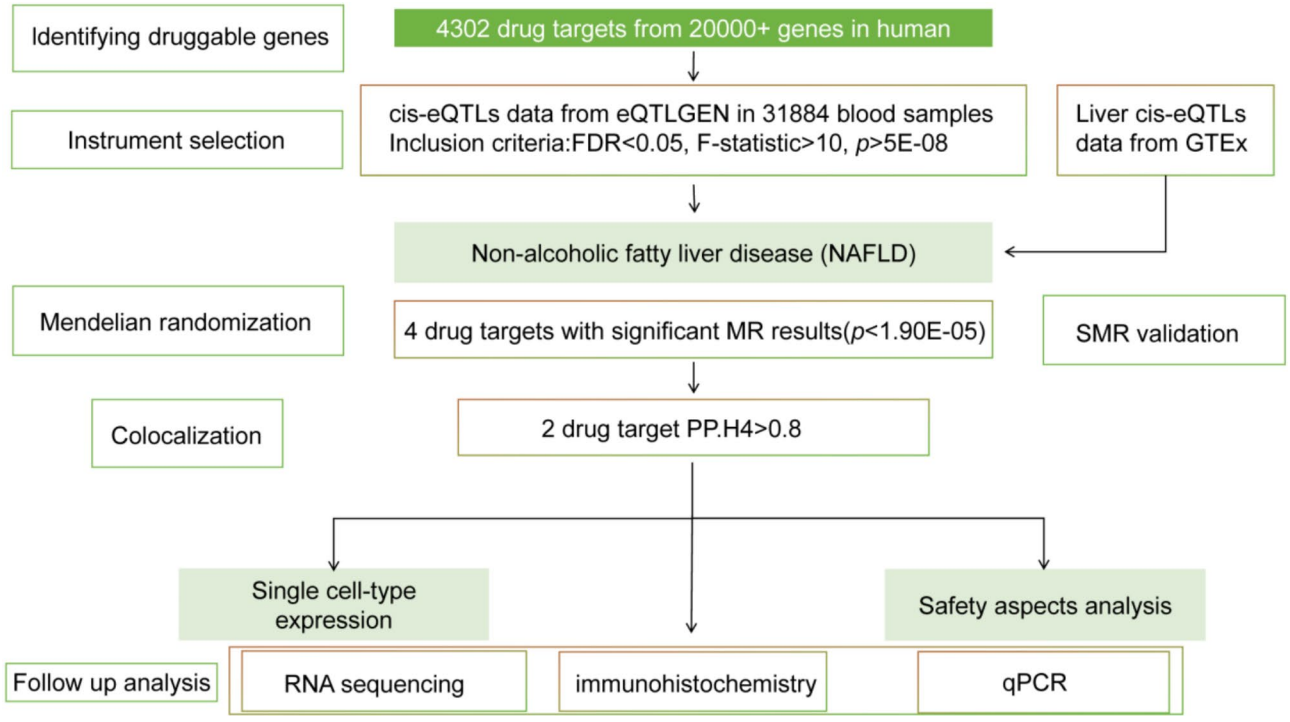

Fig. 1 Overview of the MR study design. MR, Mendelian randomization

Table 1 Mendelian randomization results

Genes	Discovery cohort					Replication cohort				
	SNPs	OR(95%CI)	IVW *p*-value	MR-Egger intercept	Egger intercept *p*-value	SNPs	OR(95% CI)	IVW *p*-value	MR-Egger intercept	Egger intercept *p*-value
CD33	18	1.17(1.10, 1.25)	1.39E-06	0.03	0.02	18	1.10(1.01–1.19)	0.02	0.06	0.45
MFGE8	24	0.89(0.85–0.94)	2.15E-06	-0.01	0.49	28	0.94(0.89–0.99)	0.04	-0.07	0.25
TEK	42	0.92 (0.89, 0.96)	1.90E-06	-0.02	0.02	45	0.98(0.94–1.02)	0.36	-3.23E-3	0.94
LILRA3	21	1.05 (1.03, 1.07)	2.03E-06	0.02	0.12	27	1.02(0.99–1.05)	0.16	0.03	0.20

MFGE8, Milk fat globule-EGF factor 8; TEK, Tie2 Endothelial Cell Kinase; LILRA3, Leukocyte Immunoglobulin-Like Receptor, Subfamily A (with TM and 2 Ig-Like Domains), Member 3; OR, odds ratio; CI, confidence interval; IVW, inverse variance weighted

Discovery analysis

A total of 2,644 druggable genes were identified by screening cis-eQTL data from the eQTLGen consortium. Subsequently, two-sample MR analyses were conducted on European MASLD patient data. In the discovery cohort, IVW meta-analysis was used in conjunction with effect estimates for each instrumental variable, with analysis performed on 8,434 patients and 770,180 controls from the meta-analysis of four cohorts. Ultimately, following a series of tests, it was determined that the predicted gene expression of four genes (*MFGE8*, *CD33*, *LILRA3*, and *TEK*) was associated with an increased risk of developing MASLD ($P < 1.90E^{-5}$ [IVW], 0.05 Bonferroni-corrected 2,644 drug targets, Table S4–8).

Replication analysis

Effect estimates for the four genes identified in the discovery phase were validated using the UK Biobank cohort ($N = 373,227$). The validation results for two of the targets (*MFGE8* and *CD33*) were statistically significant ($P < 0.05$ [IVW], Table 1, Table S9), while the direction of effect remained consistent.

Colocalization analysis

To minimize the risk of linkage disequilibrium (LD) confounding MR results, gene colocalization analyses were conducted to assess the probability of sharing causal genetic variation in SNPs associated with MASLD and eQTL. Two genes exhibited strong evidence of colocalization (PP.H4 > 0.8). Specifically, colocalization of *CD33* and *MFGE8* with MASLD suggests their potential as drug target genes. The results indicated that *MFGE8* and MASLD might share a causal variant within the *MFGE8* locus (PP.H4 = 0.85, Fig. 2), whereas *CD33* in blood was highlighted as a candidate gene for MASLD risk (PP.H4 = 0.93, Fig. 2). Consequently, MR and colocalization

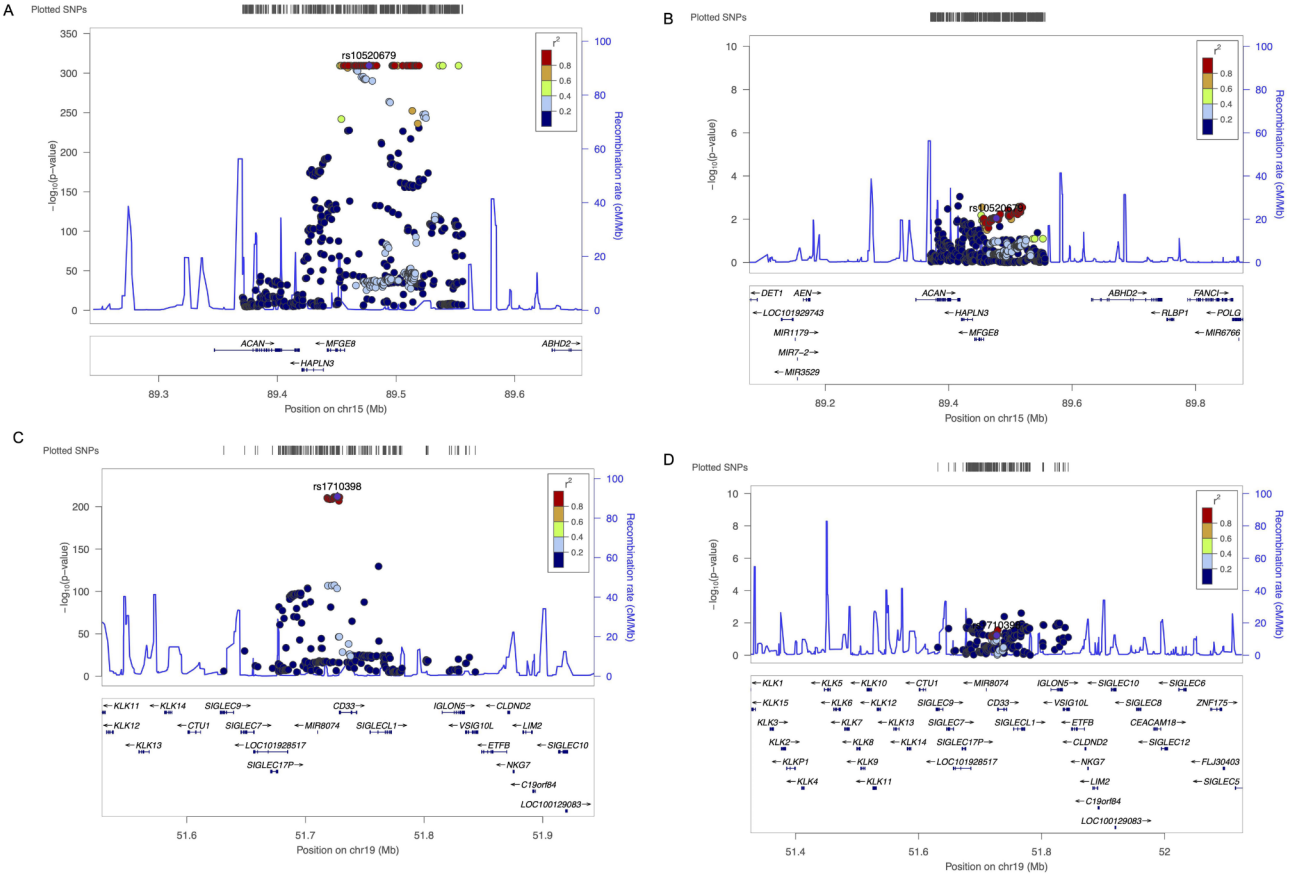

Fig. 2 Regional Manhattan plot of associations between SNPs and *MFGE8* and *CD33* locus. (**A**) rs10520679 was used to proxy serum *MFGE8* expression. (**B**) rs10520679 and its flanking 500 kb region to either side in MASLD. (**C**) rs1710398 was used to proxy serum *CD33* expression. (**D**) rs1710398 and its flanking 500 kb region to either side in MASLD. MASLD, metabolic dysfunction associated steatotic liver disease; MFGE8, milk fat globule-EGF factor 8; SNP, single-nucleotide polymorphisms

analyses identified two potential drug target genes, providing evidence of a shared genetic effect between eQTL and MASLD risk (Table S10).

Liver transcriptomics and pathological findings

RNA-seq generated a total of 285.42 G of clean data. The effective data volume for each sample ranged from 6.69 to 7.08 G, with Q30 base distribution between 95.85 and 97.27% and an average GC content of 45.35%. By aligning the reads to the reference genome, the genomic alignment for each sample was obtained, with an alignment rate of 98.82–99.12%. Based on the alignment results, expression of protein-coding gene were analyzed, and differential screening was conducted based on the expression of protein-coding genes in different samples, with a total of one differential grouping (Tables S11–12). Principal component analysis revealed different liver transcriptome profiles between patients in the non-MASH and MASH groups (Figure S1). Meanwhile, sample-sample distance clustering analysis showed that samples in the non-MASH group were more distinct from those in the MASH group (Fig. 3). A total of 1,422 genes were significantly differentially expressed in the non-MASH group compared with the MASH group, with 695 genes upregulated and 727 genes downregulated. H&E and Masson staining images showed significant macrovesicular steatosis, hepatocyte ballooning degeneration, and inflammatory infiltration in the livers of patients with MASH (Fig. 4).

CD33

The MR analysis revealed a significant correlation between increased *CD33* expression and a heightened risk of MASLD (odds ratio [OR] = 1.17, 95% confidence interval [CI]: 1.10, 1.25, $P = 1.39 \times 10^{-6}$). Consequently, antagonists targeting *CD33* are emerging as a promising therapeutic approach to mitigate the risk of MASLD. However, it is imperative to consider potential adverse effects and indications for use during the development of novel pharmacological agents. Consequently, an evaluation was performed to ascertain the causal relationship between *CD33* gene inhibitors and 16 potentially modifiable risk factors, in addition to seven metabolism-related diseases.

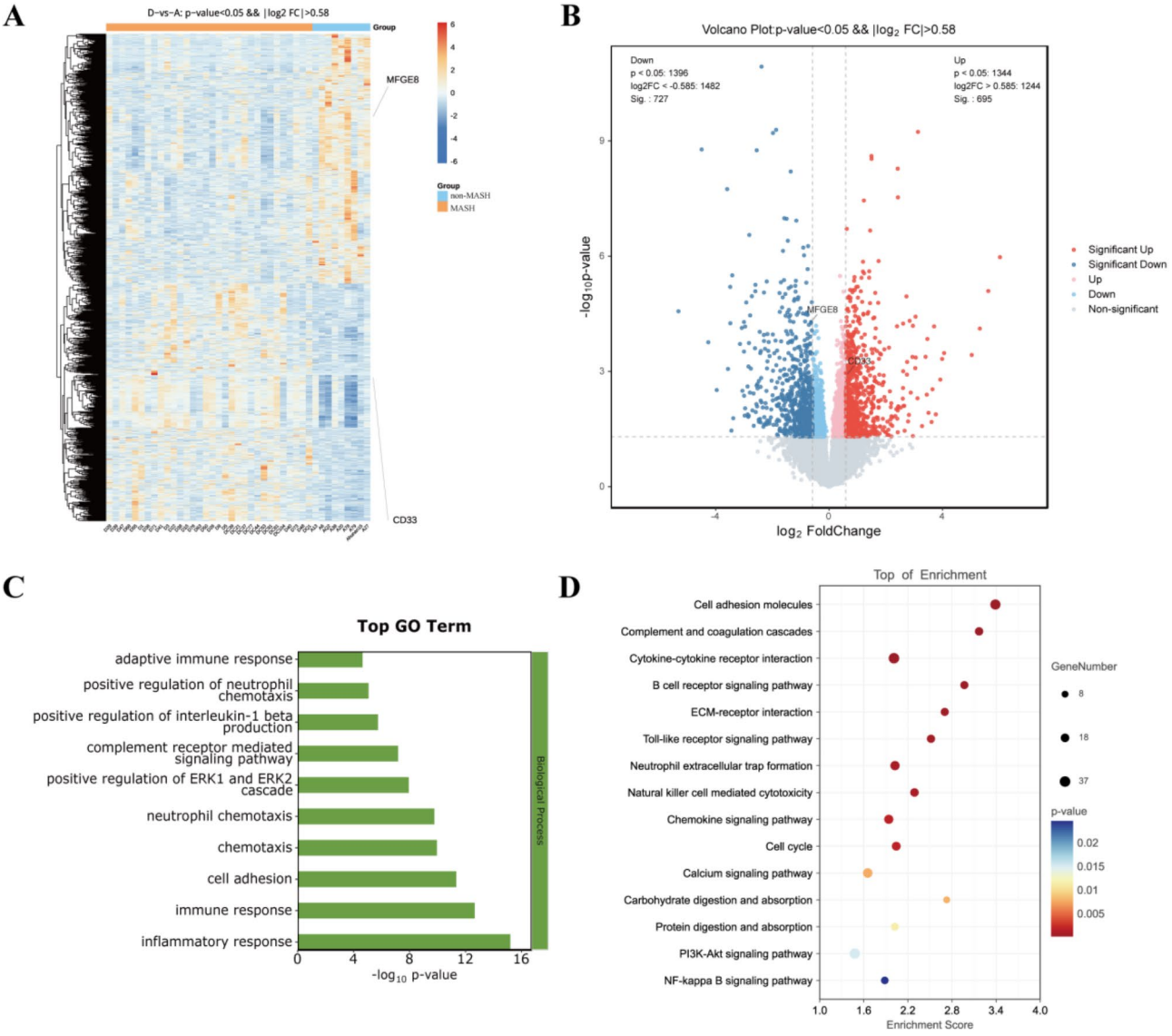

Fig. 3 Liver transcriptome profile. (**A**) Heatmap of the DEGs expression between MASH group and non-MASH group. (**B**) Volcano plot of differential expression. (**C**) Bar chart of GO enrichment analysis. (**D**) Bubble chart of KEGG pathway enrichment analysis. RNA sequencing was randomly selected for 9 samples in the non-MASH group and 32 samples in the MASH group. DEGs, differentially expressed genes; GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; MASH, metabolic dysfunction-associated steatohepatitis

The evidence was inconclusive regarding the association between proxied gene inhibitors of *CD33* and lipid metabolism markers, blood pressure, and glucose-related markers ($P > 2.17 \times 10^{-3}$ [IVW], 0.05/23 results, Fig. 5). However, weak correlations were observed between proxied gene inhibitors of *CD33* and Lp(a) ($P = 0.041$ [IVW]).

For metabolic syndrome-related diseases, genetically predicted *CD33* inhibition was significantly negatively associated with atherosclerotic heart disease (OR = 0.997, 95% CI: 0.996–0.999, $P = 1.67 \times 10^{-3}$ [IVW]) (Fig. 5). Weak correlations were presented with essential hypertension (OR = 0.999, 95% CI: 0.998–1.000, $P = 1.55 \times 10^{-2}$ [IVW]).

To explore the possible relationship between *CD33* and the risk of MASLD, gene and protein levels in liver tissues of patients with MASLD were examined. The results of liver transcriptomics showed that the mRNA of *CD33* in liver tissues of patients in the MASH group was significantly higher than those of patients in the non-MASH group. Immunohistochemical staining showed that the percentage of CD33-positive cells was significantly higher in patients in the MASH group than in those in the non-MASH group. The results showed that the expression of

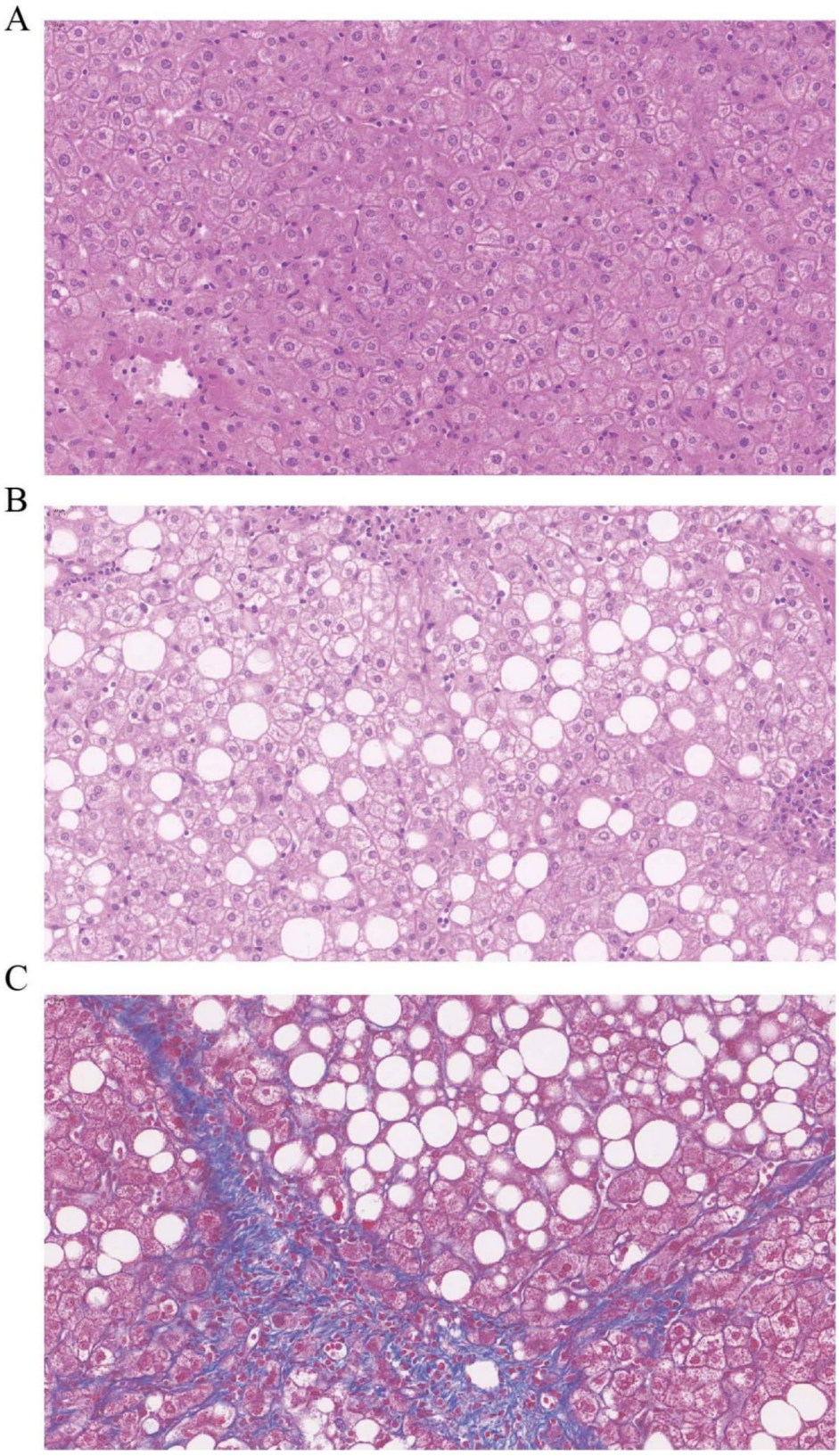

Fig. 4 Representative H&E and Masson's trichrome staining of liver from non-MASH patient and MASH patient (40×, 20 μm). (**A**) Representative H&E staining of liver from non-MASH patient. (**B**) Representative H&E staining of liver from MASH patient. (**C**) Representative Masson's trichrome staining of liver from MASH patient. H&E, hematoxylin and Eosin; MASH, metabolic dysfunction-associated steatohepatitis

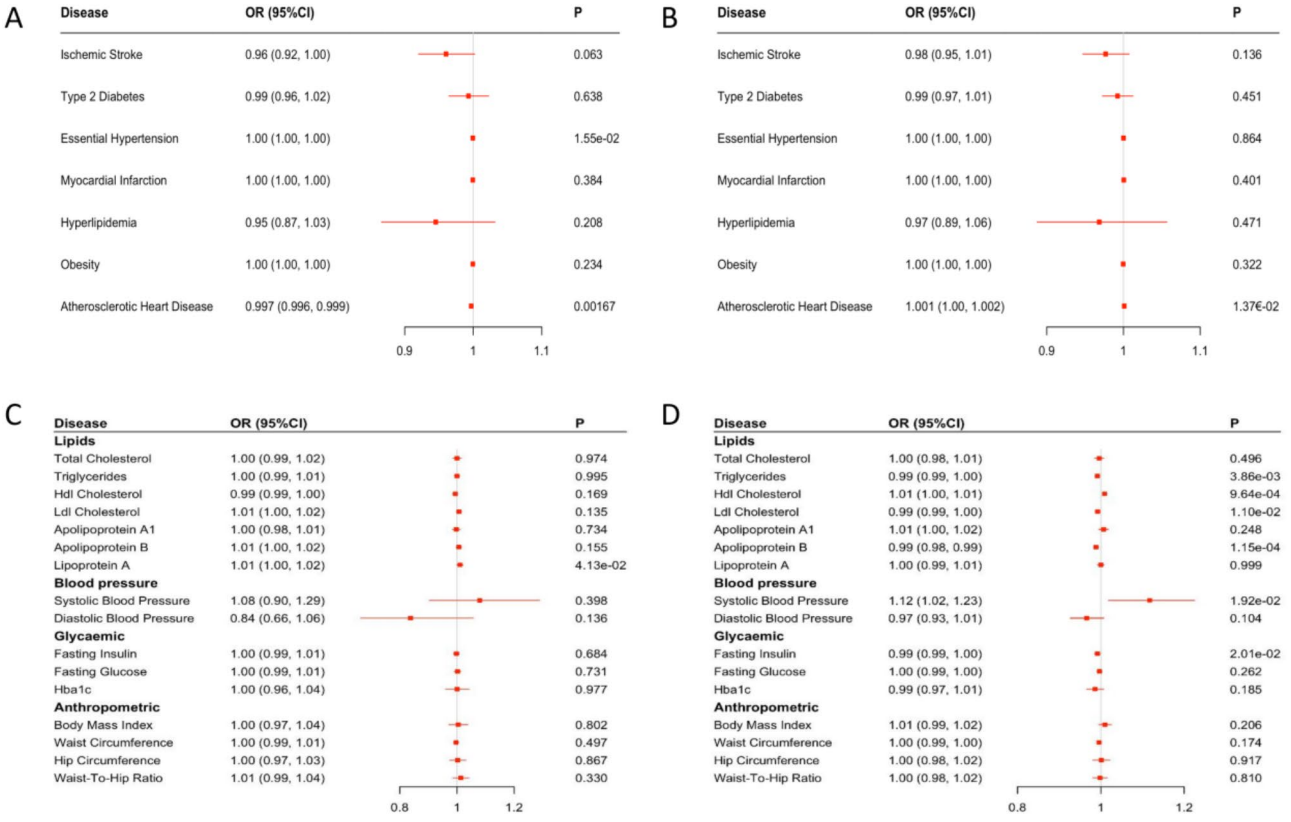

Fig. 5 Associations between genetically predicted *CD33 and MFGE8* and other metabolic-related diseases. MR forest plot showing the effect estimates and 95% CI for the genetically proxied antagonistic effect of *CD33* on 7 metabolic-related diseases (**A**) and 16 common risk factors (**C**). MR forest plot showing the effect estimates and 95% CI for the genetically proxied antagonistic effect of *MFGE8* on 7 metabolic-related diseases (**B**) and 16 common risk factors (**D**). CI, confidence interval; MR, Mendelian randomization; OR, odds ratio

CD33 was significantly higher in patients with MASH (Figs. 6 and 7). At the single-cell level, *CD33* is expressed in immune cells (myeloid cells, lymphocytes, and erythrocytes) and endothelial cells, with some expression in hepatocytes (https://singlecell.broadinstitute.org/singlecell) (Figure S2).

MFGE8

MFGE8 is another druggable gene screened for the presence of a significance threshold using colocalization and hepatic transcriptomics analysis. MR analysis showed a correlation between reduced *MFGE8* expression and increased risk of MASLD (OR = 0.89, 95% CI 0.85–0.94, $P = 2.15 \times 10^{-6}$). Therefore, *MFGE8* plays an inhibitory role in the development of MASLD. Significant negative associations were found between the surrogate gene agonists on *MFGE8* and lipid indices of triglycerides ($P = 3.86 \times 10^{-3}$ [IVW]), ApoB ($P = 1.15 \times 10^{-4}$ [IVW]), and a positive association with HDL ($P = 9.64 \times 10^{-4}$ [IVW]). No significant associations were observed between surrogate gene agonists of *MFGE8* and blood pressure, glucose, and weight-related markers ($P > 2.17 \times 10^{-3}$ [IVW], 0.05/23 results). Meanwhile, MR analyses of metabolic diseases showed that *MFGE8* activation did not significantly increase the risk of metabolic diseases. The observed correlations between surrogate gene agonists of *MFGE8* and atherosclerotic heart disease were weak ($P = 1.37 \times 10^{-2}$ [IVW] Fig. 5). However, after Bonferroni correction, this association was no longer statistically significant.

Furthermore, the RNA and protein expression levels of *MFGE8* in human liver tissues were examined. The findings indicated that the RNA expression level of *MFGE8* was lower in patients with MASH, suggesting that the level of *MFGE8* is negatively correlated with the risk of MASH (Fig. 6). Immunohistochemical staining showed that the percentage of MFGE8-positive cells was significantly higher in patients in the non-MASH group than in those in the MASH group. The results showed that the expression of *MFGE8* was significantly lower in patients with MASH (Fig. 7). At the single-cell level, *MFGE8* is expressed in endothelial, epithelial, and hepatocyte cells (https://singlecell.broadinstitute.org/singlecell) (Figure S2).

MR results of eQTL of liver tissue from GTEx

MR results of eQTL of liver tissue from GTEx (v10.0) were as follows: CD33 (OR = 1.195, 95% CI: 1.12–1.27,

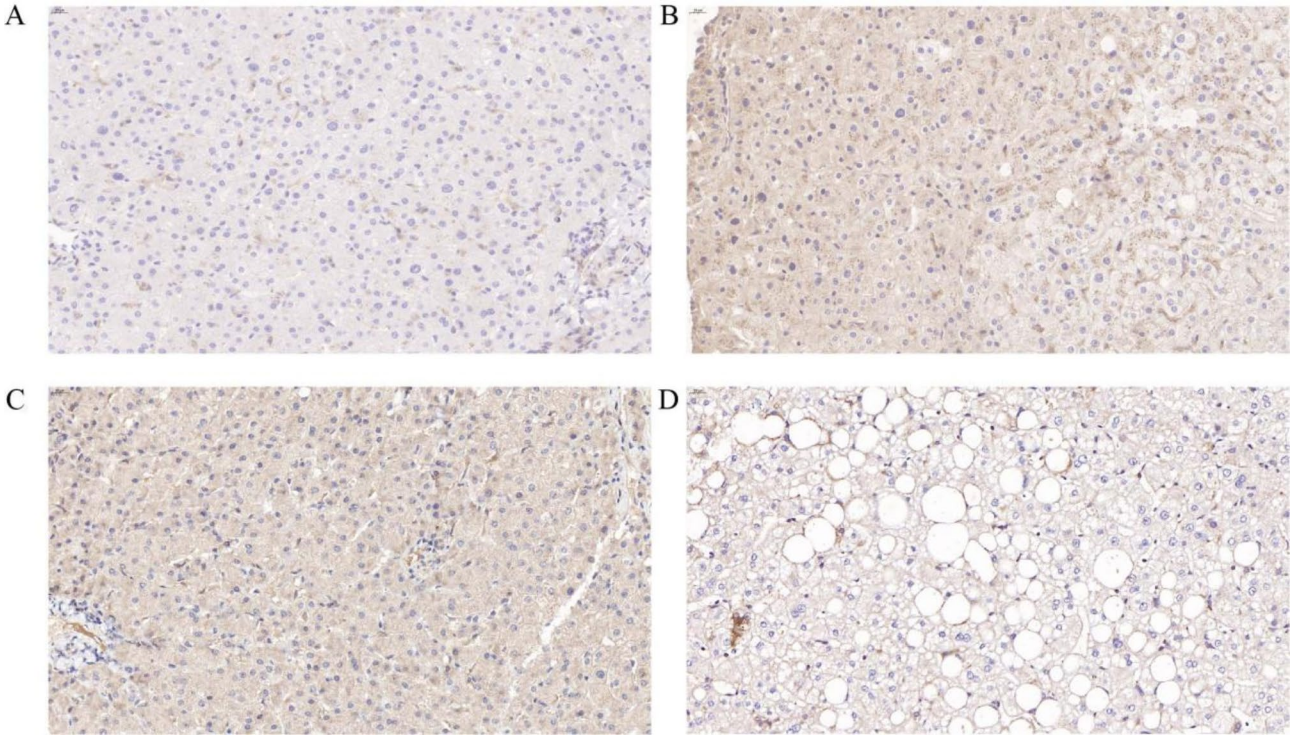

Fig. 6 Protein expression of *CD33* and *MFGE8* in liver tissue of patients with MASLD (40×, 20 μm). (**A**) *CD33* expression in the liver of patients in the non-MASH group. (**B**) *CD33* expression in the liver of patients with MASH. (**C**) *MFGE8* expression in the liver of patients in the non-MASH group. (**D**) *MFGE8* expression in the liver of patients with MASH, with six samples per group. MASH, metabolic dysfunction-associated steatohepatitis; *MFGE8*, milk fat globule-EGF factor 8

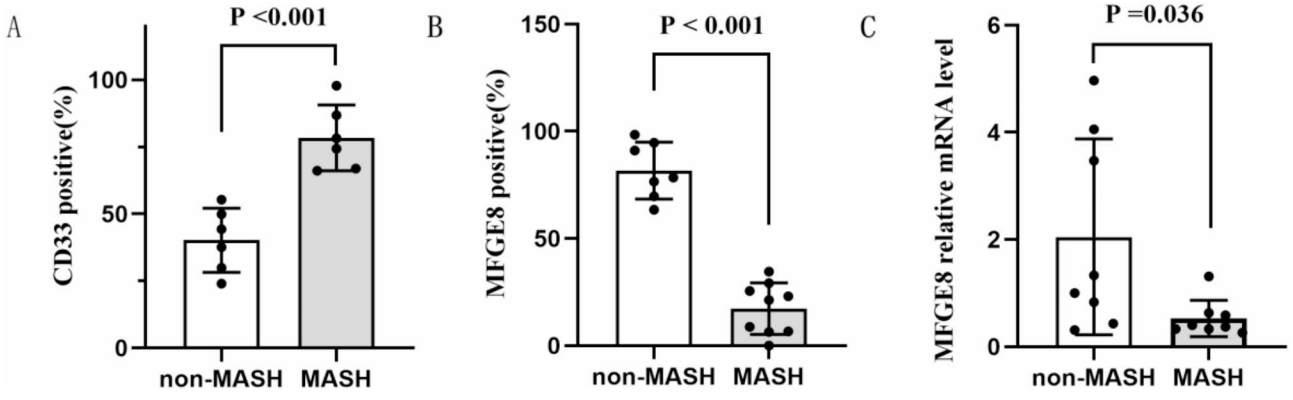

Fig. 7 Expression of *CD33* and *MFGE8* in liver tissues between the non-MASH and MASH groups. (**A**) Expression of *CD33* positive (%) in liver tissues between the non-MASH and MASH groups. (**B**) Expression of *MFGE8* positive (%) in liver tissues between the non-MASH and MASH groups. (**C**) Expression of *MFGE8* relative mRNA in liver tissues between the non-MASH and MASH groups. MASH, metabolic dysfunction-associated steatohepatitis; *MFGE8*, milk fat globule-EGF factor 8

$P = 1.05 \times 10^{-8}$ [IVW]); and MFGE8 (OR = 1.02, 95% CI: 0.997–1.03, $P = 9.38 \times 10^{-2}$ [IVW]). The OR for MFGE8 was 1.02 (95% CI: 0.997–1.03), with a *P*-value of 0.0938. This suggests a non-significant 2% increase in disease risk per MFGE8 copy, as the CI includes one and the *P*-value is >0.05, indicating no significant association with disease risk (Table S13).

SMR analysis findings for liver eQTLs from GTEx (v8.0 and v10.0)

GTEx (v8.0) results are as follows: CD33 (OR = 1.077, 95% CI: 0.967–1.187, $P = 5.20 \times 10^{-2}$) and MFGE8 (OR = 0.966, 95% CI: 0.847–1.084, $P = 7.34 \times 10^{-2}$) (Table S14). CD33 is relatively significant before FDR adjustment, and the direction is consistent with the results of the main analysis of the discovery set. However, after FDR adjustment, both are no longer significant.

GTEx (v10.0) results are as follows: CD33 (OR = 1.847, 95% CI: 1.598–2.096, $P = 3.53 \times 10^{-2}$); MFGE8: (OR = 1.185, 95% CI: 1.058–1.313, $P = 2.55 \times 10^{-1}$). CD33 is relatively significant before FDR adjustment. However, the direction of MFGE8 is not consistent with the results of the main analysis of the discovery set. Moreover, after FDR adjustment, both were no longer significant (Supplement Table S14).

Sensitivity analysis results

Sensitivity analysis results of MR results for eQTL of blood showed that CD33 and MFGE8 (Heterogeneity, steiger_direction, and MR presso tests) were robust (Supplement Table S5–7). In the liver tissue, CD33 remained robust, while MFGE8 exhibited some heterogeneity (Supplement Table S13). The HEIDI test from SMR analysis showed that results of both the target genes were robust (Supplement Table S14).

Discussion

With the rising prevalence of obesity and its associated metabolic sequelae, MASLD has emerged as the predominant chronic liver condition worldwide [40]. The etiopathogenesis of MASLD is multifaceted and not completely elucidated, contributing to a lack of effective pharmacotherapies. Therefore, the identification of pathogenic targets for MASLD is critical for the advancement of tailored treatment approaches. In this extensive MR study comprising 4,761 individuals with MASLD and 373,227 controls, GWAS data, pharmacogenomic information, and gene expression profiles were investigated. The study findings indicate that inhibition of *CD33* gene function and activation of MGFE8 are inversely associated with the risk of MASLD. Furthermore, it was observed that the genetically predicted inhibition of *CD33* is significantly inversely correlated with atherosclerotic heart disease.

CD33, a member of the sialic acid-binding immunoglobulin-like lectin (Siglec) family, is a cell surface protein that mediates cell-cell interactions and immune regulation. As a transmembrane protein, CD33 is expressed across hematopoietic and phagocytic cells, such as microglia, dendritic cells, monocytes, macrophages, granulomonocyte precursors, and hematopoietic progenitor cells [41, 42]. However, its precise associations with diseases is not fully elucidated.

CD33 is highly expressed in leukemia blasts and myeloid leukemia initiating cells, yet it is not present in primitive stem cells and multipotent progenitors. While expressed in common myeloid precursors, *CD33* levels are reduced in mature granulocytes and circulating macrophages, making it a promising target for immunotherapy in acute myeloid leukemia (AML) [43]. *CD33* is expressed on mature myeloid cells and hematopoietic progenitors, necessitating caution to prevent non-tumor toxic effects when developing *CD33* antibody-based therapies. Mylotarg (Gemtuzumab Ozogamicin, GO) is an antibody-drug conjugate (ADC) that consists of an anti-*CD33* IgG4 monoclonal antibody linked to calicheamicin. It targets the *CD33* antigen on AML cells, is internalized, and releases calicheamicin to eradicate the leukemia cells, extending survival in certain patients [44]. Currently, researchers are developing more effective *CD33*-targeted therapeutic agents, such as antibodies against different epitopes of *CD33* [45, 46].

The *CD33* gene, expressed on brain microglial cells and linked to Alzheimer's disease (AD) risk, regulates processes such as cytokine release and immune cell growth. *CD33* might facilitate AD onset through hindering microglial clearance of β-amyloid [47, 48]. Research suggests AD starts with lysosomal autophagy dysfunction in neurons, leading to β-amyloid deposits [49]. Currently, no *CD33*-related drugs for the treatment of AD exist. However, the complexity of *CD33*-expressing isoforms and safety concerns related to cytotoxicity have limited the ability to effectively target and use the drug. Some studies indicate that CD33 is an immune checkpoint receptor for HBV-induced immune tolerance [50].

The multifaceted role of *CD33* in various health and disease states underscores its importance in immune regulation. Regarding MASLD, a notable aspect of *CD33*'s function is its association with myeloid-derived suppressor cells (MDSCs), which serve as a key regulator of hepatic lipid metabolism. MDSC counts increase with lipid accumulation, and those derived from polymorphonuclear cells in the livers of obese mice have been shown to enhance liver lipid deposition through secreting proinflammatory factors (e.g., S100A9), disrupting normal liver lipid metabolism [51, 52]. In humans, both polymorphonuclear cell-derived MDSCs (PMN-MDSCs) and monocyte-derived MDSCs (M-MDSCs) express *CD33*. It is inferred that the increased expression of *CD33* in MASLD might be related to the increase in myeloid-derived suppressor cells owing to long-term chronic inflammation.

A preprint study has identified four novel genetic targets for MASLD, including *CD33*, by analyzing 2,941 plasma proteins from 43,978 individuals in the UK Biobank. Integrating human genetics, transcriptomics, liver imaging, and biopsy proteomics, the research aligns with our results. It indicates that the association of *CD33* to MASLD is mediated through body weight, with pathway analysis implicating *CD33* in the regulation of interleukin-1beta production [15]. The precise role of *CD33* in the pathogenesis of MASLD warrants further exploration. Despite the existence of *CD33*-targeted therapies, the relationship between *CD33* and MASLD has not

been fully investigated. The findings propose *CD33* as a conceivable therapeutic target for MASLD intervention.

MFGE8 was first discovered in the milk fat globules of mouse mammary epithelial cells and functions as a secreted cellular protein. *MFGE8* is an anti-inflammatory factor and is involved in a variety of physiological processes, including fertilization [53], angiogenesis [54], innate immunity [55], and tumorigenesis [56]. Studies have shown that *MFGE8* expression is associated with myocardial hypertrophy, coronary atherosclerosis [57, 58], and anti-inflammatory effects such as post-operative injury [59], and sepsis [60]. *MFGE8* is an endogenous inhibitor of inflammation-induced IL-1β production and inhibits macrophage-induced necrotic cell and ATP-dependent IL-1β production by inducing the interaction of integrin β3 and P2 × 7 receptors in cells.

Zhang Lei et al. [61] reported that *MFGE8* expression is decreased in patients with MASLD and in mouse models. During the process of lipid accumulation in hepatocytes, *MFGE8* was shown to negatively regulate inflammation and lipid accumulation through binding to apoptosis signal-regulating kinase 1 (ASK1) and inhibiting its activation. The MR analysis suggests a causal association between *MFGE8* agonists and lipid metabolism markers (TG, HDL, apo B). Under metabolic stress, a decrease in *MFGE8* promotes ASK1-dependent inflammation, suggesting that *MFGE8* may serve as a potential therapeutic target for preventing the progression of MASLD, which is consistent with our study findings. MFG-E8 participates in a series of inflammatory processes mediated by the liver-spleen axis. Some studies showed that tingible body macrophages in the germinal centers of the spleen and lymph nodes strongly express MFG-E8, and that the MFG-E8−/− mice developed splenomegaly, with the formation of numerous germinal centers [62]. Tarantino G reported all process is in agreement with the well-ascertained chronic low-grade inflammation, characteristics of NAFLD, key-process mediated by the liver spleen axis [63]. In this process, there are also some similar targets such as S1009, nd S100A8 and S100A9 heterodimers, along with S100A9 homodimers, were formed after an inflammatory challenge in protein extracts from the spleen [64].

Study strengths and limitations

The complexity of MASLD pathogenesis, which involves various pathophysiological processes and multiple factors across various tissues and organs, presents a significant challenge in identifying suitable pharmacological intervention targets. This study focused on druggable genes to improve the efficacy, safety, and success rates of drug development for MASLD, identifying *CD33* and *MFGE8* as potential targets. Additionally, this population-based research suggests that these proteins could be crucial in the development of future therapeutic targets for MASLD. MR analysis was also conducted to address potential safety concerns and explore alternative indications, which are crucial for their clinical application.

While this study offers valuable insights, it has some limitations. A population overlap was observed between the validation and discovery sets, slightly weakening the validation effect. However, MR and SMR analysis of the liver tissue from GTEx was performed. The relationship between trans-eQTLs and disease was not analyzed, although cis-eQTLs comprise an integral part of quantitative loci that regulate gene expression. The challenge in conducting MR analysis to fully mitigate the potential influence of horizontal pleiotropy should be acknowledged, as well as potential residual pleiotropy and the inability to fully capture gene-environment interactions. The investigation of side effects was confined to outcomes related to metabolic disorders, necessitating future research to broaden this scope and address potential systemic adverse effects. The study's sample population in the MR section was restricted to individuals of European descent, limiting the generalizability to other populations. Sensitivity analysis showed that only the result for MFGE8 among all MR results for eQTLs of liver tissue (GTEx) exhibited some heterogeneity, likely due to the small GTEx sample size.

Conclusion

Given the lack of effective pharmacological interventions for MASLD, developing novel therapeutics targeting *CD33* and *MFGE8* is critical. Future research is needed to determine whether *CD33* inhibitors and *MFGE8* activators produce consistent effects in preclinical and clinical settings, and whether the expression regulation of the two targets would produce the expected clinical results through the application of computer simulation and molecular design technology, as well as their fundamental biological pathways.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12944-025-02515-8.

Supplementary Material 1: **Supplementary file S1** RNA extraction and library preparation.

Supplementary Material 2: **Fig. S1**. Liver transcriptome principal component analysis.

Supplementary Material 3: **Fig. S2**. Single cell sequencing analysis.

Supplementary Material 4: **Table S1**. Data sources for MASLD in this study.

Supplementary Material 5: **Table S2.** Data sources for the analysis.

Supplementary Material 6: **Table S3.** Primer list.

Supplementary Material 7: **Table S4.** Supplementary MR results.

Supplementary Material 8: **Table S5.** Summary of the results of genetic

variants in significant genes.

Supplementary Material 9: **Table S6.** Heterogeneity test results.

Supplementary Material 10: **Table S7.** Horizontal pleiotropy test results.

Supplementary Material 11: **Table S8.** Steiger test results.

Supplementary Material 12: **Table S9**. External validation MR result.

Supplementary Material 13: **Table S10.** Colocalization results.

Supplementary Material 14: **Table S11.** Population characteristics and comparison between subjects with and without MASH.

Supplementary Material 15: **Table S12.** Transcriptomics raw data.

Supplementary Material 16: **Table S13.** MR results of liver eQTL.

Supplementary Material 17: **Table S14.** SMR analysis results of liver eQTL.

Acknowledgements
We thank the eQTLGen consortium for providing publicly available blood eQTL data, the GTEx Consortium for sharing liver tissue eQTL data, the FinnGen team, and other researchers and participants for contributing their data, all of which were crucial for this analysis.

Author contributions
Qing He, Ming Liu, and Longhao Sun conceptualized the research hypothesis and supervised the research process. Xiaohui Ma and Li Ding undertook data analysis and contributed to the composition of the manuscript, Shuo Li and Fan Yu were responsible for data access and extraction. Xin Wang, Yitong Han, and Hengjie Yuan interpreted and verified the results of the analysis. All authors collectively endorsed the final version of the manuscript. Qing He, Hengjie Yuan, and Longhao Sun took responsibility for the decision to submit the manuscript.

Funding
This work was supported by Tianjin Key Medical Discipline (Specialty) Construction Project (grant number TJYXZDXK-030 A), Major Project of Tianjin Municipal Science and Technology Bureau (grant number 21ZXJBSY00060), the National Natural Science Foundation of China (grant number 82200882), and Tianjin Major Science and Technology Projects (grant number 21ZXJBSY00050). The funding sources did not play any role in data extraction and analysis, study design, or any aspect relating to the research.

Data availability
No datasets were generated or analysed during the current study. The UK Biobank resource is available to bona fide researchers for health-related research in the public interest upon application at http://www.ukbiobank.ac.uk/using-the-resource/from. --Publicly available summary statistics are obtained https://gwas.mrcieu.ac.uk/, https://www.gtexportal.org/home/datasets, https://www.ebi.ac.uk/gwas/ and https://pan.ukbb.broadinstitute.org. --The raw sequencing data generated (and bulk liver RNA-seq data) in this study have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under the BioProject accession number PRJNA1215553 (https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA1215553) .-- All the code for the main analysis and the RNA sequencing analysis have been uploaded to the github: https://github.com/maxiaohuitianjinmedical--Other source of data or web source were clarified in the methods. Relevant detailed results generated in this study were presented in supplementary tables. Clinical data with respect to the liver biopsy samples were assessable through contacting us.

Declarations

Ethics approval and consent to participate
The GWAS dataset used in this study is a secondary analysis of publicly available data. In accordance with the original GWAS protocol, all participants provided informed consent, and all ethical approvals related to the GWAS were issued by the original authors of the GWAS. The experimental verification part of this study has been approved by the Institutional Review Board of Tianjin Medical University General Hospital (Tianjin, China) (No. IRB2020-YX-029-01/No.IRB2023-YX-204-01) and received consent from the patients.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Endocrinology and Metabolism, Tianjin Medical University General Hospital, 154 Anshan Road, Heping District, Tianjin 300052, China
[2]Department of General Surgery, Tianjin Medical University General Hospital, 154 Anshan Road, Heping District, Tianjin 300052, China
[3]Department of Pharmacy, Tianjin Medical University General Hospital, 154 Anshan Road, Heping District, Tianjin 300052, China
[4]Department of Endocrinology and Metabolism, Baotou Central Hospital, Baotou, China

Received: 6 November 2024 / Accepted: 6 March 2025
Published online: 26 March 2025

References

1. 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.
2. Driessen S, Francque SM, Anker SD, Castro Cabezas M, Grobbee DE, Tushuizen ME et al. Metabolic dysfunction-associated steatotic liver disease and the heart. Hepatology. 2023.
3. Rinella ME, Lazarus JV, Ratziu V, Francque SM, Sanyal AJ, Kanwal F, et al. A multisociety Delphi consensus statement on new fatty liver disease nomenclature. Hepatology. 2023;78(6):1966–86.
4. Younossi ZM, Paik JM, Stepanova M, Ong J, Alqahtani S, Henry L. Clinical profiles and mortality rates are similar for metabolic dysfunction-associated steatotic liver disease and non-alcoholic fatty liver disease. J Hepatol. 2024;80(5):694–701.
5. Song SJ, Lai JC, Wong GL, Wong VW, Yip TC. Can we use old NAFLD data under the new MASLD definition? J Hepatol. 2024;80(2):e54–6.
6. EASL-EASD-EASO Clinical. Practice guidelines on the management of metabolic dysfunction-associated steatotic liver disease (MASLD). J Hepatol. 2024;81(3):492–542.
7. Makri E, Goulas A, Polyzos SA, Epidemiology. Pathogenesis, diagnosis and emerging treatment of nonalcoholic fatty liver disease. Arch Med Res. 2021;52(1):25–37.
8. Anstee QM, Reeves HL, Kotsiliti E, Govaere O, Heikenwalder M. From NASH to HCC: current concepts and future challenges. Nat Rev Gastroenterol Hepatol. 2019;16(7):411–28.
9. Dufour JF, Caussy C, Loomba R. Combination therapy for non-alcoholic steatohepatitis: rationale, opportunities and challenges. Gut. 2020;69(10):1877–84.
10. Govaere O, Cockell S, Tiniakos D, Queen R, Younes R, Vacca M et al. Transcriptomic profiling across the nonalcoholic fatty liver disease spectrum reveals gene signatures for steatohepatitis and fibrosis. Sci Transl Med. 2020; 12(572).
11. Jonas W, Schürmann A. Genetic and epigenetic factors determining NAFLD risk. Mol Metab. 2021;50:101111.
12. Teo K, Abeysekera KWM, Adams L, Aigner E, Anstee QM, Banales JM, et al. rs641738C > T near MBOAT7 is associated with liver fat, ALT and fibrosis in NAFLD: A meta-analysis. J Hepatol. 2021;74(1):20–30.
13. Boehm FJ, Zhou X. Statistical methods for Mendelian randomization in genome-wide association studies: A review. Comput Struct Biotechnol J. 2022;20:2338–51.
14. Schmidt AF, Finan C, Gordillo-Marañón M, Asselbergs FW, Freitag DF, Patel RS, et al. Genetic drug target validation using Mendelian randomisation. Nat Commun. 2020;11(1):3255.
15. Liu J, Hu S, Chen L, Daly C, Prada Medina CA, Richardson TG et al. Profiling the genome and proteome of metabolic dysfunction-associated steatotic liver disease identifies potential therapeutic targets. MedRxiv. 2023.

16. Chen J, Rao H, Zheng X. Identification of novel targets associated with cholesterol metabolism in nonalcoholic fatty liver disease: a comprehensive study using Mendelian randomization combined with transcriptome analysis. Front Genet. 2024;15:1464865.
17. Xiao QA, Zhao WJ, Yu J, Qin L, Zhang XL, Yu J. Identification of novel drug targets for liver cirrhosis and its potential side-effects by human plasma proteome. Sci Rep. 2024;14(1):28884.
18. Finan C, Gaulton A, Kruger FA, Lumbers RT, Shah T, Engmann J et al. The druggable genome and support for target identification and validation in drug development. Sci Transl Med. 2017; 9(383).
19. Võsa U, Claringbould A, Westra HJ, Bonder MJ, Deelen P, Zeng B, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. Nat Genet. 2021;53(9):1300–10.
20. Ghodsian N, Abner E, Emdin CA, Gobeil É, Taba N, Haas ME, 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(11):100437.
21. Fairfield CJ, Drake TM, Pius R, Bretherick AD, Campbell A, Clark DW, et al. Genome-Wide association study of NAFLD using electronic health records. Hepatol Commun. 2022;6(2):297–308.
22. Hagström H, Adams LA, Allen AM, Byrne CD, Chang Y, Grønbaek H, et al. Administrative coding in electronic health care Record-Based research of NAFLD: an expert panel consensus statement. Hepatology. 2021;74(1):474–82.
23. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. Nat Genet. 2013;45(11):1274–83.
24. Surendran P, Feofanova EV, Lahrouchi N, Ntalla I, Karthikeyan S, Cook J, et al. Discovery of rare variants associated with blood pressure regulation through meta-analysis of 1.3 million individuals. Nat Genet. 2020;52(12):1314–32.
25. Chen J, Spracklen CN, Marenne G, Varshney A, Corbin LJ, Luan J, et al. The trans-ancestral genomic architecture of glycemic traits. Nat Genet. 2021;53(6):840–60.
26. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015;518(7538):197–206.
27. Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Mägi R, et al. New genetic loci link adipose and insulin biology to body fat distribution. Nature. 2015;518(7538):187–96.
28. Dönertaş HM, Fabian DK, Valenzuela MF, Partridge L, Thornton JM. Common genetic associations between age-related diseases. Nat Aging. 2021;1(4):400–12.
29. Trinder M, Vikulova D, Pimstone S, Mancini GBJ, Brunham LR. Polygenic architecture and cardiovascular risk of Familial combined hyperlipidemia. Atherosclerosis. 2022;340:35–43.
30. Sakaue S, Kanai M, Tanigawa Y, Karjalainen J, Kurki M, Koshiba S, et al. A cross-population atlas of genetic associations for 220 human phenotypes. Nat Genet. 2021;53(10):1415–24.
31. Ben Elsworth. (2018). [Diagnoses - main ICD10: I25.1 Atherosclerotic heart disease]. MRC Integrative Epidemiology Unit, University of Bristol. IEU Number: ukb-b-1668. Retrieved from https://gwas.mrcieu.ac.uk/datasets/ukb-b-1668/ on August 4, 2024.
32. Relton CL, Davey Smith G. Two-step epigenetic Mendelian randomization: a strategy for Establishing the causal role of epigenetic processes in pathways to disease. Int J Epidemiol. 2012;41(1):161–76.
33. 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.
34. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol. 2013;37(7):658–65.
35. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 2014;10(5):e1004383.
36. Bedossa P, Poitou C, Veyrie N, Bouillot JL, Basdevant A, Paradis V, et al. Histopathological algorithm and scoring system for evaluation of liver lesions in morbidly obese patients. Hepatology. 2012;56(5):1751–9.
37. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
38. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12(4):357–60.
39. Anders S, Pyl PT, Huber W. HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31(2):166–9.
40. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28(5):511–5.
41. Crocker PR, Paulson JC, Varki A. Siglecs and their roles in the immune system. Nat Rev Immunol. 2007;7(4):255–66.
42. Crocker PR, McMillan SJ, Richards HE. CD33-related Siglecs as potential modulators of inflammatory responses. Ann N Y Acad Sci. 2012;1253:102–11.
43. Ishii H, Yano S. New therapeutic strategies for adult acute myeloid leukemia. Cancers (Basel). 2022; 14(11).
44. Morse JW, Rios M, Ye J, Rios A, Zhang CC, Daver NG, et al. Antibody therapies for the treatment of acute myeloid leukemia: exploring current and emerging therapeutic targets. Expert Opin Investig Drugs. 2023;32(2):107–25.
45. Godwin CD, Laszlo GS, Fiorenza S, Garling EE, Phi TD, Bates OM, et al. Targeting the membrane-proximal C2-set domain of CD33 for improved CD33-directed immunotherapy. Leukemia. 2021;35(9):2496–507.
46. Albinger N, Pfeifer R, Nitsche M, Mertlitz S, Campe J, Stein K, et al. Primary CD33-targeting CAR-NK cells for the treatment of acute myeloid leukemia. Blood Cancer J. 2022;12(4):61.
47. Eskandari-Sedighi G, Jung J, Macauley MS. CD33 isoforms in microglia and Alzheimer's disease: friend and foe. Mol Aspects Med. 2023;90:101111.
48. Griciuc A, Patel S, Federico AN, Choi SH, Innes BJ, Oram MK, et al. TREM2 acts downstream of CD33 in modulating microglial pathology in Alzheimer's disease. Neuron. 2019;103(5):820–e357.
49. Lee JH, Yang DS, Goulbourne CN, Im E, Stavrides P, Pensalfini A, et al. Faulty autolysosome acidification in Alzheimer's disease mouse models induces autophagic build-up of Aβ in neurons, yielding senile plaques. Nat Neurosci. 2022;25(6):688–701.
50. Tsai TY, Huang MT, Sung PS, Peng CY, Tao MH, Yang HI et al. SIGLEC-3 (CD33) serves as an immune checkpoint receptor for HBV infection. J Clin Invest 2021; 131(11).
51. Liu YF, Wei JY, Shi MH, Jiang H, Zhou J. Glucocorticoid induces hepatic steatosis by inhibiting activating transcription factor 3 (ATF3)/S100A9 protein signaling in granulocytic Myeloid-derived suppressor cells. J Biol Chem. 2016;291(41):21771–85.
52. Ramachandran P, Dobie R, Wilson-Kanamori JR, Dora EF, Henderson BEP, Luu NT, et al. Resolving the fibrotic niche of human liver cirrhosis at single-cell level. Nature. 2019;575(7783):512–8.
53. Ensslin MA, Shur BD. Identification of mouse sperm SED1, a bimotif EGF repeat and discoidin-domain protein involved in sperm-egg binding. Cell. 2003;114(4):405–17.
54. Ni YQ, Zhan JK, Liu YS. Roles and mechanisms of MFG-E8 in vascular aging-related diseases. Ageing Res Rev. 2020;64:101176.
55. Deroide N, Li X, Lerouet D, Van Vré E, Baker L, Harrison J, et al. MFGE8 inhibits inflammasome-induced IL-1β production and limits postischemic cerebral injury. J Clin Invest. 2013;123(3):1176–81.
56. Tibaldi L, Leyman S, Nicolas A, Notebaert S, Dewulf M, Ngo TH, et al. New blocking antibodies impede adhesion, migration and survival of ovarian cancer cells, highlighting MFGE8 as a potential therapeutic target of human ovarian carcinoma. PLoS ONE. 2013;8(8):e72708.
57. Deng KQ, Li J, She ZG, Gong J, Cheng WL, Gong FH, et al. Restoration of Circulating MFGE8 (Milk fat Globule-EGF factor 8) attenuates cardiac hypertrophy through Inhibition of Akt pathway. Hypertension. 2017;70(4):770–9.
58. Dai W, Li Y, Lv YN, Wei CD, Zheng HY. The roles of a novel anti-inflammatory factor, milk fat globule-epidermal growth factor 8, in patients with coronary atherosclerotic heart disease. Atherosclerosis. 2014;233(2):661–5.
59. Cheyuo C, Jacob A, Wu R, Zhou M, Qi L, Dong W, et al. Recombinant human MFG-E8 attenuates cerebral ischemic injury: its role in anti-inflammation and anti-apoptosis. Neuropharmacology. 2012;62(2):890–900.
60. Wu W, Wang J, Chen J, Lu J, Lao Y, Huang K, et al. MFG-E8 has guiding significance for the prognosis and treatment of sepsis. Sci Rep. 2022;12(1):20916.
61. Zhang L, Tian R, Yao X, Zhang XJ, Zhang P, Huang Y, et al. Milk fat Globule-Epidermal growth Factor-Factor 8 improves hepatic steatosis and inflammation. Hepatology. 2021;73(2):586–605.
62. Hanayama R, Tanaka M, Miyasaka K, Aozasa K, Koike M, Uchiyama Y, et al. Autoimmune disease and impaired uptake of apoptotic cells in MFG-E8-deficient mice. Science. 2004;304(5674):1147–50.
63. Tarantino G, Citro V, Balsano C. Liver-spleen axis in nonalcoholic fatty liver disease. Expert Rev Gastroenterol Hepatol. 2021;15(7):759–69.

64. Källberg E, Tahvili S, Ivars F, Leanderson T. Induction of S100A9 homodimer formation in vivo. Biochem Biophys Res Commun. 2018;500(3):564–8.

Publisher's note