HEPATIC EPIDEMIOLOGY

Lifestyle and metabolic factors for nonalcoholic fatty liver disease: Mendelian randomization study

Shuai Yuan1 · Jie Chen2,3 · Xue Li4,5 · Rongrong Fan6 · BenoitArsenault7,8 · DipenderGill9,10,11,12 · EdwardL. Giovannucci13,14 · Ju‑shengZheng15 · SusannaC. Larsson1,16

Received: 14February 2022 / Accepted: 4 April 2022 / Published online: 30 April 2022
© The Author(s) 2022

Abstract
The risk factors for nonalcoholic fatty liver disease (NAFLD) have not been clearly identified. We conducted a Mendelian randomization (MR) study to explore this. Independent genetic variants strongly associated with 5 lifestyle and 9 metabolic factors were selected as instrumental variables from corresponding genome-wide association studies (GWASs). Summary-level data for NAFLD were obtained from a GWAS meta-analysis of 8434 cases and 770,180 non-cases (discovery dataset) and another GWAS meta-analysis of 1483 cases and 17,781 non-cases (replication dataset). Univariable and multivariable MR analyses were performed. There were associations with NAFLD for lifetime smoking index (odds ratio (OR) 1.59, 95% confidence interval (CI) 1.31–1.93 per SD-increase), body mass index (BMI, OR 1.33, 95% CI 1.23–1.43 per SD-increase), waist circumference (OR 1.82; 95% CI 1.48–2.24 per SD-increase), type 2 diabetes (OR 1.21, 95% CI 1.15–1.27 per unit increase in log-transformed odds), systolic blood pressure (OR 1.17; 95% CI 1.15–1.27 per 10 mmHg increase), high-density lipoprotein cholesterol (OR 0.84, 95% CI 0.77–0.90 per SD-increase), and triglycerides (OR 1.23, 95% CI 1.15–1.33 per SD-increase). The associations for type 2 diabetes, systolic blood pressure, triglycerides, but not for high-density lipoprotein cholesterol remained strong after adjusting for genetically-predicted BMI. Genetic liability to type 2 diabetes mediated 51.4% (95% CI 13.4–89.3%) of the BMI-effects on NAFLD risk. There were suggestive inverse associations of genetically-predicted alcohol, coffee, and caffeine consumption, and vigorous physical activity with NAFLD risk. This study identified several lifestyle and metabolic factors that may be causally implicated in NAFLD.

Keywords Lifestyle · Mendelian randomization · Metabolic factor · Nonalcoholic fatty liver disease

Introduction
Nonalcoholic fatty liver disease (NAFLD) has become the most common form of chronic liver disease afflicting ~ 25% of the global population [1]. NAFLD has merged as the second leading indication for liver transplantation in the United States [2]. Due to obesity and diabetes epidemics, the disease burden of NAFLD is projected to increase 2 to threefold in Western countries as well as in several Asian areas by 2030 [3]. Although obesity is an important risk factor for NAFLD development, non-obese NAFLD patients have been identified as a large cluster making up ~ 20% of worldwide NAFLD population [4], which implies the etiological complexity of this disease as well as the possibility of prevention strategies targeting at other modifiable factors. Previous observational studies have identified modifiable factors for NAFLD, including metabolic traits, smoking, alcohol drinking, coffee consumption, and physical activity [5–7]. However, certain modifiable exposures, such as cigarette smoking [8] and alcohol drinking [9], have been inconsistently associated with NAFLD risk. The mutual relationship between metabolic syndrome and NAFLD are intertwined, especially in cross-sectional and case–control studies [10]. In addition, whether the associations of above factors with NAFLD risk are causal remains undermined due to potential residual confounding and reverse causality issues in observational studies.

Utilizing genetic variants as instrumental variables, Mendelian randomization (MR) is an epidemiological technique aimed at strengthening causal inference [11]. The approach
carries two merits of minimizing confounding and diminishing reverse causality because genetic variants are randomly allocated at conception (thus unrelated to self-adopted and environmental factors) and cannot be modified by the development and progression of the disease [11]. Here, we conducted an MR study to investigate the associations of metabolic and lifestyle factors with risk of NAFLD. We also examined the obesity-independent effects of metabolic features on NAFLD as well as explored the mediators in the association between obesity and NAFLD.

**Methods**

**Study design**

Figure 1 shows the study design. The present MR study included 14 modifiable factors (5 lifestyle and 9 metabolic factors). We firstly examined the associations of these factors with NAFLD in a large discovery dataset and then performed a replication analysis in an independent population. To increase power of the analysis, we combined estimates from two data sources. Multivariable MR method and mediation analysis were used. The analysis was conducted using summary-level data from published genome-wide association studies (GWASs) and the analytic process was in accordance with the STROBE-MR guidelines [12]. All studies included in cited GWASs had been approved by a relevant review board and all participants had provided the consent forms. The present MR analyses were approved by the Swedish Ethical Review Authority (2019-02793).

**Genetic instrument selection**

Single-nucleotide polymorphisms (SNPs) associated with 14 modifiable factors at the genome-wide significance level ($p \leq 5 \times 10^{-8}$) were obtained from corresponding GWASs (Table 1). We estimated linkage disequilibrium among these SNPs based on the 1000 Genomes European reference panel [13]. SNPs in linkage disequilibrium ($r^2 \geq 0.01$) were excluded and the SNP with the smallest $p$ value for the genome-wide association was attained. Genetic instrument selection for multivariable MR analysis followed the same criteria. For smoking behaviors, two sets of instruments (SNPs for smoking initiation and for lifestyle smoking index) were used for validation. Detailed information on GWASs of studied exposures, including number of participants and adjusted covariates, is presented in Table 1.

**Data sources for NAFLD**

Summary-level data (i.e., beta coefficient and corresponding standard error) for the associations of exposure-associated SNPs with NAFLD were extracted from a GWAS meta-analysis including 8434 NAFLD cases and 770,180 non-cases (discovery stage) [14] and another GWAS including 1483 NAFLD cases and 17,781 non-cases (replication stage) [15] (Supplementary Table 1). Four GWASs (the Electronic Medical Records and Genomics, UK Biobank, FinnGen, and Estonian Biobank) were included in the discovery dataset [14]. The replication GWAS comprised data from 11 leading European tertiary liver centers [15]. Case definition and exclusion criteria in included NAFLD GWASs are shown.
### Table 1: Detailed information on used studies

| Exposure or outcome                  | Unit                                         | Participants included in analysis                    | Adjustments                                                                                                                                                                                                                                                                                                                                 | IVs | PubMed ID            |
|-------------------------------------|----------------------------------------------|------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|----------------------|
| **Lifestyle factor**                |                                              |                                                      |                                                                                                                                                                                                                                                                                                                                                                                                     |     |                      |
| Smoking initiation                  | SD in prevalence of smoking initiation       | 1,232,091 European-descent individuals               | Age, sex, and the first ten genetic principal components                                                                                                                                                                                                                                                                                                                                           | 311 | 30,643,251           |
| Lifetime smoking index              | SD change of lifetime smoking index          | 462,690 European-descent individuals                 | Genotyping chip and sex                                                                                                                                                                                                                                                                                                                                                                | 126 | 31,689,377           |
| Alcohol drinking                    | SD increase of log-transformed alcoholic drinks/week | 941,280 European-descent individuals                 | Age, sex, and the first ten genetic principal components                                                                                                                                                                                                                                                                                                                                   | 84  | 30,643,251           |
| Coffee consumption                  | 50% change                                   | 375,833 European-descent individuals                 | Age, sex, body mass index, total energy, proportion of typical food intake, and 20 genetic principal components                                                                                                                                                                                                                                                                             | 12  | 31,046,077           |
| Caffeine consumption                | 80 mg increase (equivalent to dose from 1 cup of coffee) | 98,76 European-descent individuals | Age, sex, study-site, fasting status, smoking status, and genetic principal components                                                                                                                                                                                                                                                                                                      | 2   | 27,702,941           |
| Vigorous physical activity          | ≥ 3 versus 0 day/week                        | 98,060 cases and 162,995 controls of European descent | Age, sex, genotyping chip, first ten genomic principal components, and center                                                                                                                                                                                                                                                                                                              | 5   | 29,899,525           |
| **Metabolic factor**                |                                              |                                                      |                                                                                                                                                                                                                                                                                                                                                                                                     |     |                      |
| Body mass index                     | SD                                           | 806,834 European-descent individuals                | Age, sex, and genetic 1–5 principal components                                                                                                                                                                                                                                                                                                                                               | 312 | 30,239,722           |
| Waist circumference                 | SD                                           | 224,459 European-descent individuals                | Age and study-specific covariates                                                                                                                                                                                                                                                                                                                                                           | 44  | 25,673,412           |
| Type 2 diabetes                     | One-unit in log-transformed odds             | 228,499 type 2 diabetes cases and 1,178 non-cases of multi-ancestries | Age, sex, and the first ten genetic principal components                                                                                                                                                                                                                                                                                                                                 | 497 | 32,541,925           |
| Systolic blood pressure             | 10 mm Hg                                     | Up to 1,006,863 European-descent individuals        | Age, sex, BMI, genotyping chips                                                                                                                                                                                                                                                                                                                                                          | 228 | 30,224,653           |
| High-density lipoprotein cholesterol| SD                                           | 403,943 European-descent individuals                | Age, sex, and genotyping chips                                                                                                                                                                                                                                                                                                                                                           | 473 | 32,203,549           |
| Apolipoprotein A-1                  | SD                                           | 393,193 European-descent individuals                | Age, sex, and genotyping chips                                                                                                                                                                                                                                                                                                                                                           | 383 | 32,203,549           |
| Low-density lipoprotein cholesterol | SD                                           | 440,546 European-descent individuals                | Age, sex, and genotyping chips                                                                                                                                                                                                                                                                                                                                                           | 199 | 32,203,549           |
| Triglycerides                       | SD                                           | 441,016 European-descent individuals                | Age, sex, and genotyping chips                                                                                                                                                                                                                                                                                                                                                           | 392 | 32,203,549           |
| Apolipoprotein B                    | SD                                           | 439,214 European-descent individuals                | Age, sex, and genotyping chips                                                                                                                                                                                                                                                                                                                                                           | 225 | 32,203,549           |
| **Outcome**                         |                                              |                                                      |                                                                                                                                                                                                                                                                                                                                                                                                     |     |                      |
| NAFLD                               | Odds ratio                                   | 8,434 NAFLD cases and 770,180 non-cases of European ancestry | Age, sex, and genetic principal components                                                                                                                                                                                                                                                                                                                                               | –   | 34,841,290           |
| NAFLD                               | Odds ratio                                   | 14,83 NAFLD cases and 17,781 non-cases of European ancestry | Top 5 genetic principal components                                                                                                                                                                                                                                                                                                                                                 | –   | 32,203,549           |

*ID* identifier, *IV* instrumental variables, *NAFLD* nonalcoholic fatty liver disease, *SD* standard deviation
in Supplementary Table 1. Detailed information on quality control refers to the cited GWAS papers [14, 15].

**Statistical analysis**

SNPs in the exposure and outcome datasets were harmonized by coded and reference alleles to omit ambiguous SNPs with non-concordant alleles. We defined palindromic SNPs with ambiguous minor allele frequency > 0.45 and < 0.55 and all possible palindromic SNPs were excluded in a sensitivity analysis. A few missing instruments were not replaced by proxy SNPs given that a small proportion of missing generates limited influences on the results.

The inverse variance weighted (IVW) method was used as the main statistical analysis method and supplemented by four sensitivity analyses, including the weighted median [16], MR-Egger [17], MR-PRESSO [18], and contamination mixture [19] methods. The assumptions and strengths of used methods are summarized in Supplementary Table 2. Given previously identified shared genetic risk between obesity and metabolic traits [20], we used multivariable MR analysis with adjustment for genetically predicted BMI to assess the independent effects of metabolic traits on NAFLD. Likewise, we performed multivariable MR analysis to disentangle the effects of different blood lipids fractions on NAFLD. The analysis of mediation of metabolic traits on the association between obesity and NAFLD was performed under network MR framework where multivariable MR analysis was applied to adjust for the genetic association of the instruments with BMI. The mediation effect was calculated using the formula: (total effect—direct effect)/total effect and standard error of the mediation estimate was calculated using the propagation of error method [21]. Reverse MR analyses were performed for type 2 diabetes, BMI, and blood lipids based on 6 SNPs as genetic instruments for NAFLD [22]. Cochran’s Q statistic was used to assess the heterogeneity of SNP-estimates in each MR association. The \( p \) value of intercept test from MR-Egger regression was used to assess the horizontal pleiotropy [17]. The association with the \( p \) value < 0.004 (0.05/14 exposures) were deemed a significant association, and the association with the \( p \) value ≥ 0.004 and ≤ 0.05 were regarded as a suggestive association. The \( F \) statistic was calculated to measure the strength of used instruments and power was estimated using an online tool [23]. All tests were two-sided and performed using the TwoSampleMR [24], MR-PRESSO [18] and MendelianRandomization [25] packages in the R software (version 4.0.2).

**Results**

The \( F \) statistic for instruments and estimated power for all analyses are shown in Table 2. All \( F \) statistics for the overall instruments were over 10, indicating a good strength of used genetic instruments. The power was low in the analysis of

| Table 2: \( F \) statistic and power estimation |
|---|---|---|---|---|---|---|---|---|---|
| Class | Exposure | SNPs | Variance explained | OR at 80% power for discovery analysis | OR at 80% power for replication analysis | OR at 80% power for combined analysis | \( F \) statistic |
| | | | | \( \geq 1.21 \) | \( \leq 0.80 \) | \( \geq 1.52 \) | \( \leq 0.53 \) | \( \geq 1.19 \) | \( \leq 0.81 \) |
| Lifestyle | Smoking initiation | 311 | 0.023 | | | | | | 60 |
| Lifestyle | Lifetime smoking index | 126 | NA | NA | NA | NA | NA | NA |
| Lifestyle | Alcohol drinking | 84 | 0.003 | \( \geq 1.55 \) | \( \leq 0.45 \) | \( \geq 2.45 \) | \( \leq 0.01 \) | \( \geq 1.52 \) | \( \leq 0.49 \) |
| Lifestyle | Coffee consumption | 12 | 0.005 | \( \geq 1.44 \) | \( \leq 0.57 \) | \( \geq 2.12 \) | \( \leq 0.05 \) | \( \geq 1.40 \) | \( \leq 0.60 \) |
| Lifestyle | Caffeine consumption | 2 | 0.013 | \( \geq 1.27 \) | \( \leq 0.73 \) | \( \geq 0.38 \) | \( \leq 0.73 \) | \( \geq 1.25 \) | \( \leq 0.75 \) |
| Lifestyle | Vigorous physical activity | 5 | NA | NA | NA | NA | NA | NA |
| Metabolic | Body mass index | 312 | 0.027 | \( \geq 1.19 \) | \( \leq 0.81 \) | \( \geq 1.48 \) | \( \leq 0.56 \) | \( \geq 1.18 \) | \( \leq 0.82 \) |
| Metabolic | Waist circumference | 44 | 0.012 | \( \geq 1.28 \) | \( \leq 0.72 \) | \( \geq 1.72 \) | \( \leq 0.35 \) | \( \geq 1.26 \) | \( \leq 0.74 \) |
| Metabolic | Type 2 diabetes | 497 | NA | NA | NA | NA | NA | NA |
| Metabolic | Fasting insulin | 91 | 0.036 | \( \geq 1.16 \) | \( \leq 0.84 \) | \( \geq 1.41 \) | \( \leq 0.62 \) | \( \geq 1.15 \) | \( \leq 0.85 \) |
| Metabolic | Fasting glucose | 58 | 0.005 | \( \geq 1.43 \) | \( \leq 0.57 \) | \( \geq 2.12 \) | \( \leq 0.05 \) | \( \geq 1.40 \) | \( \leq 0.60 \) |
| Metabolic | Systolic blood pressure | 228 | 0.050 | \( \geq 1.14 \) | \( \leq 0.86 \) | \( \geq 1.35 \) | \( \leq 0.67 \) | \( \geq 1.13 \) | \( \leq 0.87 \) |
| Metabolic | High-density lipoprotein cholesterol | 473 | 0.142 | \( \geq 1.08 \) | \( \leq 0.91 \) | \( \geq 1.21 \) | \( \leq 0.80 \) | \( \geq 1.08 \) | \( \leq 0.91 \) |
| Metabolic | Apolipoprotein A-1 | 383 | 0.119 | \( \geq 1.09 \) | \( \leq 0.91 \) | \( \geq 1.23 \) | \( \leq 0.78 \) | \( \geq 1.09 \) | \( \leq 0.91 \) |
| Metabolic | Low-density lipoprotein cholesterol | 199 | 0.078 | \( \geq 1.11 \) | \( \leq 0.89 \) | \( \geq 1.28 \) | \( \leq 0.73 \) | \( \geq 1.11 \) | \( \leq 0.89 \) |
| Metabolic | Triglycerides | 392 | 0.105 | \( \geq 1.10 \) | \( \leq 0.90 \) | \( \geq 1.24 \) | \( \leq 0.77 \) | \( \geq 1.10 \) | \( \leq 0.90 \) |
| Metabolic | Apolipoprotein B | 225 | 0.082 | \( \geq 1.11 \) | \( \leq 0.89 \) | \( \geq 1.27 \) | \( \leq 0.74 \) | \( \geq 1.11 \) | \( \leq 0.89 \) |

\( NA \) not available, \( OR \) odds ratio
alcohol, coffee, and caffeine consumption, but adequate for the other studied exposures.

Genetic predisposition to smoking was significantly associated with an increased risk of NAFLD in the discovery dataset and the association remained directionally consistent in the replication dataset (Fig. 1). The odds ratios (ORs) of NAFLD were 1.09 (95% confidence interval (CI) 1.02, 1.16; \( p = 0.008 \)) for genetically predicted one-standard deviation (SD) increase in prevalence of smoking initiation and 1.59 (95% CI 1.31, 1.93; \( p = 3.42 \times 10^{-5} \)) for genetically predicted one-SD change of lifetime smoking index in the combined analysis. Other studied lifestyle factors, including alcohol drinking, coffee and caffeine consumption, and vigorous physical activity, were suggestively inversely associated with the risk of NAFLD in the combined dataset (Fig. 2). The ORs of NAFLD were 0.61 (95% CI 0.38, 0.96; \( p = 0.032 \)) for genetically predicted one-SD increase of log-transformed alcoholic drinks/week, 0.74 (95% CI 0.55, 1.00; \( p = 0.047 \)) for genetically predicted 50% increase in coffee consumption, 0.87 (95% CI 0.75, 1.00; \( p = 0.050 \)) for genetically predicted 80 mg increase in caffeine consumption, and 0.77 (95% CI 0.61, 0.97; \( p = 0.026 \)) for genetic predisposition to vigorous physical activity.

Six out of nine metabolic factors were significantly associated with NAFLD risk in the combined dataset (Fig. 2). There were associations for BMI (OR 1.33, 95% CI 1.23, 1.43; \( p = 1.75 \times 10^{-12} \) per genetically predicted one-SD increase), waist circumference (OR 1.82; 95% CI 1.48, 2.24; \( p = 1.28 \times 10^{-8} \) per genetically predicted one-SD increase), type 2 diabetes (OR 1.21, 95% CI 1.15, 1.27; \( p = 7.15 \times 10^{-15} \) per one-unit increase in log-transformed odds), systolic blood pressure (OR 1.17; 95% CI 1.07, 1.26; \( p = 2.29 \times 10^{-4} \) per genetically predicted 10 mm Hg increase), high-density lipoprotein cholesterol (OR 0.84, 95% CI 0.77, 0.90; \( p = 6.88 \times 10^{-6} \) per genetically predicted one-SD increase), and triglycerides (OR 1.23, 95% CI 1.15, 1.33; \( p = 3.08 \times 10^{-8} \) per genetically predicted one-SD increase). In the reverse MR analysis, genetic liability to NAFLD was associated with lower levels of high-density lipoprotein cholesterol, but not associated with other blood lipids, BMI, or risk of type 2 diabetes (Supplementary Table 3). High heterogeneity was observed in these analyses.

The observed associations were overall consistent across sensitivity analyses and between discovery and replication datasets (Supplementary Table 4). Moderate-to-high heterogeneity was observed in the analyses for alcohol drinking, type 2 diabetes, and lipid traits (Table 3 and Supplementary Table 4). We detected possible pleiotropy in the analyses for lifetime smoking index, type 2 diabetes, low-density lipoprotein cholesterol in the discovery analysis and in the analysis for body mass index in the replication analysis (Table 3 and Supplementary Table 4). However, these associations remained consistent after removal of outlier variants in MR-PRESSO analysis (Table 3 and Supplementary Table 4).

The associations for type 2 diabetes, systolic blood pressure, and triglycerides, but not for high-density lipoprotein cholesterol, remained similar in MR analysis with adjustment for genetically predicted BMI (Figs. 2 and 3). The inverse association for genetically predicted levels of high-density lipoprotein cholesterol and the positive

---

Fig. 2 Associations of genetically predicted lifestyle and metabolic factors with risk of nonalcoholic fatty liver disease in discovery, replication, and combined datasets. CI confidence interval, OR odds ratio

\[
\begin{array}{llll}
\text{Exposure} & \text{Discovery (9,634 cases)} & \text{P-value} & \text{Replication (14,482 cases)} & \text{P-value} & \text{Combined effect} & \text{OR (95% CI)} & \text{P-value} \\
\hline
\text{Lifestyle factor} & & & & & & & \\
\text{Smoking initiation} & 1.24 (1.10, 1.39) & 3.58e-4 & 1.03 (0.96, 1.11) & 0.361 & 1.09 (1.02, 1.16) & 0.008 \\
\text{Lifetime smoking index} & 1.57 (1.25, 1.96) & 7.50e-5 & 1.67 (1.11, 2.53) & 0.014 & 1.59 (1.31, 1.93) & 3.42e-3 \\
\text{Alcohol drinking} & 0.72 (0.41, 1.25) & 0.241 & 0.41 (0.16, 0.95) & 0.036 & 0.61 (0.38, 0.96) & 0.032 \\
\text{Coffee consumption} & 0.73 (0.53, 1.01) & 0.055 & 0.78 (0.30, 2.00) & 0.600 & 0.74 (0.55, 1.00) & 0.047 \\
\text{Caffeine consumption} & 0.85 (0.71, 1.01) & 0.071 & 0.90 (0.70, 1.15) & 0.395 & 0.87 (0.75, 1.00) & 0.050 \\
\text{Vigorous physical activity} & 0.95 (0.74, 1.23) & 0.691 & 0.30 (0.17, 0.51) & 1.19e-5 & 0.77 (0.61, 0.97) & 0.026 \\
\text{Metabolic factor} & & & & & & & \\
\text{Body mass index} & 1.69 (1.49, 1.92) & 5.11e-15 & 1.14 (1.03, 1.26) & 0.010 & 1.33 (1.23, 1.43) & 1.75e-12 \\
\text{Waist circumference} & 1.78 (1.41, 2.24) & 1.02e-6 & 2.01 (1.26, 3.19) & 0.003 & 1.82 (1.48, 2.24) & 1.28e-8 \\
\text{Type 2 diabetes} & 1.22 (1.13, 1.30) & 4.86e-11 & 1.19 (1.10, 1.29) & 2.79e-5 & 1.21 (1.15, 1.27) & 7.15e-15 \\
\text{Fasting insulin} & 2.26 (1.49, 3.40) & 0.004 & 1.14 (0.78, 1.67) & 0.527 & 1.42 (1.03, 1.94) & 0.030 \\
\text{Fasting glucose} & 0.91 (0.68, 1.23) & 0.496 & 0.57 (0.32, 0.94) & 0.600 & 0.90 (0.73, 1.15) & 0.397 \\
\text{Systolic blood pressure} & 1.26 (1.12, 1.42) & 9.33e-5 & 1.06 (0.96, 1.21) & 0.181 & 1.17 (1.07, 1.26) & 2.28e-4 \\
\text{High-density lipoprotein cholesterol} & 0.76 (0.69, 0.85) & 7.22e-7 & 0.93 (0.83, 1.04) & 0.203 & 0.84 (0.77, 0.90) & 6.88e-6 \\
\text{Apolipoprotein A-1} & 0.85 (0.76, 0.96) & 0.011 & 0.98 (0.88, 1.09) & 0.729 & 0.92 (0.85, 1.00) & 0.053 \\
\text{Low-density lipoprotein cholesterol} & 0.93 (0.82, 1.07) & 0.327 & 0.98 (0.84, 1.14) & 0.816 & 0.96 (0.86, 1.06) & 0.376 \\
\text{Triglycerides} & 1.32 (1.20, 1.46) & 2.40e-8 & 1.13 (1.01, 1.26) & 0.033 & 1.23 (1.15, 1.33) & 3.08e-6 \\
\text{Apolipoprotein B} & 1.08 (0.95, 1.23) & 0.245 & 1.07 (0.91, 1.25) & 0.409 & 1.07 (0.97, 1.19) & 0.155 \\
\end{array}
\]
association for genetically predicted levels of triglycerides became stronger in multivariable MR analyses with adjustment for genetically predicted levels of other related lipid traits (Supplementary Table 5). The association between BMI and NAFLD attenuated after adjusting for genetic liability to type 2 diabetes, and genetically predicted levels of high-density lipoprotein cholesterol and triglycerides (Table 4). Genetic liability to type 2 diabetes mediated 51.4% (95% CI 13.4%-89.3%) of BMI-effects on NAFLD risk (Table 4).

**Table 3** Heterogeneity and pleiotropy assessment

| Stage          | Exposure                            | IVs | Cochran’s Q | Intercept | P for intercept | Outliers |
|----------------|-------------------------------------|-----|-------------|-----------|-----------------|----------|
| **Discovery**  |                                     |     |             |           |                 |          |
| Lifestyle factor | Smoking initiation                  | 311 | 345        | −0.003    | 0.524           | 0        |
|                | Lifetime smoking index              | 126 | 124        | 0.017     | 0.013           | 0        |
|                | Alcohol drinking                    | 84  | 165        | 0.009     | 0.329           | 4        |
|                | Coffee consumption                  | 12  | 19         | −0.009    | 0.552           | 0        |
|                | Caffeine consumption                | 2   | 0          | −         | −               | −        |
|                | Vigorous physical activity          | 5   | 2          | 0.071     | 0.303           | 0        |
| Metabolic factor | Body mass index                     | 312 | 342        | 0.001     | 0.720           | 0        |
|                | Waist circumference                 | 44  | 69         | −0.002    | 0.847           | 0        |
|                | Type 2 diabetes                     | 497 | 929        | 0.007     | 0.006           | 6        |
|                | Fasting insulin                    | 58  | 128        | 0.005     | 0.639           | 2        |
|                | Fasting glucose                    | 91  | 159        | 0.008     | 0.152           | 1        |
|                | Systolic blood pressure             | 228 | 188        | 0.001     | 0.907           | 0        |
|                | High-density lipoprotein cholesterol| 473 | 798        | −0.003    | 0.167           | 7        |
|                | Apolipoprotein A-1                  | 383 | 730        | −0.002    | 0.530           | 6        |
|                | Low-density lipoprotein cholesterol | 199 | 396        | 0.008     | 0.029           | 7        |
|                | Triglycerides                       | 392 | 583        | 0.004     | 0.065           | 4        |
|                | Apolipoprotein B                    | 225 | 415        | 0.006     | 0.095           | 6        |
| **Replication**|                                     |     |             |           |                 |          |
| Lifestyle factor | Smoking initiation                  | 311 | 333        | −0.003    | 0.590           | 0        |
|                | Lifetime smoking index              | 126 | 118        | 0.001     | 0.968           | 0        |
|                | Alcohol drinking                    | 84  | 151        | 0.005     | 0.719           | 3        |
|                | Coffee consumption                  | 12  | 60         | −0.012    | 0.871           | 1        |
|                | Caffeine consumption                | 2   | 2          | −         | −               | −        |
|                | Vigorous physical activity          | 5   | 1          | 0.107     | 0.530           | 0        |
| Metabolic factor | Body mass index                     | 312 | 416        | −0.004    | 0.045           | 0        |
|                | Waist circumference                 | 44  | 74         | −0.018    | 0.355           | 0        |
|                | Type 2 diabetes                     | 497 | 925        | −0.001    | 0.882           | 5        |
|                | Fasting insulin                    | 58  | 112        | 0.043     | 0.033           | 1        |
|                | Fasting glucose                    | 91  | 148        | 0.007     | 0.514           | 1        |
|                | Systolic blood pressure             | 228 | 268        | −0.006    | 0.493           | 2        |
|                | High-density lipoprotein cholesterol| 473 | 829        | −0.001    | 0.831           | 4        |
|                | Apolipoprotein A-1                  | 383 | 744        | −0.001    | 0.689           | 4        |
|                | Low-density lipoprotein cholesterol | 199 | 337        | 0.006     | 0.151           | 6        |
|                | Triglycerides                       | 392 | 685        | 0.003     | 0.178           | 8        |
|                | Apolipoprotein B                    | 225 | 337        | 0.003     | 0.410           | 5        |

*IVs* instrumental variables

**Discussion**

This MR study found that genetic predisposition to smoking, obesity, type 2 diabetes, high blood pressure, and dyslipidemia (low levels of low-density lipoprotein cholesterol and high levels of triglycerides) was associated with an increased risk of NAFLD. The associations for type 2 diabetes, systolic blood pressure, and triglycerides were independent of genetically predicted BMI. Low-density lipoprotein cholesterol and triglycerides appeared to be robust risk factors
Lifestyle and metabolic factors for nonalcoholic fatty liver disease: Mendelian randomization

for NAFLD risk among lipid biomarkers. Type 2 diabetes mediated a large proportion of BMI-effects on NAFLD risk. There were suggestive inverse associations of genetically predicted moderate alcohol drinking, coffee and caffeine consumption, and vigorous physical activity with NAFLD risk.

An animal study found that four-week cigarette smoking worsened liver injury in obese rats with histological features of NAFLD via increased oxidative stress [26]. Observational studies of smoking and NAFLD have been inconsistent, with no association in a cross-sectional study with 90 NAFLD patients [8]. In another cross-sectional study including 2811 participants, individuals with one more pack of cigarette smoked per day had 1% higher risk of NAFLD [7]. This positive association was confirmed in a subsequent cohort study of 199,468 Korean adults where current smoking, pack-years, and urinary cotinine levels (a marker of tobacco smoke exposure) were positively associated with the risk of incident NAFLD [27]. Being in line with this cohort, our study found robust MR associations of smoking initiation and lifetime smoking index with NAFLD risk in two independent datasets, which strengthened the causal nature of this association. In addition, passive smoking in child and early adult lives has been associated with an increased later-life risk of developing NAFLD [28]. The underlying mechanisms behind the association between smoking and NAFLD may related to insulin resistance, hyperinsulinemia, dyslipidemia, hepatic steatosis, inflammation, and increased levels of catecholamine and glucagon, which can be induced by long-term smoking and nicotine use [27, 29].

Epidemiological evidence on the association between alcohol drinking and NAFLD risk is conflicting. A three-year longitudinal study with 4 waves’ repeated measurements including up to 3773 Japanese adults found that light to moderate alcohol consumption was associated with a reduced risk of NAFLD in both sexes [30]. This inverse association was also identified in a recent meta-analysis of 14 observational studies [31]. However, moderate alcohol use compared to nondrinking was associated with less improvement in steatosis and level of aspartate transaminase in 285 NAFLD patients after a mean follow-up period of 47 months [9]. An MR study including 266 NAFLD cases and 200 non-cases found that lifetime moderate alcohol consumption proxied by one genetic variant located in alcohol...
dehydrogenase (ADHIB) gene had no beneficial effects on NAFLD disease severity [32]. Nevertheless, animal data revealed that aldehyde dehydrogenase 2 deficiency ameliorated alcoholic fatty liver but worsened liver inflammation and fibrosis in mice [33]. Our MR analysis found a borderline inverse association between moderate alcohol drinking and NAFLD. Additional studies with a large sample size, the ability to assess the nonlinear associations, and with a comprehensive consideration of confounders, especially healthy lifestyle factors, will be needed to verify our findings and elaborate on underlying mechanisms.

Habitual coffee consumption has been consistently associated with a reduced risk of NAFLD and the association appears to be in a dose–response way in observational studies [34]. A recent MR study found no statistically significant association of coffee consumption instrumented by 4 SNPs (OR 0.76; 95% CI 0.51, 1.14) and 6 SNPs (OR 0.77; 95% CI 0.48, 1.25) with NAFLD risk using data from the UK Biobank with 1122 cases (one sub-dataset of our analysis) [35]. Compared with this study, our MR analysis with larger power (12 SNPs explaining larger variance in coffee consumption and ~9 times more cases) revealed a possible inverse association between coffee consumption and NAFLD risk. Besides, we detected a possible inverse association for caffeine consumption.

An inverse association between physical activity and NAFLD risk was observed in a meta-analysis including 6 cohort and 4 case–control studies [36] and subsequent studies [37, 38]. Compared with individuals with lowest physical activity levels, those with highest levels (similar to vigorous physical activity) had a lower odds of developing NAFLD (risk ratio, 0.79; 95% CI 0.71, 0.89) [36]. This association was in line with our MR finding. Future well-powered MR analysis is warranted to verify our finding.

Metabolic disorders, like obesity [39], type 2 diabetes [10], increased systolic blood pressure [10], and dyslipidemia [40], have been associated with NAFLD in observational studies. However, whether these metabolic traits are causally associated with NAFLD risk is unknown given that most associations were based on observational data. A previous MR study with 1122 NAFLD cases found harmful causal effects of overall and central obesity (represented by BMI and waist-to-hip ratio adjusted for BMI, respectively), type 2 diabetes, and triglycerides levels on NAFLD [41]. These associations were replicated in this updated MR study with more cases. Notably, increased systolic blood pressure and decreased levels of low-density lipoprotein cholesterol were identified as two new causal risk factors for NAFLD in our analysis. Even though we had limited evidence in support of the reverse impact of having NAFLD on type 2 diabetes, BMI, and blood lipids except for high-density lipoprotein cholesterol, which is not in line with observational studies [42], these null findings should be cautiously interpreted given high heterogeneity in these analyses as well as a few genetic instruments for NAFLD [22]. Thus, the effects of NAFLD on metabolic profiles need to be further explored.

Metabolic traits are usually tightly correlated with overweight and obesity. Estimating BMI-independent effect of these metabolic traits is of clinical importance especially in identifying high-risk group especially given a large number of lean NAFLD patients [4]. By subtracting effects of BMI, our multivariable MR analysis found independent roles of type 2 diabetes, elevated systolic blood pressure, and increased levels of triglycerides in the development of NAFLD. These findings are in line with associations observed in the normal weight population [43] and meanwhile convey the information that it may be beneficial to promote NAFLD screening as well as lifestyle intervention among individuals with an abnormal profile of glycemic traits, blood pressure or triglycerides. We also performed multivariable MR analysis among lipid fractions to pinpoint the robust lipid biomarker associated with NAFLD [44]. We found that low-density lipoprotein cholesterol and triglycerides appeared to be better indicators compared to other studied lipid traits. Additionally, mediation analysis found that type 2 diabetes and dyslipidemia might be intermediators in the pathway from BMI to NAFLD. The finding highlights that a good management of glucose and lipids levels in obese individuals may be an effective way to somehow neutralize the detrimental effects of overweight and obesity on NAFLD.

MR analysis has three important assumptions [11]. First, the selected instrumental variables should be strongly associated with the exposure of interest (assumption 1, relevance). In the present study, we selected SNPs that were associated with the exposures at the genome-wide significance level ($P < 5 \times 10^{-8}$) as instrumental variables from genome-wide association studies with large sample sizes. Second, used instrumental variables should not be associated with any confounders in the association between the studied exposure and outcome (assumption 2, independence). Given the study was based on summary-level data, a thorough examination of the associations between exposures and possible confounders was not possible. However, these instruments were widely used in previous MR studies on metabolic diseases [45, 46]. Third, the genetic instruments should influence the outcome only via the exposure, not via other alternative pathways (assumption 3, exclusion restriction). Although we could not completely rule out the possibility that our findings might be biased by horizontal pleiotropy, our results remained consistent across several sensitivity analyses and the MR-Egger and MR-PRESSO analyses detected limited evidence in support of strong pleiotropic effects.

There are several strengths of this MR study. The major merit is MR design which can minimize confounding and
reverse causality to a large extent [11]. We explored the associations in two independent datasets to examine the consistency and then combined the associations from two data sources to increase the number of cases. Together with more SNPs used as instrumental variables compared to previous MR studies [32, 35, 41], our established associations should be better powered even though we might still have overlooked weak associations. The results remained overall consistent across several sensitivity analyses. In addition, we confined our analysis to the population of European descent, which effectively reduce the bias caused by the population structure bias. However, this study population of consistent ancestry may limit the generalizability of our findings to other populations.

Limitations need consideration when interpreting our results. The major issue for any MR study is horizontal pleiotropy that means selected genetic instrument variables influence the risk of outcome not via the exposure but other alternative pathways. However, this pleiotropic effect should not bias our results for two reasons. First, we detected limited evidence on pleiotropy from MR-Egger intercept test for most associations in the present MR analysis. For certain associations with significant indication of horizontal pleiotropy, there were few outliers detected by MR-PRESSO analysis and the association remained consistent or became even stronger after removal of outlying SNPs. Second, we performed multivariable MR for traits with strong phenotypic and genetic correlations and the associations remained stable after adjustment. Sample overlap between the exposure and outcome data sources might exist and thus bias our causal estimates towards observational associations by inflating the weak instrument bias [47]. Nonetheless, our SNPs were selected at the genome-wide threshold (strongly associated with the exposure) and all estimated F statistics were over 10, which indicates that the bias introduced by partial sample overlap should be minimal. Associations for certain exposures differed between the discovery and replication datasets, which might be caused population differences in certain features, like prevalence of obesity and vigorous physical activity. In addition, differences in NAFLD definition might cause heterogeneity in meta-analysis of associations across used data sources. For certain exposures, like alcohol consumption, the nonlinear association could not be estimated in the present MR analysis based on summary-level genetic statistics. Likewise, the gene-environmental interaction could not be assessed in summary-level data. The prevalence of different lifestyle and metabolic factors for nonalcoholic fatty liver disease: Mendelian randomization…

In conclusion, the present study provides MR data in support of causal roles of smoking, obesity, high systolic blood pressure, and dyslipidemia featured by low levels of low-density lipoprotein cholesterol and high levels of triglycerides, in the development of NAFLD. Mediation effects of type 2 diabetes and dyslipidemia in the association between BMI and NAFLD suggest the important role of the management of blood glucose and lipids in obesity in NAFLD prevention. The inverse associations for moderate alcohol drinking, coffee and caffeine consumption, and vigorous physical activity need confirmation in well-powered studies.

Supplementary Information The online version contains supplemental material available at https://doi.org/10.1007/s10654-022-00868-3.

Acknowledgements Genetic instruments for studied exposures were obtained from cited genome-wide association studies. Genetic association estimates for nonalcoholic fatty liver disease were obtained from the two published genome-wide association studies. The authors thank all investigators for sharing these data.

Author contributions Study conception and design: SY, JSZ and SCL; data acquisition and analysis: SY, BA, and SCL; drafting the manuscript and figures: SY; reviewing the manuscript: SY, JC, XL, BA, DG, ELG, JSZ and SCL.

Funding Open access funding provided by Karolinska Institute. This study is supported by research grants from the Swedish Heart–Lung Foundation (Hjärt-Lungfonden; Grant No. 20210351), the Swedish Research Council for Health, Working Life and Welfare (Forte; Grant No. 2018-00123), and the Swedish Research Council (Vetenskapsrådet; Grant No. 2019-00977). DG is supported by the British Heart Foundation Research Centre of Excellence (RE/18/4/34215) at Imperial College London and by a National Institute for Health Research Clinical Lectureship (CL-2020-16-001) at St George’s, University of London. Funders had no roles in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication.

Declarations

Conflict of interest DG is employed part-time by Novo Nordisk. All authors declare no support from any organization for the submitted work.

Ethical approval: All studies had been approved by a relevant ethical review board and participants had given informed consent.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.
References

1. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. Hepatology. 2016;64(1):73–84. https://doi.org/10.1002/hep.28431.

2. Kim WR, Lake JR, Smith JM, et al. OPTN/SRTK 2016 annual data report: liver. Am J Transplant. 2018;18(Suppl 1):172–253. https://doi.org/10.1111/ajt.14559.

3. Estes C, Anstee QM, Arias-Loste MT, et al. Modeling NAFLD disease burden in China, France, Germany, Italy, Japan, Spain, United Kingdom, and United States for the period 2016–2030. J Hepatol. 2018;69(4):896–904. https://doi.org/10.1016/j.jhep.2018.05.036.

4. Ye Q, Zou B, Yeo YH, et al. Global prevalence, incidence, and outcomes of non-obese or lean non-alcoholic fatty liver disease: a systematic review and meta-analysis. Lancet Gastroenterol Hepatol. 2020;5(8):739–52. https://doi.org/10.1016/s2468-1253(20)30077-7.

5. Cotter TG, Rinella M. Nonalcoholic fatty liver disease 2020: the state of the disease. Gastroenterology. 2020;158(7):1851–64. https://doi.org/10.1053/j.gastro.2020.01.052.

6. Brunt EM, Wong VW, Nobili V, et al. Nonalcoholic fatty liver disease. Nat Rev Dis Primers. 2015;1:15080. https://doi.org/10.1038/nrdp.2015.80.

7. Koehler EM, Schouten JN, Hansen BE, et al. Prevalence and risk factors of non-alcoholic fatty liver disease in the elderly: results from the Rotterdam study. J Hepatol. 2012;57(6):1305–11. https://doi.org/10.1016/j.jhep.2012.07.028.

8. Yilmaz Y, Yonal O, Kurt R, Avsar E. Cigarette smoking is not associated with specific histological features or severity of nonalcoholic fatty liver disease. Hepatology. 2010;52(1):391. https://doi.org/10.1002/heap.23718.

9. Ajmera V, Belt P, Wilson LA, et al. Among patients with nonalcoholic fatty liver disease, modest alcohol use is associated with less improvement in histologic steatosis and steatohepatitis. Clin Gastroenterol Hepatol. 2018;16(9):1511–20.e5. https://doi.org/10.1016/j.cgh.2018.01.026.

10. Lonardo A, Nascimbeni F, Mantovani A, Targher G. Hypertension, diabetes, atherosclerosis and NASH: cause or consequence? J Hepatol. 2018;68(2):335–52. https://doi.org/10.1016/j.jhep.2017.09.021.

11. Burgess S, Foley CN, Allara E, Staley JR, Howson JMM. A robust mendelian randomisation and meta-analysis of clinical trials. Hypertension. 2017;72(2):383–92. https://doi.org/10.1161/hypertensionaha.120.16547.

12. Yuan S, Larsson SC. Inverse association between serum 25-hydroxyvitamin D and nonalcoholic fatty liver disease. Clin Gastroenterol Hepatol. 2022. https://doi.org/10.1016/j.cgh.2022.01.021.

13. Brion MJ, Shahkhabazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. Int J Epidemiol. 2013;42(5):497–501. https://doi.org/10.1093/ije/dyt179.

14. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018. https://doi.org/10.7554/eLife.34408.

15. Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. Int J Epidemiol. 2017;46(6):1734–9. https://doi.org/10.1093/ije/dyx034.

16. Azzalini L, Ferrer E, Ramalho LN, et al. Cigarette smoking exacerbates nonalcoholic fatty liver disease in obese rats. Hepatology. 2010;51(5):1567–76. https://doi.org/10.1002/hep.23516.

17. Jung HS, Chang Y, Kwon MJ, et al. Smoking and the risk of non-alcoholic fatty liver disease: a cohort study. Am J Gastroenterol. 2019;114(5):453–63. https://doi.org/10.1038/s41395-018-0283-5.

18. Wu F, Pahkala K, Juonala M, et al. Childhood and adulthood passive smoking and nonalcoholic fatty liver in midlife: a 31-year cohort study. Am J Gastroenterol. 2021;116(6):1256–63. https://doi.org/10.1039/c9mg00000c0001141.

19. Malferttheiner P, Schütte K. Smoking—a trigger for chronic inflammation and cancer development in the pancreas. Am J Gastroenterol. 2006;101(1):160–2. https://doi.org/10.1111/j.1572-0241.2005.00402.x.

20. Moriya A, Iwasaki Y, Ohguchi S, et al. Roles of alcohol consumption in fatty liver: a longitudinal study. J Hepatol. 2015;62(4):921–7. https://doi.org/10.1016/j.jhep.2014.11.025.

21. Wongtrakul W, Niltwat S, Charachatronwitthaya P. The effects of modest alcohol consumption on non-alcoholic fatty liver disease: a systematic review and meta-analysis. Front Med (Lausanne). 2021;8:744713. https://doi.org/10.3389/fmed.2021.744713.

22. Sookoian S, Flichman D, Castaño GO, Pirola CJ. Mendelian randomisation suggests no beneficial effect of moderate alcohol consumption on the severity of nonalcoholic fatty liver disease. Aliment Pharmacol Ther. 2016;44(11–12):1224–34. https://doi.org/10.1111/apt.13828.

23. Kwon HJ, Won YS, Park O, et al. Aldehyde dehydrogenase 2 deficiency ameliorates alcoholic fatty liver but worsens liver
inflammation and fibrosis in mice. Hepatology. 2014;60(1):146–57. https://doi.org/10.1002/hep.27036.
34. Chen YP, Lu FB, Hu YB, Xu LM, Zheng MH, Hu ED. A systematic review and a dose-response meta-analysis of coffee dose and nonalcoholic fatty liver disease. Clin Nutr. 2019;38(6):2552–7. https://doi.org/10.1016/j.clnu.2018.11.030.
35. Zhang Y, Liu Z, Choudhury T, Cornelis MC, Liu W. Habitual coffee intake and risk for nonalcoholic fatty liver disease: a two-sample Mendelian randomization study. Eur J Nutr. 2021;60(4):1761–7. https://doi.org/10.1007/s00394-020-02369-z.
36. Qiu S, Cai X, Sun Z, et al. Association between physical activity and risk of nonalcoholic fatty liver disease: a meta-analysis. Therap Adv Gastroenterol. 2017;10(9):701–13. https://doi.org/10.1177/1756283x17725977.
37. Vilar-Gomez E, Nephew LD, Vuppalanchi R, et al. High-quality diet, physical activity, and college education are associated with low risk of NAFLD among the US population. Hepatology. 2021. https://doi.org/10.1002/hep.32207.
38. Li Y, He F, He Y, et al. Dose-response association between physical activity and non-alcoholic fatty liver disease: a case-control study in a Chinese population. BMJ Open. 2019;9(3):e026854. https://doi.org/10.1136/bmjopen-2018-026854.
39. Li L, Liu DW, Yan HY, Wang ZY, Zhao SH, Wang B. Obesity is an independent risk factor for non-alcoholic fatty liver disease: evidence from a meta-analysis of 21 cohort studies. Obes Rev. 2016;17(6):510–9. https://doi.org/10.1111/obr.12407.
40. Deprince A, Haas JT, Staels B. Dysregulated lipid metabolism links NAFLD to cardiovascular disease. Mol Metab. 2020;42:101092. https://doi.org/10.1016/j.molmet.2020.101092.
41. Liu Z, Zhang Y, Graham S, et al. Causal relationships between NAFLD, T2D and obesity have implications for disease subphenotyping. J Hepatol. 2020;73(2):263–76. https://doi.org/10.1016/j.jhep.2020.03.006.
42. Yki-Järvinen H. Non-alcoholic fatty liver disease as a cause and a consequence of metabolic syndrome. Lancet Diabetes Endocrinol. 2014;2(11):901–10. https://doi.org/10.1016/s2213-8587(14)70032-4.
43. Sookoian S, Pirola CJ. Systematic review with meta-analysis: risk factors for non-alcoholic fatty liver disease suggest a shared altered metabolic and cardiovascular profile between lean and obese patients. Aliment Pharmacol Ther. 2017;46(2):85–95. https://doi.org/10.1111/apt.14112.
44. Yuan S, Tang B, Zheng J, Larsson SC. Circulating lipoprotein lipids, apolipoproteins and ischemic stroke. Ann Neurol. 2020;88(6):1229–36. https://doi.org/10.1002/ana.25916.
45. Yuan S, Larsson SC. An atlas on risk factors for type 2 diabetes: a wide-angled Mendelian randomisation study. Diabetologia. 2020;63(11):2359–71. https://doi.org/10.1007/s00125-020-05253-x.
46. van Oort S, Beulens JWJ, van Balleghoijen AJ, Handoko ML, Larsson SC. Modifiable lifestyle factors and heart failure: a Mendelian randomization study. Am Heart J. 2020;227:64–73. https://doi.org/10.1016/j.ahj.2020.06.007.
47. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. Genet Epidemiol. 2016;40(7):597–608. https://doi.org/10.1002/gepi.21998.
48. Lonardo A, Nascimbini F, Ballestri S, et al. Sex differences in nonalcoholic fatty liver disease: state of the art and identification of research gaps. Hepatology. 2019;70(4):1457–69. https://doi.org/10.1002/hep.30626.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Shuai Yuan1 · Jie Chen2,3 · Xue Li4,5 · Rongrong Fan6 · Benoit Arsenault7,8 · Dipender Gill9,10,11,12 · Edward L. Giovannucci13,14 · Ju-sheng Zheng15 · Susanna C. Larsson1,16

1 Unit of Cardiovascular and Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Nobelsväg 13, 17177 Stockholm, Sweden
2 Centre for Global Health, Zhejiang University School of Medicine, Hangzhou, China
3 Department of Gastroenterology, The Third Xiangya Hospital, Central South University, Changsha, China
4 School of Public Health and the Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China
5 Centre for Global Health, Usher Institute, University of Edinburgh, Edinburgh, UK
6 Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden
7 Centre de Recherche de l’Institut Universitaire de Cardiologie et de Pneumologie de Québec, Quebec City, QC, Canada
8 Department of Medicine, Faculty of Medicine, Université Laval, Quebec, QC, Canada
9 Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK
10 Clinical Pharmacology and Therapeutics Section, Institute for Infection and Immunity, St George’s, University of London, London, UK
11 Clinical Pharmacology Group, Pharmacy and Medicines Directorate, St George’s University Hospitals NHS Foundation Trust, London, UK
12 Novo Nordisk Research Centre Oxford, Old Road Campus, Oxford, UK
13 Department of Epidemiology, Harvard T H Chan School of Public Health, Boston, MA, USA
14 Department of Nutrition, Harvard T H Chan School of Public Health, Boston, MA, USA
15 Key Laboratory of Growth Regulation and Translational Research of Zhejiang Province, School of Life Sciences, Westlake University, Hangzhou, China
16 Department of Surgical Sciences, Uppsala University, Uppsala, Sweden