# Shared genetic architecture and mendelian randomisation between simvastatin-related medication use and cholelithiasis: a multi-method GWAS-based analysis

Danxi Feng[1], Xia Tong[2], Zhicun Wang[1], Xin Zhang[2]*

[1]GCP Office, The First People's Hospital of Shuangliu District (West China Airport Hospital of Sichuan University), Chengdu, China
[2]Department of Gastroenterology, The First People's Hospital of Shuangliu District (West China Airport Hospital of Sichuan University), Chengdu, China

**Submitted:** 14 April 2026; **Accepted:** 24 September 2026
**Online publication:** 21 October 2026

Arch Med Sci 2026; 22 (5): 2905–2915
DOI: https://doi.org/10.5114/aoms/240289
Copyright © 2026 Termedia & Banach

***Corresponding author:**
Xin Zhang PhD
Department of Gastroenterology
The First People's Hospital of Shuangliu District (West China Airport Hospital of Sichuan University)
120, Chengbei Shangjie St
Dongsheng Town
Shuangliu District
Chengdu 610200, China
E-mail: xinzhang201618@gmail.com

## Abstract

**Introduction:** Statins are widely prescribed, but their relationship with cholelithiasis (CL) remains uncertain. We evaluated shared genetic architecture and MR associations between CL and UK Biobank medication-use phenotypes for simvastatin (SIM) and cholesterol-lowering medication (LLA).
**Material and methods:** GWAS summary statistics were analysed using LDSC, Genomic SEM, MAGMA, univariable MR, lipid-adjusted multivariable MR (MVMR), and cis-drug-target MR for HMGCR, NPC1L1, and PCSK9. MVMR adjusted SIM and LLA separately for LDL-C, HDL-C, and triglycerides.
**Results:** CL showed inverse genetic correlations with SIM (rg = –0.277, $p = 1.54 \times 10^{-4}$) and LLA (rg = –0.223, $p = 2.47 \times 10^{-3}$). Gene and pathway analyses implicated cholesterol–bile acid metabolism and enterohepatic signalling. On the native linear-model outcome scale, univariable MR supported an inverse SIM association (β = –0.050, 95% CI: –0.077 to –0.023, $p < 0.001$), whereas LLA was not significant (β = –0.029, 95% CI: –0.062 to 0.004, $p = 0.090$). MVMR direct effects were near null for SIM (β = 0.0051, $p = 0.858$) and LLA (β = 0.0060, $p = 0.631$), with conditional $F$-statistics < 10 and marked heterogeneity. Cis-drug-target MR showed no clear association for HMGCR, NPC1L1, or PCSK9.
**Conclusions:** Genetic liability to simvastatin medication use was modestly inversely associated with CL in univariable MR but not after lipid adjustment. Together, these findings support a potential protective genetic relationship between simvastatin-related medication use and CL and highlight lipid-related and target-specific mechanisms for further mechanistic and clinical validation.

**Key words:** cholelithiasis, simvastatin, cholesterol-reducing agents, medication-use phenotype, drug-target Mendelian randomisation, genetic correlation, genomic SEM, MAGMA, bile acid metabolism, FXR signalling.

## Introduction

Cholelithiasis (CL) is a common hepatobiliary disorder in which cholesterol supersaturation, impaired gallbladder motility, and dysregulated bile acid metabolism promote gallstone formation [1–4]. Because cholesterol-rich bile is central to most gallbladder stones, genetic and pharmacologic factors that influence systemic cholesterol synthesis, intestinal

Creative Commons licenses: This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY -NC -SA 4.0). License (http://creativecommons.org/licenses/by-nc-sa/4.0/).

cholesterol absorption, biliary cholesterol secretion, and bile acid cycling are biologically relevant to gallstone susceptibility.

Statins are widely prescribed lipid-lowering drugs, yet their relationship with gallstone disease remains clinically unsettled. A simple mechanistic chain is plausible: inhibition of hepatic 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGCR) lowers cholesterol synthesis and circulating low-density lipoprotein cholesterol (LDL-C), which may in turn reduce hepatic cholesterol availability for biliary secretion and decrease biliary lithogenicity [5–7]. This pathway is also connected to bile acid synthesis, intestinal bile acid reabsorption, and farnesoid X receptor (FXR)-regulated enterohepatic feedback.

Observational studies and meta-analyses have reported inverse, null, and exposure-duration-dependent associations between statin use and gallstone outcomes [8–10]. Prior drug-target MR found that genetically proxied HMGCR inhibition was associated with a greater reduction in gallstone risk (OR ≈ 0.72 per SD decrease in LDL-C), whereas proxies for other lipid-lowering targets showed different directions of effect [11]. The smaller effect observed in the present study is therefore not expected to be directly comparable, because our exposures are UK Biobank medication-use phenotypes rather than molecular HMGCR inhibition or measured on-treatment LDL-C response.

Pharmacologic specificity further supports the need to separate medication classes. Statins and HMGCR inhibition may reduce cholesterol synthesis, whereas fibrates can increase lithogenicity by suppressing bile acid synthesis through peroxisome proliferator-activated receptor-α–mediated downregulation of cholesterol 7α-hydroxylase and sterol 27-hydroxylase [12–14]. Thus, a broad cholesterol-reducing medication phenotype may combine agents with different or even opposing hepatobiliary effects.

No genome-wide study has directly asked whether genetic variants associated with SIM-related medication use also predict lower CL risk while simultaneously mapping shared genes and pathways. Accordingly, we used a multi-method GWAS-based framework – including LDSC, Genomic SEM, MAGMA, medication-use MR, MVMR, and complementary cis-drug-target MR – to quantify shared genetic architecture, prioritise biological pathways, evaluate medication-use liability, and directly compare LDL-C lowering proxied through HMGCR, NPC1L1, and PCSK9. Recent work in *Archives of Medical Science* has likewise combined two-sample MR with NHANES analyses to investigate a modifiable determinant of gallstone disease, illustrating the expanding use of causal-genetic designs in this field [15].

## Material and methods

### Data sources

Genetic instruments and cross-trait analyses were based on publicly available UK Biobank GWAS summary statistics, with exact phenotype definitions provided in Supplementary Table SI. The CL dataset (OpenGWAS ID: ukb-a-559; *N* = 337,199; 6986 cases and 330,213 controls) has the exact trait label 'Diagnoses – main ICD10: K80 Cholelithiasis'; cases had a main-diagnosis K80 record and controls did not have that record in the analysed GWAS. The SIM dataset (ukb-b-11268; *N* = 462,933; 52,427 cases and 410,506 controls) has the exact label 'Treatment/medication code: simvastatin' and was derived from UK Biobank field 20003, code 1140861958, which records regularly used self-reported medications during the assessment-centre interview. SIM cases carried this code and controls did not. The LLA dataset (ukb-b-11740; *N* = 209,638; 48,450 cases and 161,188 controls) has the exact label 'Medication for cholesterol, blood pressure or diabetes: Cholesterol lowering medication'. It is a PHESANT-derived, class-level binary medication-use phenotype, and the OpenGWAS record explicitly reports the analysed sample as including males and females. Cases selected the cholesterol-lowering medication response, and controls did not select that response in the analysed PHESANT sample. The public GWAS metadata do not identify the agent taken, and the available summary statistics cannot determine the proportions attributable to statins, fibrates, ezetimibe, or other lipid-lowering drugs. Accordingly, SIM and LLA are interpreted as liabilities to self-reported medication use, not as measures of drug efficacy, adherence, dose, duration, on-treatment lipid change, or target-specific pharmacologic inhibition.

### LDSC

GWAS summary statistics for SIM-related medication use, LLA use, and CL were analysed using bivariate linkage disequilibrium score regression (LDSC) [16]. Precomputed linkage disequilibrium (LD) scores from the 1000 Genomes Project European reference panel were applied. LDSC was used to estimate SNP-based heritability and genetic correlation between traits, a method robust to sample overlap [17]. Analyses were conducted using the LDSC software with the baseline-LD model. The regression framework accounted for LD-related confounding and genomic inflation, with statistical significance of genetic correlations (rg) assessed using standard errors (SEs) from the regression estimates.

### Genomic SEM

Genomic structural equation modelling (Genomic SEM) was used as a descriptive multivariate representation of the LDSC-derived covariance matrix for SIM medication use, LLA use, and CL [18]. The model asked whether one common genetic dimension could parsimoniously summarise the pairwise covariance pattern; it was not used as an independent causal test. Exploratory factor analysis informed the one-factor specification, and confirmatory analysis reported the standardised loadings and model-fit indices. SIM was used as the anchor indicator, with its loading fixed at 0.99 for model identification after alternative specifications showed a near-unity SIM loading. This fixed value is not a freely estimated effect size and cannot be interpreted as evidence that SIM is biologically dominant. Because the model contains only three observed indicators and an identification constraint, its incremental contribution beyond LDSC is limited to a joint, compact representation of the covariance structure; fit indices and near-unity loadings were therefore interpreted cautiously.

### MAGMA

Multi-marker Analysis of GenoMic Annotation (MAGMA v1.10) was used to identify gene- and pathway-level genetic associations between drug-related medication-use phenotypes and CL [19]. Gene-level association analyses were conducted using multiple linear principal component regression models. SNP summary statistics were mapped to 19,205 protein-coding genes based on genomic position, with LD accounted for using the 1000 Genomes Project European reference panel [20]. For exploratory shared-gene prioritisation, genes were retained when they showed nominal evidence of association (MAGMA $p < 0.05$) in both traits of a pair; these values should not be interpreted as Bonferroni-significant gene-level discoveries. Competitive gene set enrichment analysis was subsequently performed to identify overrepresented biological pathways using curated gene sets from the Molecular Signatures Database (MSigDB). Tissue-specific enrichment analyses were conducted across 54 tissues using the Genotype-Tissue Expression (GTEx) v8 tissue expression reference.

### MR

Two-sample univariable MR analyses assessed associations of genetic liability to SIM medication use and LLA use with CL risk. The analysis was interpreted under the three core instrumental-variable assumptions: the variants are associated with the exposure (relevance), are not associated with exposure–outcome confounders (independence), and influence CL only through the exposure pathway (exclusion restriction). Exposure-associated SNPs were selected at $p < 5 \times 10^{-8}$ and clumped against the 1000 Genomes European reference panel at $r^2 < 0.001$ within a 10,000-kb window. Effect alleles were harmonised across exposure and outcome datasets, and incompatible or strand-ambiguous variants were excluded according to the harmonisation rules. For each exposure-associated SNP, the variance explained was calculated as $R^2 = 2 \times \beta X^2 \times EAF \times (1 - EAF)$, and the per-variant $F$-statistic was calculated as $F = (N - 2) \times R^2/(1 - R^2)$, where EAF was the effect-allele frequency and $N$ was the corresponding exposure GWAS sample size. When per-SNP sample size was unavailable, N was obtained from OpenGWAS metadata using the reported total sample size or the sum of cases and controls. SNPs with $F > 10$ were retained. Seven SIM SNPs ($F$ range: 11.96–45.22) and 22 LLA SNPs ($F$ range: 10.57–72.40) passed the exposure-side selection criteria and are listed in Supplementary Table SV. After outcome extraction and harmonisation, all seven SIM SNPs and 21 LLA SNPs entered the primary MR analysis; the selection workflow is summarised in Supplementary Table SV A. The primary estimator was multiplicative random-effects IVW [21], with weighted median [22] and MR-Egger [23] as sensitivity estimators. Cochran's Q assessed heterogeneity, and the MR-Egger intercept assessed directional pleiotropy. Because only seven SIM instruments were available, the MR-Egger analysis and intercept test were considered low powered. The medication-use analyses and the separate cis-drug-target analyses estimate different exposures and were therefore interpreted independently. The binary GWAS exposure scale has no direct clinical unit and is not interpretable as a dose, duration, or treatment-initiation effect. Because the Neale Lab CL GWAS was fitted using linear regression for a binary outcome, native outcome-scale β values and their 95% confidence intervals were primary; exponentiating these coefficients does not produce formal odds ratios.

### Classic cis-drug-target MR and molecular-QTL triangulation

To proxy pharmacologic LDL-C lowering at established drug targets, we used the European, males-and-females LDL-C GWAS (ieu-b-110; $N = 440{,}546$) [24] and the fixed CL outcome (ukb-a-559). GRCh37 cis windows were defined as ±100 kb around HMGCR, NPC1L1, and PCSK9. Within each window, variants were required to associate with LDL-C at $p < 5 \times 10^{-8}$, have minor-allele frequency $\geq 0.01$, and have $F > 10$. The primary instrument set retained partially correlated variants after 1000

Genomes European clumping at $r^2 < 0.10$; signed LD matrices aligned to the LDL-C effect allele were incorporated in generalised IVW with multiplicative random-effects standard errors. Missing outcome associations were replaced only by European proxies at $r^2 \geq 0.80$. Effects were oriented to a 1-SD decrease in LDL-C. A strict independent-instrument analysis used $r^2 < 0.001$, and Holm correction was applied across the three targets.

For molecular triangulation, cis-eQTLs from eQTLGen whole blood [25], relevant-tissue GTEx v8 data [26], and plasma cis-pQTLs from INTERVAL [27] were screened in ±1-Mb windows using the same *p*-value, minor-allele-frequency, and *F*-statistic criteria. MR was performed only when eligible instruments were available; unavailable or subthreshold target–QTL combinations were reported rather than forced into estimation. HMGCR eQTL instruments were additionally tested against LDL-C as a positive control. Regional colocalisation used ±500-kb windows and Wakefield approximate Bayes factors [28], with $p1 = p2 = 1 \times 10^{-4}$ and $p12 = 1 \times 10^{-5}$; $p12$ values of $5 \times 10^{-6}$ and $5 \times 10^{-5}$ were examined as prior-sensitivity analyses. Because the Neale Lab binary CL GWAS was fitted using linear regression, native linear-model β values were the primary effect scale. First-order logit conversions using $\beta/[K(1 - K)]$, where $K = 6986/337{,}199$, were reported only as sensitivity estimates and were not treated as formal odds ratios. Full specifications and instrument-level results are provided in Supplementary Tables SIX, SX. Additional harmonised instrument-level data, analysis-ready files, and analysis code are available from the corresponding author upon reasonable request.

### Multivariable Mendelian randomisation

To assess whether the medication-use associations persisted after accounting for established lipid traits, we prespecified two separate four-exposure MVMR models: SIM (ukb-b-11268) plus LDL-C (ieu-b-110; $N = 440{,}546$), HDL-C (ieu-b-109; $N = 403{,}943$), and triglycerides (ieu-b-111; $N = 441{,}016$), and LLA (ukb-b-11740) plus the same three lipid traits, with CL (ukb-a-559) as the outcome. All datasets included European-ancestry males and females. Within each model, variants associated with any exposure at $p < 5 \times 10^{-8}$ were pooled, clumped against the 1000 Genomes European reference at $r^2 < 0.001$ within 10,000 kb both within and across traits, and re-extracted for every exposure and the outcome. Missing associations were sought using proxies at $r^2 \geq 0.8$. Harmonisation used action 2; incompatible, multiallelic, and palindromic variants with ambiguous minor-allele frequency (> 0.42) were excluded. Direct effects were estimated by multivariable IVW-weighted least squares through the origin with inverse outcome-variance weights, equivalent to TwoSampleMR::mv_multiple (intercept = FALSE, instrument_specific = FALSE) and MVMR::ivw_mvmr [29]. Because the same linear-model CL outcome was used, MVMR direct effects were reported as native β values with 95% confidence intervals and *p*-values; exponentiated coefficients were not interpreted as odds ratios.

Conditional instrument strength was assessed using MVMR::strength_mvmr, residual heterogeneity using the multivariable Q statistic, directional pleiotropy using the MVMR-Egger intercept, and collinearity using variance inflation factors and the condition number [30]. Sensitivity analyses excluded variants involving a proxy, all palindromic variants, and influential variants with Cook's distance > $4/n$. Because the medication and lipid GWASs overlap in UK Biobank and per-SNP exposure-effect covariances were unavailable, conditional F and Q were calculated under zero covariance ($\rho = 0$) and repeated across equicorrelation assumptions from $\rho = -0.20$ to 0.30. The MVMR-IVW point estimates do not depend on the assumed covariance, whereas conditional F and Q do; these diagnostics were therefore interpreted as sensitivity scenarios rather than known-covariance estimates.

## Results

### LDSC

Bivariate LDSC identified inverse genetic correlations between CL and both medication-use phenotypes. The genetic correlation was $rg = -0.2774$ (SE = 0.0733, $p = 1.54 \times 10^{-4}$) for SIM–CL and $rg = -0.2228$ (SE = 0.0736, $p = 2.47 \times 10^{-3}$) for LLA–CL (Table I). These estimates indicate modest inverse genetic overlap: alleles associated with the medication-use phenotypes tended to be associated with lower genetic liability to CL.

### Genomic SEM

Genomic SEM provided a one-factor representation of the LDSC covariance matrix (CFI = 0.995; SRMR = 0.077). SIM served as the reference indicator, and its loading was fixed at 0.99

**Table I.** LD score regression results

| Trait pair | *rg* | SE | *P*-value |
|---|---|---|---|
| SIM-CL | –0.2774 | 0.0733 | 0.00015405 |
| LLA-CL | –0.2228 | 0.0736 | 0.002468519 |

for identification; this value was not estimated. LLA loaded positively on the factor (β = 0.99, SE = 0.08), whereas CL had a smaller inverse loading (β = –0.25, SE = 0.07). Thus, the model summarises the two medication-use traits and CL within a common genetic dimension, with CL oriented in the opposite direction. Taken together, the SEM model provides a compact representation of the LDSC-derived genetic covariance structure, showing that the two medication-use phenotypes and CL can be summarised by a single shared genetic dimension, with CL loading in the opposite direction. Given the three-indicator structure and fixed loading, this result is descriptive and does not provide independent causal evidence beyond the pairwise LDSC correlations (Figure 1).

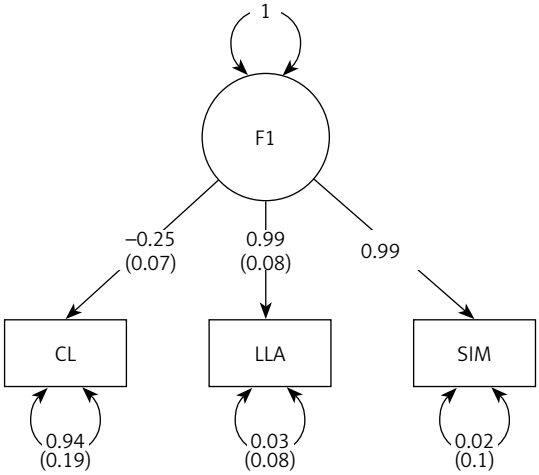



**Figure 1.** Genomic SEM results. Diagram depicting the standardised factor loadings (path coefficients ± standard error) of a single common latent genetic factor (F1) on three traits: CL, LLA, and SIM

## MAGMA

MAGMA analyses identified shared nominal gene-level signals between drug-related medication-use phenotypes and CL. For SIM–CL, 299 genes showed MAGMA $p < 0.05$ in both traits, and for LLA–CL, 353 genes met the same exploratory paired-trait threshold. Prominent biologically relevant genes included HMGCR, ABCG5, ABCG8, CYP7A1, SLCO1B1, and NPC1L1; genomic coordinates and full gene-level statistics are provided in Supplementary Table SII rather than being emphasised in the main text.

Pathway enrichment analyses supported convergent involvement of cholesterol synthesis, cholesterol transport, bile acid synthesis, enterohepatic circulation, FXR signalling, and bile acid signalling pathways. For SIM–CL, 135 pathways were enriched, whereas 141 pathways were enriched for LLA–CL; database-specific pathway identifiers are listed in Supplementary Table SIII. Tissue enrichment highlighted the terminal ileum and spleen for SIM–CL, while LLA–CL enrichment was confined to the terminal ileum. The terminal ileum finding is biologically consistent with intestinal bile acid reabsorption and FXR–FGF15/19 feedback. Gene-level Manhattan plots are shown in Figure 2, and tissue-specific enrichment results are presented in Supplementary Figure S1 and Supplementary Table SIV.

## MR

Univariable MR supported a modest inverse association between genetic liability to SIM medication use and CL on the native linear-model outcome scale. The IVW estimate was β = –0.0501 (95% CI: –0.0774 to –0.0229, $p < 0.001$) using seven instruments with F-statistics of 11.96–45.22. The weighted median estimate was similar (β = –0.0501, 95% CI: –0.0844 to –0.0158, $p = 0.004$), whereas MR-Egger was directionally consistent but imprecise (β = –0.0283, 95% CI: –0.0929 to 0.0362, $p = 0.429$). Heterogeneity was not detected (Cochran's Q = 2.8085, d$f$ = 6, $p = 0.832$), and the MR-Egger intercept was not significant ($p = 0.498$); however, both pleiotropy assessments have limited power with seven instruments. Because the CL GWAS used linear regression for a binary outcome, these β values are primary and their exponentiation is not interpreted as a formal odds ratio.

For the broader LLA phenotype, 22 exposure-side instruments passed the selection criteria with *F*-statistics of 10.57–72.40; 21 harmonised SNPs entered the primary MR analysis. The primary IVW estimate was directionally inverse but not statistically significant (β = –0.0285, 95% CI: –0.0616 to 0.0045, $p = 0.090$). The weighted median estimate was β = –0.0172 (95% CI: –0.0303 to –0.0042, $p = 0.009$), and MR-Egger was null (β = –0.0144, 95% CI: –0.0854 to 0.0566, $p = 0.694$). Substantial heterogeneity was present (Cochran's Q = 239.823, d$f$ = 20, $p < 0.001$), whereas the MR-Egger intercept was not significant ($p = 0.664$). The SIM and LLA confidence intervals overlap, and no formal between-exposure difference test was performed; therefore, statistical significance for SIM but not LLA is not evidence that their MR estimates differ. Full univariable MR and sensitivity statistics are provided in Supplementary Tables SVI and SVII (Figure 3).

## Classic cis-drug-target MR and molecular-QTL triangulation

The primary LDL-C cis analyses retained 11 HMGCR, 2 NPC1L1, and 19 PCSK9 instruments after outcome extraction and harmonisation. Per 1-SD lower LDL-C, the native-scale gener-

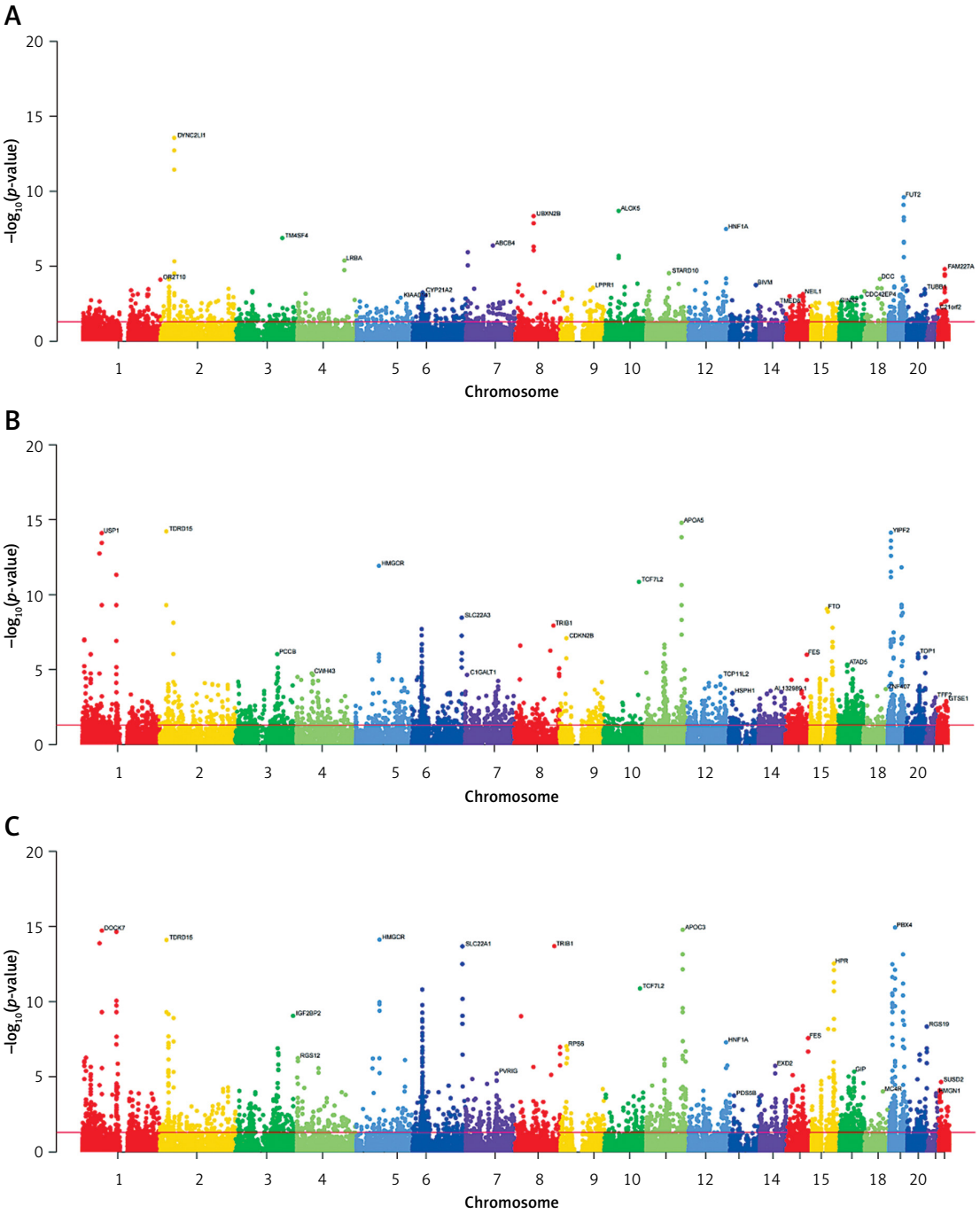



**Figure 2.** MAGMA gene-level Manhattan plots. Manhattan plots showing the distribution of MAGMA gene-level association p-values across autosomes (chromosomes 1–22) for CL (**A**), SIM-related medication use (**B**), and LLA use (**C**)

alised-IVW estimates were β = –0.000596 (95% CI: –0.011414 to 0.010222; $p$ = 0.914) for HMGCR, β = –0.004908 (95% CI: –0.063239 to 0.053424; $p$ = 0.869) for NPC1L1, and β = 0.002780 (95% CI: –0.003216 to 0.008776; $p$ = 0.363) for PCSK9. Holm-adjusted $p$-values were 1.000 for all three targets. NPC1L1 was imprecisely estimated and showed heterogeneity (Q = 6.37, $p$ = 0.0116). The strict $r^2$ < 0.001 analyses were also nonsignificant for HMGCR (one SNP, $p$ = 0.955), NPC1L1 (one SNP, $p$ = 0.618), and PCSK9 (three SNPs, $p$ = 0.436); the strict PCSK9 instruments remained heterogeneous (Q = 7.35, $p$ = 0.0254).

For molecular-QTL triangulation, seven eligible eQTLGen HMGCR cis-eQTL instruments produced no clear association between higher HMGCR expression and CL (β = –0.000275, 95% CI: –0.003774 to 0.003225; $p$ = 0.878). The LDL-C positive control was directionally coherent and statistically significant (β = 0.219 SD LDL-C per

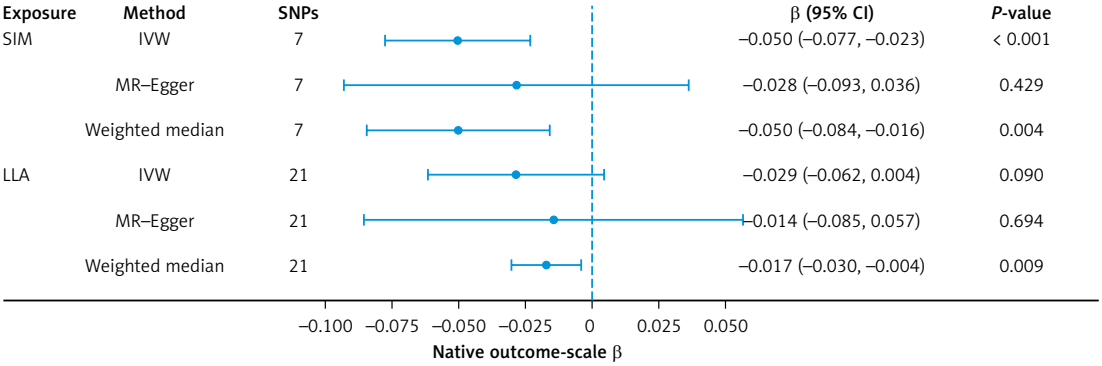



**Figure 3.** MR analysis results. The figure presents univariable Mendelian randomisation estimates for the associations of genetic liability to SIM-related medication use and LLA use with CL on the native linear-model outcome scale

SD higher HMGCR expression; $p = 0.00108$), supporting engagement of the HMGCR–LDL-C pathway. NPC1L1 had no cis-eQTL at $p < 5 \times 10^{-8}$, no eligible strict-European PCSK9 eQTL dataset was identified, the HMGCR INTERVAL cis-pQTL minimum $p$-value was 0.00110, and no eligible public NPC1L1 or provenance-complete PCSK9 plasma pQTL instrument was available. Accordingly, these combinations were not estimated.

Colocalisation did not support a shared causal variant between target-related LDL-C signals and CL: PP.H4 was 0.000176 for HMGCR, 0.000257 for NPC1L1, and 0.000656 for PCSK9 under the primary prior. HMGCR eQTL–CL, NPC1L1 eQTL–CL, and HMGCR pQTL–CL comparisons were likewise weak (PP.H4 ≤ 0.000311). These conclusions were unchanged across the prespecified p12 prior range. Complete primary, sensitivity, availability, and colocalisation results are reported in Supplementary Table SIX.

### Multivariable Mendelian randomisation

Joint clumping yielded 424 candidate variants for the SIM model, of which 405 remained after harmonisation, and 426 candidates for the LLA model, of which 406 remained. After joint adjustment for LDL-C, HDL-C, and triglycerides, the SIM medication-use direct effect was near null (β = 0.0051, SE = 0.0284; 95% CI: –0.0506 to 0.0608; $p = 0.858$). The LLA direct effect was also near null (β = 0.0060, SE = 0.0124; 95% CI: –0.0184 to 0.0303; $p = 0.631$). These are native linear-model outcome-scale coefficients, not formal odds ratios. Thus, the inverse SIM association observed in univariable MR did not persist in the lipid-adjusted model.

Under the zero-covariance scenario, conditional F-statistics for SIM, LDL-C, HDL-C, and triglycerides were 2.59, 3.62, 31.90, and 14.27, respectively; the corresponding values in the LLA model were 3.39, 4.81, 42.03, and 25.45. Across covariance scenarios (ρ = –0.20 to 0.30), the medication-trait conditional F-statistic remained < 10 (SIM, 2.34–3.07; LLA, 3.02–4.16). Residual heterogeneity was marked at ρ = 0 (SIM: Q = 1280.77, $df = 400$, $p = 8.53 \times 10^{-93}$; LLA: Q = 1277.09, $df = 401$, $p = 5.44 \times 10^{-92}$) and remained so across all covariance scenarios. Collinearity was not severe (maximum variance inflation factor 3.72 for SIM and 3.59 for LLA; condition numbers 3.76 and 3.65). MVMR-Egger intercepts were not significant ($p = 0.200$ and 0.500), and the medication effects remained nonsignificant after excluding variants involving proxies, excluding all palindromic variants, or excluding variants with Cook's distance > 4/n. Full results and harmonisation details are provided in Supplementary Table SVIII and the analysis audit files.

## Discussion

This study found modest inverse genetic correlations between CL and two phenotypically distinct medication-use traits: a specific simvastatin medication code and a class-level cholesterol-lowering medication response. Shared gene and pathway signals converged on cholesterol–bile acid and FXR-enterohepatic biology. Univariable MR supported a modest inverse association for SIM medication-use liability, whereas the LLA primary IVW estimate was not significant. In MVMR, neither medication-use direct effect remained significant after joint adjustment for LDL-C, HDL-C, and triglycerides. Complementary LDL-C-based cis-drug-target MR did not identify clear associations for HMGCR, NPC1L1, or PCSK9, and regional colocalisation was nonsupportive. These results concern distinct genetic estimands and do not directly estimate the effect of prescribing or initiating a drug.

LDSC and Genomic SEM characterise shared architecture rather than treatment effects or independent causal proof. Pairwise LDSC quantified inverse genetic correlations, while Genomic SEM provided a joint one-factor representation

in which the two medication-use traits loaded positively and CL loaded inversely. Its additional insight is therefore organisational rather than inferential: it shows that the covariance matrix can be summarised by one common dimension. The fixed SIM loading and favourable fit indices should not be read as evidence of biological dominance, especially given the three-indicator model and identification constraint.

Gene-level enrichment analyses further highlighted shared involvement of ABCG5/G8, CYP7A1, NPC1L1, HMGCR, and SLCO1B1, overlapping known gallstone susceptibility genes (ABCG5/8, CYP7A1) [31] and established statin pharmacogenes, particularly SLCO1B1 (e.g. rs4149056 influencing hepatic uptake and toxicity risk) [32] and HMGCR variants affecting LDL-C response and dose requirements [33]. Collectively, these genes form an integrated absorption–synthesis–secretion network: NPC1L1 regulates intestinal cholesterol uptake with compensatory HMGCR/SREBP2 feedback [34]; ABCG5/8 mediates cholesterol efflux into bile and the intestinal lumen [35]; CYP7A1 controls bile acid pool size and hepatic cholesterol disposal [36]; and SLCO1B1 modulates hepatic statin uptake and LDL-C lowering [37].

The FXR-enterohepatic axis provides a plausible biological bridge between the genetic and pathway findings. In the distal ileum, bile acids activate FXR, which induces fibroblast growth factor 15/19 (FGF15/19) signalling to the liver and suppresses CYP7A1-mediated bile acid synthesis through enterohepatic feedback [38, 39]. The apical sodium-dependent bile acid transporter (ASBT) maintains bile acid reabsorption in the terminal ileum, and disruption of this system changes bile acid pool size, faecal sterol excretion, and biliary lithogenicity [40, 41]. Therefore, terminal ileum enrichment and FXR pathway signals are consistent with a mechanism in which genetic variation influences both lipid-lowering medication-use liability and gallstone susceptibility through bile acid recycling and cholesterol disposal.

The apparent SIM–LLA contrast requires cautious interpretation. The SIM and LLA IVW point estimates were both modestly inverse and their confidence intervals overlapped; a significant result for one exposure and a nonsignificant result for another does not by itself demonstrate a true difference. Several features may nevertheless explain the different precision and robustness. SIM is a specific field 20003 medication code, whereas LLA is a broader, class-level PHESANT-derived medication-use phenotype whose exact drug composition is unavailable. The LLA instruments showed extreme heterogeneity ($Q$ = 239.823, $p < 0.001$), unlike SIM ($Q$ = 2.8085, $p$ = 0.832), consistent with heterogeneous prescribing indications, medication classes, and genetic pathways. The new cis-drug-target analyses did not replicate a clear protective association for HMGCR in the current CL dataset and were also null for NPC1L1 and PCSK9. This differs from the stronger inverse HMGCR estimate reported in a prior multi-biobank target-MR study [11], but the wide confidence intervals, different outcome construction, instrument selection, LD modelling, and effect scales preclude interpreting the difference as a contradiction or as evidence that HMGCR inhibition has no effect.

The attenuation of the SIM estimate in MVMR is compatible with lipid-related indication or shared lipid biology contributing to its univariable association. It is not, however, proof that LDL-C, HDL-C, or triglycerides mediate or fully explain that association. Summary-data MVMR estimates conditional direct effects only under its instrumental-variable assumptions. Here, conditional F-statistics were below 10 for both medication traits and LDL-C, and the multivariable Q statistics indicated marked residual heterogeneity. Accordingly, the near-null medication estimates are exploratory and cannot definitively exclude a direct effect.

The HMGCR molecular analysis adds a mechanistic qualification. Its eQTL instruments strongly predicted LDL-C in the positive-control analysis, indicating that the selected variants engaged the expected cholesterol pathway, yet the same instruments did not show a clear association with CL. This pattern weakens a simple interpretation in which any genetically proxied reduction in HMGCR activity necessarily reduces gallstone risk, but it remains inconclusive because whole-blood expression may not proxy hepatic target activity, the outcome regional signal was weak, and no qualifying HMGCR pQTL instrument was available. Likewise, the absence of eligible QTL instruments for several target–molecule combinations is a data limitation rather than negative biological evidence.

Clinically, these results should be viewed as hypothesis generating. Convergence on ABCG5/8, SLCO1B1, HMGCR, NPC1L1, CYP7A1, and FXR-related pathways suggests candidate mechanisms for future pharmacogenomic and bile acid studies, but it does not justify changing statin prescribing for gallstone prevention. Similarly, FXR-targeted therapeutics remain mechanistically relevant, yet no FXR agonist is currently approved specifically for gallstone prevention; translational claims therefore require functional validation and outcome-focused clinical studies.

Several limitations warrant consideration. First, the exposure GWASs capture liability to self-reported medication use. Their instruments may reflect the indication for treatment (including genetically influenced lipid levels and cardiovascu-

lar risk), health-care access, medication selection, reporting behaviour, adherence, dose, duration, comorbidity, or pharmacodynamic pathways. Such pathways can violate the MR independence or exclusion-restriction assumptions, so the estimates cannot be interpreted as the clinical effect of starting simvastatin. Second, the exact agent composition of the LLA phenotype is unavailable, and the public summary metadata do not provide drug-specific or sex-stratified effects, which limits detailed phenotype reconstruction and may contribute to heterogeneity. Third, CL was obtained from Neale Lab round 1 while the medication-use phenotypes were obtained from IEU round 2, introducing differences in phenotype construction and quality control. Fourth, European ancestry limits transportability because gallstone prevalence, medication-use patterns, and statin pharmacokinetics, including SLCO1B1 allele frequencies, vary across populations. Fifth, the limited SIM instrument count reduces power for pleiotropy detection and small-effect sensitivity analyses. Finally, pathway enrichments are computational and require experimental validation, ideally in hepatocyte, cholangiocyte, enterocyte, or organoid systems under lipid and bile acid perturbations.

The MVMR analysis has additional limitations. The medication and lipid GWASs include overlapping UK Biobank participants, but per-SNP covariances among exposure-effect estimates were unavailable. The medication, lipid, and CL GWASs were also derived wholly or partly from UK Biobank, so exposure–outcome sample overlap was likely, although exact participant-level overlap was unavailable. In the presence of conditionally weak instruments, such overlap may increase susceptibility to weak-instrument bias. We therefore reported zero-covariance diagnostics together with equicorrelation sensitivity scenarios; the conditional *F*-statistics and *Q* tests, unlike the IVW point estimates, depend on this assumption. Medication-trait and LDL-C conditional *F*-statistics remained < 10 across the examined covariance scenarios, and heterogeneity was substantial. These findings limit conditional identification and preclude definitive causal partitioning of medication-use and lipid-trait effects.

The cis-drug-target analysis also has specific limitations. The primary $r^2 < 0.10$ models used correlated instruments and therefore depended on the accuracy of the European LD matrices; strict clumping reduced the instrument counts to one, one, and three for HMGCR, NPC1L1, and PCSK9, respectively. The Neale Lab CL GWAS used a linear model for a binary outcome, so native β values – not exponentiated β values – are primary, and the reported logit/OR conversions are only first-order sensitivity approximations. NPC1L1 and strict PCSK9 analyses showed heterogeneity, several eQTL/pQTL combinations lacked qualifying European instruments, and relevant-tissue QTL sample sizes were limited. Finally, the single-causal-variant colocalisation model and weak regional CL signals limit the ability to rule in or rule out shared causal variants.

In conclusion, genetic liability to reporting/using simvastatin was modestly and inversely associated with CL in univariable MR, whereas the class-level LLA phenotype showed substantial instrument heterogeneity and no significant primary IVW association. After joint adjustment for LDL-C, HDL-C, and triglycerides, neither medication-use direct effect remained significant. However, weak conditional instruments for the medication traits and LDL-C, together with marked residual heterogeneity, mean that this attenuation cannot establish complete lipid mediation or definitively exclude a direct effect. In complementary cis-drug-target MR, HMGCR-, NPC1L1-, and PCSK9-mediated LDL-C lowering showed no clear association with CL, and colocalisation was nonsupportive; these null estimates remain inconclusive because of outcome-scale, instrument, and molecular-QTL limitations. The shared architecture converged on cholesterol–bile acid and enterohepatic pathways, but the medication-use GWASs do not estimate drug efficacy or treatment initiation; these findings remain hypothesis generating.

## Acknowledgments

The authors gratefully acknowledge the technical support and professional guidance provided by Knowledge-Vision Co., Ltd. during this research.

## Data availability statement

Publicly available GWAS summary statistics can be accessed using the identifiers and source links listed in the Supplementary Material. Additional harmonised instruments, analysis-ready summary data, analysis code, request manifests, audit trails, and validation outputs generated in this study are available from the corresponding author upon reasonable request.

## Funding

No external funding.

## Institutional review board statement

This study utilised exclusively summary-level genetic information obtained from the publicly available and anonymised IEU Open GWAS database (https://gwas.mrcieu.ac.uk/). Under Articles 32(1) and 32(2) of China's Ethical Review Measures for Life Sciences and Medical Research Involving Humans (National Health Commission *et al*., 2023), no further ethical approval is required

for secondary analysis of legally acquired and fully de-identified public datasets. Each original genome-wide association study that contributed data to this resource had received ethical approval from an applicable Institutional Review Board and provided explicit documented consent allowing the sharing and reuse of anonymised data for scientific purposes.

## Conflict of interest

The authors declare no conflict of interest.

## References


1. Wang X, Yu W, Jiang G, et al. Global epidemiology of gallstones in the 21st century: a systematic review and meta-analysis. Clin Gastroenterol Hepatol 2024; 22: 1586-95.
2. Tazuma S. Gallstone disease: epidemiology, pathogenesis, and classification of biliary stones (common bile duct and intrahepatic). Best Pract Res Clin Gastroenterol 2006; 20: 1075-83.
3. Pak M, Lindseth G. Risk factors for cholelithiasis. Gastroenterol Nurs 2016; 39: 297-309.
4. Di Ciaula A, Wang DQ, Portincasa P. An update on the pathogenesis of cholesterol gallstone disease. Curr Opin Gastroenterol 2018; 34: 71-80.
5. Istvan ES, Deisenhofer J. Structural mechanism for statin inhibition of HMG-CoA reductase. Science 2001; 292: 1160-4.
6. Han T, Lv Y, Wang S, Hu T, Hong H, Fu Z. Pioglitazone prevents cholesterol gallstone formation through the regulation of cholesterol homeostasis in guinea pigs with a lithogenic diet. Lipids Health Dis 2019; 18: 218.
7. Smith JL, Roach PD, Wittenberg LN, et al. Effects of simvastatin on hepatic cholesterol metabolism, bile lithogenicity and bile acid hydrophobicity in patients with gallstones. J Gastroenterol Hepatol 2000; 15: 871-9.
8. Ginting Z. The effect of statin in managing cholelithiasis: a systematic review. Int J Res Rev 2024; 11: 607-12.
9. Martin D, Schmidt R, Mortensen EM, Mansi I. Association of statin therapy and risks of cholelithiasis, biliary tract diseases, and gallbladder procedures: retrospective cohort analysis of a US population. Ann Pharmacother 2016; 50: 161-71.
10. Kan HP, Guo WB, Tan YF, Zhou J, Liu CD, Huang YQ. Statin use and risk of gallstone disease: a meta-analysis. Hepatol Res 2015; 45: 942-8.
11. Yang G, Mason AM, Gill D, Schooling CM, Burgess S. Multi-biobank Mendelian randomization analyses identify opposing pathways in plasma low-density lipoprotein-cholesterol lowering and gallstone disease. Eur J Epidemiol 2024; 39: 857-67.
12. Post SM, Duez H, Gervois PP, Staels B, Kuipers F, Princen HM. Fibrates suppress bile acid synthesis via peroxisome proliferator-activated receptor-alpha-mediated downregulation of cholesterol 7alpha-hydroxylase and sterol 27-hydroxylase expression. Arterioscler Thromb Vasc Biol 2001; 21: 1840-5.
13. Reihnér E, Rudling M, Ståhlberg D, et al. Influence of pravastatin, a specific inhibitor of HMG-CoA reductase, on hepatic metabolism of cholesterol. N Engl J Med 1990; 323: 224-8.
14. Tomlinson B, Chan P, Lan W. How well tolerated are lipid-lowering drugs? Drugs Aging 2001; 18: 665-83.
15. Chen X, Liu X. Weekend warrior physical activity patterns can reduce the risk of gallstone formation: NHANES 2017-2020 and Mendelian randomization analyses. Arch Med Sci 2026; 22: 45-55.
16. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet 2015; 47: 1236-41.
17. Zhang Y, Cheng Y, Jiang W, Ye Y, Lu Q, Zhao H. Comparison of methods for estimating genetic correlation between complex traits using GWAS summary statistics. Brief Bioinform 2021; 22: bbaa442.
18. Grotzinger AD, Rhemtulla M, de Vlaming R, et al. Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. Nat Hum Behav 2019; 3: 513-25.
19. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol 2015; 11: e1004219.
20. Gerring ZF, Mina-Vargas A, Gamazon ER, Derks EM. E-MAGMA: an eQTL-informed method to identify risk genes using genome-wide association study summary statistics. Bioinformatics 2021; 37: 2245-9.
21. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol 2013; 37: 658-65.
22. 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: 304-14.
23. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol 2015; 44: 512-25.
24. Richardson TG, Sanderson E, Palmer TM, et al. Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: a multivariable Mendelian randomisation analysis. PLoS Med 2020; 17: e1003062.
25. Võsa U, Claringbould A, Westra HJ, 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: 1300-10.
26. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 2020; 369: 1318-30.
27. Sun BB, Maranville JC, Peters JE, et al. Genomic atlas of the human plasma proteome. Nature 2018; 558: 73-9.
28. Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet 2014; 10: e1004383.
29. Sanderson E, Davey Smith G, Windmeijer F, Bowden J. An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. Int J Epidemiol 2019; 48: 713-27.
30. Sanderson E, Spiller W, Bowden J. Testing and correcting for weak and pleiotropic instruments in two-sample multivariable Mendelian randomization. Stat Med 2021; 40: 5434-52.
31. Costa CJ, Nguyen MTT, Vaziri H, Wu GY. Genetics of gallstone disease and their clinical significance: a narrative review. J Clin Transl Hepatol 2024; 12: 316-26.
32. Cooper-DeHoff RM, Niemi M, Ramsey LB, et al. The clinical pharmacogenetics implementation consortium guideline for SLCO1B1, ABCG2, and CYP2C9 genotypes and statin-associated musculoskeletal symptoms. Clin Pharmacol Ther 2022; 111: 1007-21.

33. Khaleqsefat E, Rasul KH, Kheder RK, Baban S, Baban J. Frameshift variation in the HMG-CoA reductase gene and unresponsiveness to cholesterol-lowering drugs in type 2 diabetes mellitus patients. Sci Rep 2025; 15: 288.
34. Sané AT, Sinnett D, Delvin E, et al. Localization and role of NPC1L1 in cholesterol absorption in human intestine. J Lipid Res 2006; 47: 2112-20.
35. Li T, Matozel M, Boehme S, et al. Overexpression of cholesterol 7αhydroxylase promotes hepatic bile acid synthesis and secretion and maintains cholesterol homeostasis. Hepatology 2011; 53: 996-1006.
36. Li X, Xiao H, Jian X, et al. Epigenetic regulation of key enzymes CYP7a1 and HMGCR affect hepatic cholesterol metabolism in different breeds of piglets. Front Vet Sci 2020; 7: 231.
37. Dagli-Hernandez C, Zhou Y, Lauschke VM, et al. Pharmacogenomics of statins: lipid response and other outcomes in Brazilian cohorts. Pharmacol Rep 2022; 74: 47-66.
38. Chiang JYL, Ferrell JM. Discovery of farnesoid X receptor and its role in bile acid metabolism. Mol Cell Endocrinol 2022; 548: 111618.
39. Jiang L, Zhang H, Xiao D, Wei H, Chen Y. Farnesoid X receptor (FXR): Structures and ligands. Comput Struct Biotechnol J 2021; 19: 2148-59.
40. Ticho AL, Malhotra P, Dudeja PK, Gill RK, Alrefai WA. Intestinal absorption of bile acids in health and disease. Compr Physiol 2019; 10: 21-56.
41. Dawson PA, Haywood J, Craddock AL, et al. Targeted deletion of the ileal bile acid transporter eliminates enterohepatic cycling of bile acids in mice. J Biol Chem 2003; 278: 33920-7.