**Supplementary Methods**

***Study participants***

UK Biobankis a large, prospective study that recruited more than 500,000 participants from 40 to 70 years of age1. All participants who provided written informed consent were interviewed using a touch-screen questionnaire to collect social demographic, lifestyle, and health-related information, as well as completed a range of anthropometric and physiological measurements. The NJHE Cohort is an ongoing population-based cohort study which recruited 12,007 participants ≥ 20 years of age in the Health Promotion Center of the First Affiliated Hospital of Nanjing Medical University between May 2019 and July 2021. All participants who provided written informed consent were interviewed using a structured questionnaire to collect sociodemographic characteristics, lifestyle behaviors, and personal and family medical histories; routine health examination was then processed.

***Definition of NAFLD outcome***

NAFLD diagnosis was based on the presence of three findings, (1) evidence of hepatic steatosis by either histology or imaging, (2) without heavy alcohol consumption (women >20 g/day, men >30 g/day in the UK; women > 10 g/day, men > 20 g/day in China), (3) without a history of specific diseases that could lead to steatosis, such as viral hepatitis, drug-induced liver disease, Wilson's disease and autoimmune liver disease2. In the UK Biobank, subjects were scanned in a Siemens MAGNETOM Aera 1.5-T MRI scanner (Siemens Healthineers, Erlangen, Germany), providing a single-slice multiecho Dixon acquisition for PDFF assessment in the liver3. Liver PDFF measured as the average PDFF in up to nine (at least three) regions of interest in the liver while avoiding any inhomogeneities, major vessels, and bile ducts. A high PDFF value denotes a high level of liver fat content and fatty liver was defined as PDFF ≥ 5%4. While in the NJHE Cohort, ultrasound was used to diagnose fatty liver by at least two professional physicians with the presence of at least two of three abnormal findings: diffusely increased echogenicity of the liver relative to the kidney, ultrasound beam attenuation, and poor visualization of intrahepatic structures.

***Assessment of smoking status***

Smoking status was classified as non-, current, former, and passive smokers. Nonsmokers were defined as those who smoked < 100 cigarettes during their whole lifetime. Current smokers were those who smoked on most or all days ≥ 12 months. Former smokers were those who smoked on most or all days in the past but quit smoking ≥ 6 months. Passive smokers in the UK Biobank were defined as those who were exposed to secondhand smoke for more than 1 h per week at home or outside, based on responses to questions of about how many hours per week are you exposed to other people's tobacco smoke at home and outside the home, about how many hours per week are you exposed to other people's tobacco smoke. While in the NJHE Cohort, passive smokers were defined based on whether they were exposed to secondhand smoke at home for more than 1 year, or in the workplace more than 1 h per day. If they answered in the affirmative, further information was obtained on details (total years of exposure at home, hours of exposure per day at work).

***Covariates***

We collected the information on age, sex, ethnicity, education level, physical activity, and drinking status at enrollment. Ethnicity was grouped as white ethnicity and other in the UK Biobank, Han nationality and other in the NJHE Cohort. Education level was classified as a college or university degree and other levels. Physical activity was assessed using adapted questions from the validated short International Physical Activity Questionnaire (IPAQ)5, regular physical activity indicates whether a person met the guidelines of 150 m of moderate activity per week or 75 m of vigorous activity in both cohorts. Drinking status was classified as never, former, and current, and the BMI (kg/m2) was calculated for each participant. Waist-hip ratio was calculated in the UK Biobank as waist circumference (cm)/hip circumference (cm). Blood samples were drawn after an overnight fast to measure serum triglycerides, high-density lipoprotein cholesterol (HDL-C) and fasting blood glucose using standard enzymatic methods. Cut-offs for stratified analysis were based on the definition of metabolic disorder6, 7: overweight (BMI ≥ 25 in the UK Biobank and ≥ 24 in the NJHE Cohort), triglycerides ≥ 1.7 mmol/L, HDL-C < 1.03 mmol/L in men and < 1.29 mmol/L in women, fasting blood glucose ≥ 5.6 mmol/L and waist-hip ratio ≥ 0.95 in men and ≥ 0.80 in women.

***Statistical analysis***

To estimate the dose-response relationship between smoking exposure with NAFLD prevalence, we further explored the associations of duration of secondhand smoke exposure, duration of smoking cessation and cumulative cigarette consumption with risk of NAFLD. Based on the significant results of the regression model, we performed mediation analysis for the association of smoking status with NAFLD using the following variables as potential mediators: BMI, triglycerides, fasting blood glucose, HDL-C, and waist-hip ratio, which were confirmed by previous studies to be associated with both cigarette exposure and NAFLD8, 9. Consistent with prior study10, we used a single binary variable and compared participants in the specific smoking group with the others when we calculated PAFs.

MR analysis between smoking initiation and NAFLD were performed by the multiplicative random-effects IVW method, which provides the most precise estimates though assuming that all SNPs are valid instruments. Regarding instrumental variable selection, single nucleotide polymorphisms (SNPs) associated with smoking initiation at the genome-wide significance threshold (*p* < 5 × 10–8) were identified in 607,291 individuals of European ancestry in the Sequencing Consortium of Alcohol and Nicotine use (GSCAN) study11. SNPs in linkage disequilibrium (defined as r2> 0.01 or clump distance < 1000 kb) and with the weaker associations with the exposure were removed. We used the UK Biobank genome-wide association study (GWAS) summary statistics of European ancestry conducted in by Lee lab12, where the NAFLD were defined by codes of the International Classification of Diseases 9th Revision (ICD-9) and ICD-10 (571.5, 571.8, 571.9, K74.0, K74.1, K74.2, K76.0). Genetic associations were estimated by logistic regression with adjustment for sex, birthdate, and the first four genetic principal components. Data were harmonized to omit ambiguous SNPs with nonconcordant alleles and palindromic SNPs with ambiguous minor allele frequency (> 0.42 and < 0.58) were removed from the analysis, leaving 102 independent SNPs as instrumental variables. To detect potential unbalanced pleiotropy (horizontal pleiotropy) and examine the consistency of the associations, three extra analyses including the weighted median, MR Egger, and MR pleiotropy residual sum and outlier (MR-PRESSO) analyses were performed.13,14 The MR Egger intercept test detects unmeasured pleiotropy and the MR-PRESSO method can identify SNP outliers and provide results identical to that from IVW after removal of outliers.

In case of missing information, sex-specific means were imputed for continuous covariates and a missing indicator was used for categorical ones (all covariates ≤ 5% missing except for education level (8.12%), physical activity (15.28%), pack-years of smoking (5.89%), fasting blood glucose (15.29%), triglycerides (7.28%), and HDL-C (15.23%) in UK Biobank. We performed a series of sensitivity analyses to test the stability of the results. First, we restricted the analysis to participants with complete covariates. Second, in order to consist with the definition in UK Biobank, we redefined passive smoker as those who were exposed to secondhand smoke at home for more than 1 year, or in the workplace more than 1 h per week in NJHE Cohort. Third, we restricted the analysis to those aged 40 years or older in NJHE Cohort to coincide with the age distribution in UK Biobank, and to reduce the bias that the risk of NAFLD is relatively lower in younger adults. Fourth, we adopted a more rigorous exclusion criteria that we excluded participants with cancer, respiratory diseases (asthma, chronic bronchitis, emphysema, chronic pneumonia), cardiovascular disease (stroke, heart attack, angina), or diabetes at baseline, to ensure the specificity of smoking with NAFLD.

All *p*-values were two-sided and *p*< 0.05 was considered statistically significant. Statistical analyses in this study were performed using STATA 16 (STATA Corp, College Station, TX, USA). The mediation analysis was fitted using the R package Lavaan and the MR analysis was assessed by the R package TwoSampleMR and MR-PRESSO in R 4.1.0 (R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/).

**References**

1. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 2015;12(3):e1001779. doi: 10.1371/journal.pmed.1001779. PMID: 25826379.

2. Cotter TG, Rinella M. Nonalcoholic Fatty Liver Disease 2020: The State of the Disease. Gastroenterology 2020;158(7):1851-1864. doi: 10.1053/j.gastro.2020.01.052. PMID: 32061595.

3. Linge J, Borga M, West J, Tuthill T, Miller MR, Dumitriu A, et al. Body Composition Profiling in the UK Biobank Imaging Study. Obesity 2018;26(11):1785-1795. doi: 10.1002/oby.22210. PMID: 29785727.

4. Caussy C, Alquiraish MH, Nguyen P, Hernandez C, Cepin S, Fortney LE, et al. Optimal threshold of controlled attenuation parameter with MRI-PDFF as the gold standard for the detection of hepatic steatosis. Hepatology 2018;67(4):1348-1359. doi: 10.1002/hep.29639. PMID: 29108123.

5. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003;35(8):1381-1395. PMID: 12900694.

6. Alberti KGMM, Zimmet P, Shaw J. The metabolic syndrome--a new worldwide definition. Lancet 2005;366(9491):1059-1062. PMID: 16182882.

7. Lear SA, James PT, Ko GT, Kumanyika S. Appropriateness of waist circumference and waist-to-hip ratio cutoffs for different ethnic groups. Eur J Clin Nutr 2010;64(1):42-61. doi: 10.1038/ejcn.2009.70. PMID: 19672278.

8. Willi C, Bodenmann P, Ghali WA, Faris PD, Cornuz J. Active smoking and the risk of type 2 diabetes: a systematic review and meta-analysis. JAMA 2007;298(22):2654-2664. PMID: 18073361.

9. Byrne CD, Targher G. NAFLD: a multisystem disease. J Hepatology 2015;62(1 Suppl):S47-S64. doi: 10.1016/j.jhep.2014.12.012. PMID: 25920090.

10. Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. Am J Public Health 1998;88(1):15-19. PMID: 9584027.

11. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nat Genet 2019;51(2):237-244. doi: 10.1038/s41588-018-0307-5. PMID: 30643251.

12. Zhou W, Zhao Z, Nielsen JB, Fritsche LG, LeFaive J, Gagliano Taliun SA, et al. Scalable generalized linear mixed model for region-based association tests in large biobanks and cohorts. Nat Genet 2020;52(6):634-639. doi: 10.1038/s41588-020-0621-6. PMID: 32424355.

13. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet 2018;50(5):693-698. doi: 10.1038/s41588-018-0099-7. PMID: 29686387.

14. Wootton RE, Richmond RC, Stuijfzand BG, Lawn RB, Sallis HM, Taylor GMJ, et al. Evidence for causal effects of lifetime smoking on risk for depression and schizophrenia: a Mendelian randomisation study. Psychol Med 2020;50(14):2435-2443. doi: 10.1017/S0033291719002678. PMID: 31689377.