

Cross-ethnic Molecular Signatures Underpin the Adverse Impact of Statin Use on Type 2 Diabetes

Fengzhe Xu ^{1,#}, Min Yang ^{1,#}, Wei Hu ^{2,#}, Shuai Yuan ³, Xue Cai ^{1,4,5},
Wanglong Gou ^{1,4,5}, Zelei Miao ^{1,4,5}, Bang-yan Li ², Liang Yue ^{1,4,5}, Zhangzhi Xue ^{1,4,5},
Menglei Shuai ^{1,4}, Luqi Shen ^{1,4,5}, Yuanqing Fu ^{1,4,5}, Tiannan Guo ^{1,4,5,*},
Yu-ming Chen ^{2,*}, Ju-Sheng Zheng ^{1,4,5,*}

¹Westlake Center for Intelligent Proteomics, Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou 310024, China

²Guangdong Provincial Key Laboratory of Food, Nutrition and Health, Department of Epidemiology, School of Public Health, Sun Yat-sen University, Guangzhou 510275, China

³Unit of Cardiovascular and Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm 17177, Sweden

⁴School of Medicine and School of Life Sciences, Westlake University, Hangzhou 310024, China

⁵Institute of Basic Medical Sciences, Westlake Institute for Advanced Study, Hangzhou 310024, China

*Corresponding authors: zhengjusheng@westlake.edu.cn (Zheng JS), chenyum@mail.sysu.edu.cn (Chen YM), guotianann@westlake.edu.cn (Guo T).

#Equal contribution.

Handling Editor: Minxian Wang

Abstract

The use of statins as the primary therapy for reducing low-density lipoprotein cholesterol has raised concerns regarding their potential side effects in increasing the risk of type 2 diabetes (T2D). However, the underlying mechanism remains largely unknown. In this study, we utilized multi-omics molecular signatures to unravel the etiology of statin-induced T2D. Through systematic screening of 102 gut microbial features, 40 blood metabolites, and 131 circulating proteins in East Asians and Europeans, we identified a set of blood metabolites and proteins potentially influenced by genetically proxied statin use. Notably, Mendelian randomization analyses provided evidence that elevated circulating levels of gastric inhibitory polypeptide (GIP) were associated with an increased risk of T2D. This association between genetically proxied statin use and GIP was consistently observed across East Asian and European populations, highlighting the pivotal role of GIP in modulating the risks of statin-induced T2D. Furthermore, this study established an extensive atlas of multi-omics molecular signatures associated with statin-induced T2D, offering valuable insights for prioritizing intervention targets.

Key words: Statin; T2D; Multi-omics; Cross-ethnic; Mendelian randomization.

Graphical abstract

Key question

Why statin therapy could raise the risk of type 2 diabetes (T2D)? Does the underlying mechanism apply to populations with different ancestries?

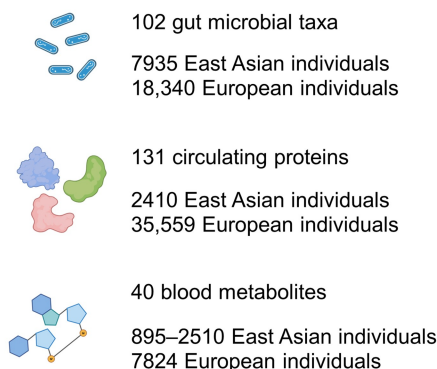
Key finding

In East Asians and Europeans, genetically proxied statin use was associated with multi-omics biomarkers. The model integrating statin-altered signatures outperformed conventional T2D risk factors for the prediction of T2D.

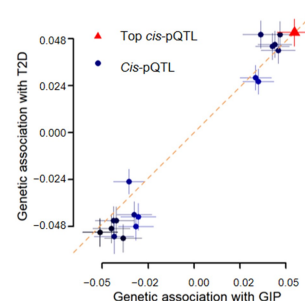
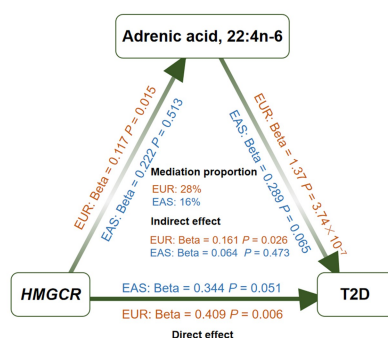
Take-home message

These findings suggest that the identified signatures may be used as potential intervention targets for T2D prevention among statin users.

Multi-omics biological signatures



Important mediation factors



Introduction

Statin therapy is the most potent treatment for lowering low-density lipoprotein cholesterol (LDL-C) and is widely used for primary prevention of cardiovascular diseases [1,2]. However, concerns persist about its side effects that increase the risk of type 2 diabetes (T2D). Large-scale population-based observational studies have unveiled an association between statin therapy and dose-dependent reductions in insulin sensitivity and secretion [3], accompanied by high blood glucose levels [4]. A Mendelian randomization (MR) study also supports a positive association between statin use and risk of T2D [5], which has been further confirmed in a meta-analysis of randomized controlled trials (RCTs) [6]. Despite these findings, the underlying mechanisms remain poorly understood.

Statins may have diverse influences on the biological system, as reflected by the changes in different omics signatures. For example, statins could alter the gut microbiome composition, fecal bile acids, and blood metabolites in animal models [7,8]. A small-population-based observational study suggests that statin therapy is associated with the modification of the gut microbiome composition [9]. Statin therapy is also linked to changes in blood metabolites, such as lower levels of docosahexaenoic acid [10]. To date, studies regarding the molecular signatures of statin use are limited to individual omics in a single observational cohort with small sample sizes, underscoring the need for a holistic blueprint of multi-omics constellations.

Integration of human genetic data with multi-omics signatures enables a comprehensive investigation into the effects of genetically predicted drug targets on biological systems and disease pathways [11,12]. These studies mimic the desired

drug response using genetic instruments within or near the genes encoding the drug target proteins [13]. The genetic variants in the hydroxymethyl glutaryl coenzyme A reductase (*HMGCR*) gene, which modulate a rate-limiting step in the biosynthesis of cholesterol, have been utilized as genetic instruments to mimic the effects of *HMGCR* inhibitors (statins) on complex diseases [5,11,14].

Here, we combined multi-omics features, including 102 gut microbial taxa, 40 blood metabolites, and 131 blood proteins, collected from East Asians ($n = 895$ to 7935) and Europeans ($n = 7824$ to 35,559), to identify shared and generalizable etiology of statin-induced (proxied by *HMGCR* variants) T2D across these two ethnic populations (Figure 1). Moreover, in a deeply phenotyped human cohort study, we used the machine learning model to validate that combining statin-altered molecular signatures outperformed traditional T2D risk factors for T2D prediction, providing potential targets for the prevention of statin-induced T2D.

Results

Genetically proxied statin use is positively associated with T2D risk

Previous studies have consistently affirmed a positive correlation between statin use and the risk of T2D in European populations [5,6]. The proprotein convertase subtilisin/kexin type 9 (PCSK9) inhibitors are another class of widely used lipid-lowering agents. Therapies targeting PCSK9, block the interaction between PCSK9 and low-density lipoprotein receptor (LDLR), thereby attenuating LDLR degradation. We further compared the associations of PCSK9 inhibitors and statin therapy with T2D across populations with diverse

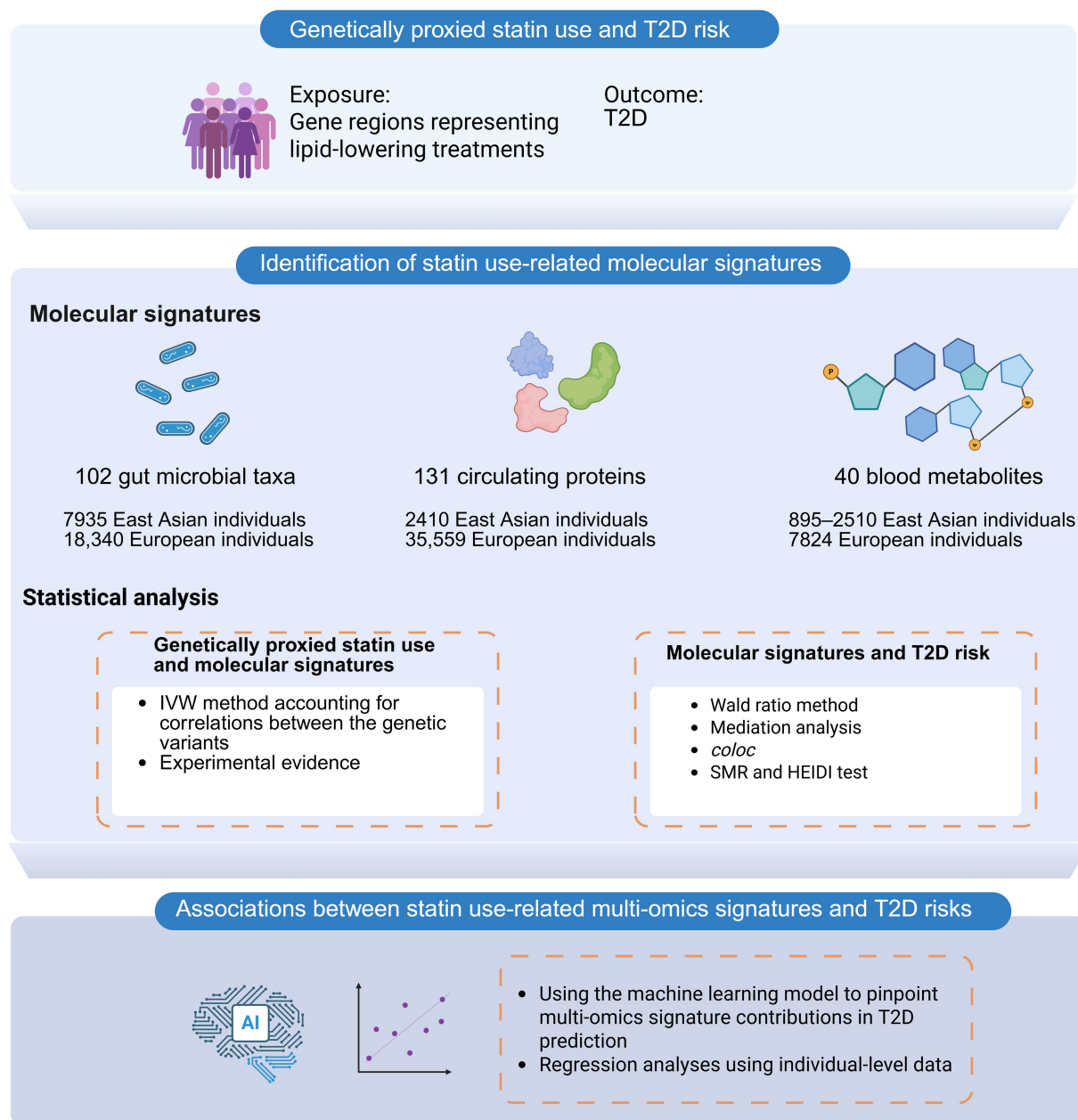


Figure 1 Study design for investigating cross-ethnic multi-omics signatures of statin therapy

The genetic variant–LDL associations were drawn from the GLGC. The microbiome GWAS summary statistics from meta-analyses on Chinese populations and the MicroGen consortium were used to determine statin–microbiota associations. GWAS summary-level data for blood metabolites were obtained from the GNHS, TwinsUK, and KORA datasets. The genetic variant–protein associations were acquired from the GNHS, the Icelandic Cancer Project, and deCODE genetics. In the subsequent analyses, the associations between genetically proxied statin use and protein targets in the Icelandic Cancer Project and deCODE genetics were examined. Based on the individual-level GNHS data, a machine learning model was developed to prioritize targets based on feature importance in the prediction model. LDL, low-density lipoprotein; T2D, type 2 diabetes; GWAS, genome-wide association study; GLGC, Global Lipids Genetics Consortium; GNHS, Guangzhou Nutrition and Health Study; IVW, inverse-variance weighted; SMR, summary-data-based Mendelian randomization; HEIDI, heterogeneity in dependent instruments.

ancestries employing genetic instruments [15–17]. We performed clumping based on genome-wide association study (GWAS) summary statistics of LDL-C, using a threshold of $P < 1 \times 10^{-8}$ and linkage disequilibrium (LD) < 0.3 for loci within ± 1000 kb of the transcription start sites of the *HMGCR*, *LDLR*, and *PCSK9* genes. The inverse-variance weighted (IVW) method, accounting for SNP correlation, was used to investigate the effect of genetically predicted lipid-lowering treatments. The effect size estimates for lipid-related genetic variants were scaled to a one standard

deviation (SD) decrease in LDL-C. Our findings confirmed that statin use, represented by *HMGCR* variants, exhibited a positive association with T2D risk in both East Asian and European ethnic populations. The odds ratio (OR) was 1.43 [95% confidence interval (CI) = 1.25–1.62] for East Asians and 1.66 (95% CI = 1.31–2.11) for Europeans (Figure 2; Table S3).

Additionally, we extended our investigation to two other drug targets, *LDLR* and *PCSK9*, to compare the potential impacts of alternative lipid-lowering treatments on T2D.

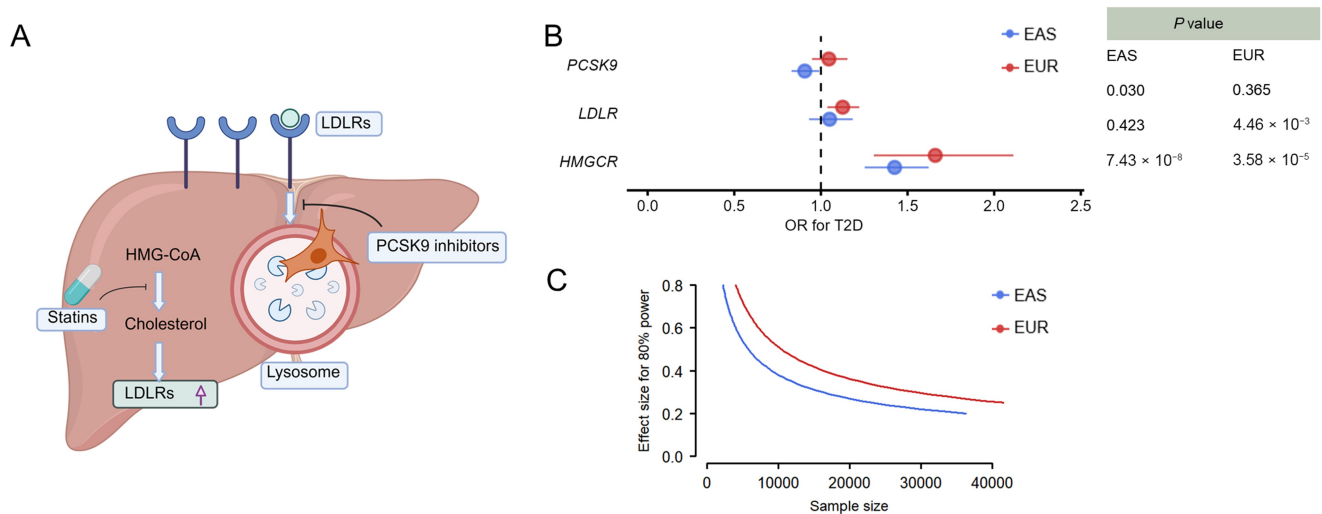


Figure 2 Association of lipid-lowering treatment targets with T2D and the power for MR analyses

A. Pharmacology of statins and PCSK9 inhibitors. Statins inhibit HMG-CoA to prevent cholesterol production in the liver, while PCSK9 inhibitors prevent LDLRs from being degraded by PCSK9-mediated processes. **B.** Associations of LDL-related genes with T2D. The LDL-related genes, including *HMGCR*, *LDLR*, and *PCSK9*, in two populations were analyzed. **C.** Power calculations for MR analyses of the *HMGCR* gene. The plot shows the detectable estimates at 80% power. MR, Mendelian randomization; EUR, European; EAS, East Asian; HMG-CoA, hydroxymethyl glutaryl coenzyme A; HMGCR, hydroxymethyl glutaryl coenzyme A reductase; PCSK9, proprotein convertase subtilisin/kexin type 9; LDLR, low-density lipoprotein receptor; OR, odds ratio.

Genetic variants within the genes encoding these lipid-lowering drug targets served as instrumental variables. Results for *PCSK9* revealed inconsistency between the two populations (OR = 0.91, 95% CI = 0.83–0.99 for East Asians; OR = 1.05, 95% CI = 0.95–1.15 for Europeans). Regarding *LDLR*, the effect estimates exhibited the same direction across both populations (OR = 1.05, 95% CI = 0.93–1.18 for East Asians; OR = 1.13, 95% CI = 1.04–1.22 for Europeans), although they were notably weaker than those observed for *HMGCR*.

The consistent effect of genetically proxied statin use on increased T2D risk observed in both populations raised our interest in the shared underlying mechanisms. We systematically screened multiple microbial taxa, metabolites, and proteins that overlap between the two populations. Concurrently, the MR approach was employed to mitigate biases arising from confounding factors and bolster causal inference, mimicking an RCT.

Gut microbial signatures of statin therapy across East Asian and European populations

We examined the associations between the *HMGCR* gene region and the gut microbiome among 7935 Chinese and 18,340 Europeans [18] by analyzing 102 overlapping taxa across the two ethnic datasets (6 phyla, 9 classes, 11 orders, 22 families, 54 genera) from 16S rRNA sequencing data (Figure 1). After multiple testing correction based on microbial taxonomic levels, we did not observe significant results [false discovery rate (FDR) < 0.05]. Genetically proxied statin use showed a modest negative correlation with gut microbiota in both cohorts (Pearson correlation = -0.11). Notably, genetically proxied statin use was positively associated with the relative abundance of the genus *Faecalibacterium* in both populations (Beta = 0.36, 95% CI = 0.03–0.71 for East Asians; Beta = 0.30, 95% CI = 0.002–0.60 for Europeans). Meta-analysis suggested that genetically

proxied statin use was negatively associated with the order Actinomycetales, the family Actinomycetaceae, and the genera *Actinomyces*, *Ruminococcus gnavus* group, *Eisenbergiella*, and *Erysipelatoclostridium* at a nominally significant level ($P < 0.05$) (Figure 3A; Table S4). The association with the order Actinomycetales remained significant after FDR correction (FDR < 0.05).

Genetically proxied statin use has a profound impact on fatty acid metabolism

We subsequently sought to identify blood metabolites that statin therapy may influence among up to 2510 Chinese and 7824 Europeans [19]. We included 40 blood metabolites in the analysis, featuring 10 bile-acid-related metabolites and 10 long-chain fatty acids. Genetically proxied statin use positively correlated with blood metabolites in both populations (Pearson correlation = 0.27), while no significant results were observed after multiple testing correction (FDR < 0.05). At nominal significance ($P < 0.05$), the MR results showed that genetically proxied statin use was positively associated with adrenic acid (22:4n-6), oleic acid (18:1n-9), linoleic acid (18:2n-6), and palmitic acid in both populations (Figure 3A; Table S5), whereas it was negatively associated with creatine, glycocholic acid, and threonic acid across the two ethnic groups.

We further found that, in the univariable MR analysis, adrenic acid was positively associated with higher T2D risk across ethnic populations (OR = 1.29, 95% CI = 1.07–1.55 for East Asians; OR = 2.28, 95% CI = 1.51–3.44 for Europeans; meta-analysis: OR = 1.42, 95% CI = 1.19–1.68) (Table S6). The multivariable MR analysis adjusting for the genetic variants in the *HMGCR* gene region showed a concordant result with the univariable MR (Figure 3C). Mediation analysis revealed that adrenic acid was a putative mediator through which statin therapy increased T2D risk (see Materials and methods), and the proportion of

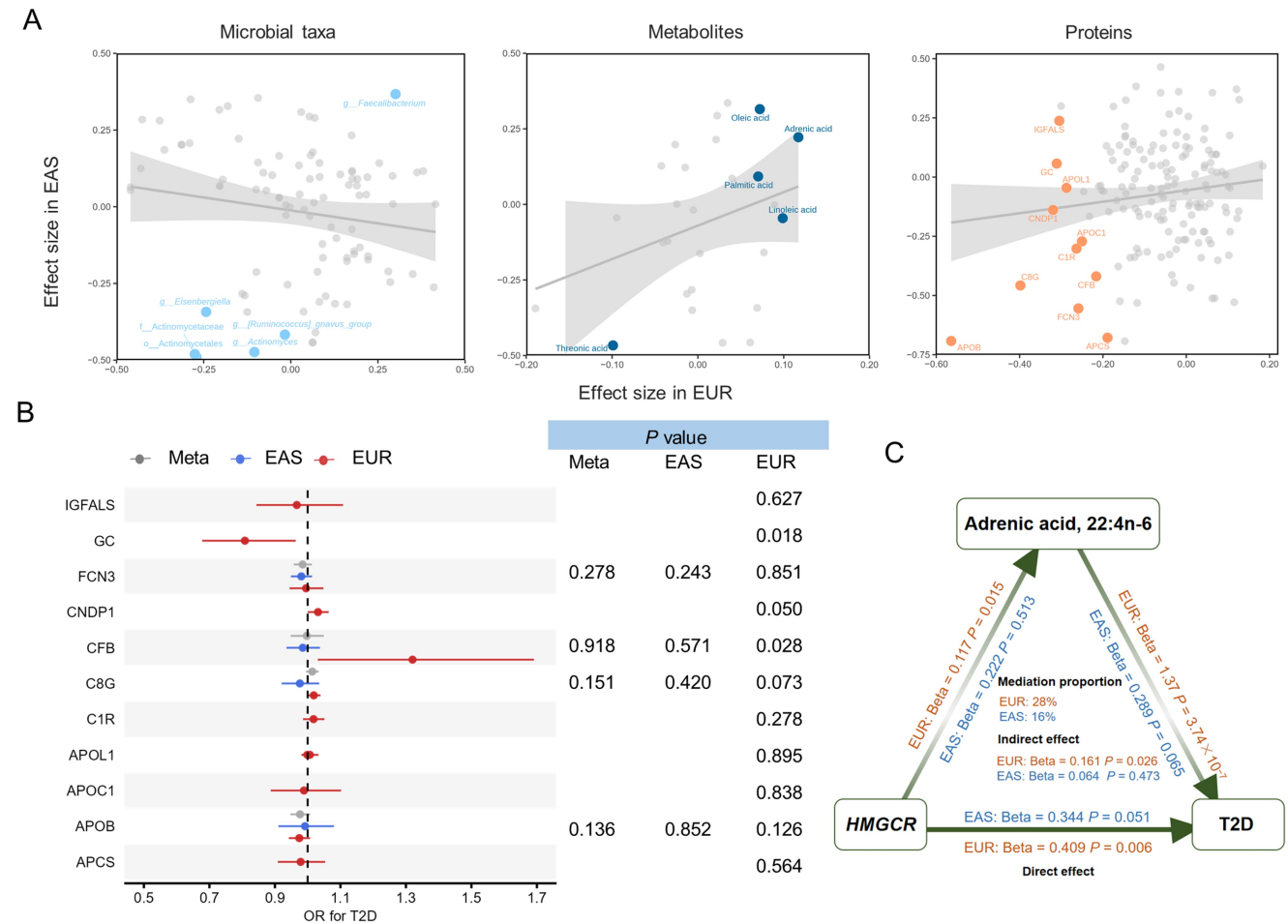


Figure 3 Association of genetically proxied statin use with gut microbiota, blood metabolites, and proteins
A. Potential causal effects of genetically proxied statin use on gut microbial taxa, blood metabolites, and proteins. Results from EAS and EUR populations are shown, with a nominal significance highlighted in the fixed-effect meta-analysis. **B.** Univariable MR analyses for proteins with available genetic instruments in the corresponding populations. **C.** Further multivariable MR analysis of adrenic acid.

mediation effect was 16% in East Asians and 28% in Europeans (Figure 3C).

Blood proteomic signatures underlie the positive link between statin use and T2D risk

To explore proteins modulated by statin use, blood proteomics GWAS summary statistics from our public dataset of Chinese participants ($n = 2410$) [20] and published summary statistics of Europeans ($n = 35,559$) were utilized, examining 131 overlapping proteins. The results revealed a negative correlation between genetically proxied statin use and apolipoprotein B (APOB) levels, aligning with statin in lowering LDL-C (Figure 3A; Table S7). Meta-analysis of East Asian and European data showed negative associations between genetically proxied statin use and several proteins involved in the complement system, such as complement component C8 gamma chain (C8G), serum amyloid P-component (APCS), complement factor B (CFB), and complement C1r subcomponent (C1R), which are important components in the complement activation, with C8G remained significant after multiple testing correction (FDR = 0.038). MR analysis in the European population showed that genetically proxied statin use was negatively associated with

vitamin D-binding protein (VDBP/GC) levels (Beta = -0.31 , 95%CI = $-0.56 - -0.07$). GC was involved in the transport and storage of vitamin D, and it was negatively associated with T2D risk in subsequent univariable MR analysis (OR = 0.81, 95% CI = 0.68–0.96) (Figure 3B; Table S6).

To comprehensively understand statin-induced T2D, we assessed 4752 protein targets using the GWAS summary statistics from individuals of European ancestry. Among these, 151 proteins exhibited nominal associations with T2D risk ($P_{IVW} < 0.05$ or $P_{SMR} < 0.05$; SMR, summary-data-based Mendelian randomization) (Figure 4A; Table S8). However, the majority of proteins were regulated by a single locus, potentially resulting in a pleiotropic effect. To address this concern, we applied colocalization analysis to distinguish causality from pleiotropy [21,22]. The analysis included SNPs located within ± 1000 kb of each protein-coding gene's transcription start site. Among the 151 proteins, 12 showed associations with T2D, supported by evidence from colocalization analysis [posterior probability for H4 (PP₄) > 0.7]. To assess whether the two traits share a common genetic signal, we performed heterogeneity in dependent instruments (HEIDI) tests [21]. The null hypothesis of the HEIDI test posits that a single causal variant underlies both the protein

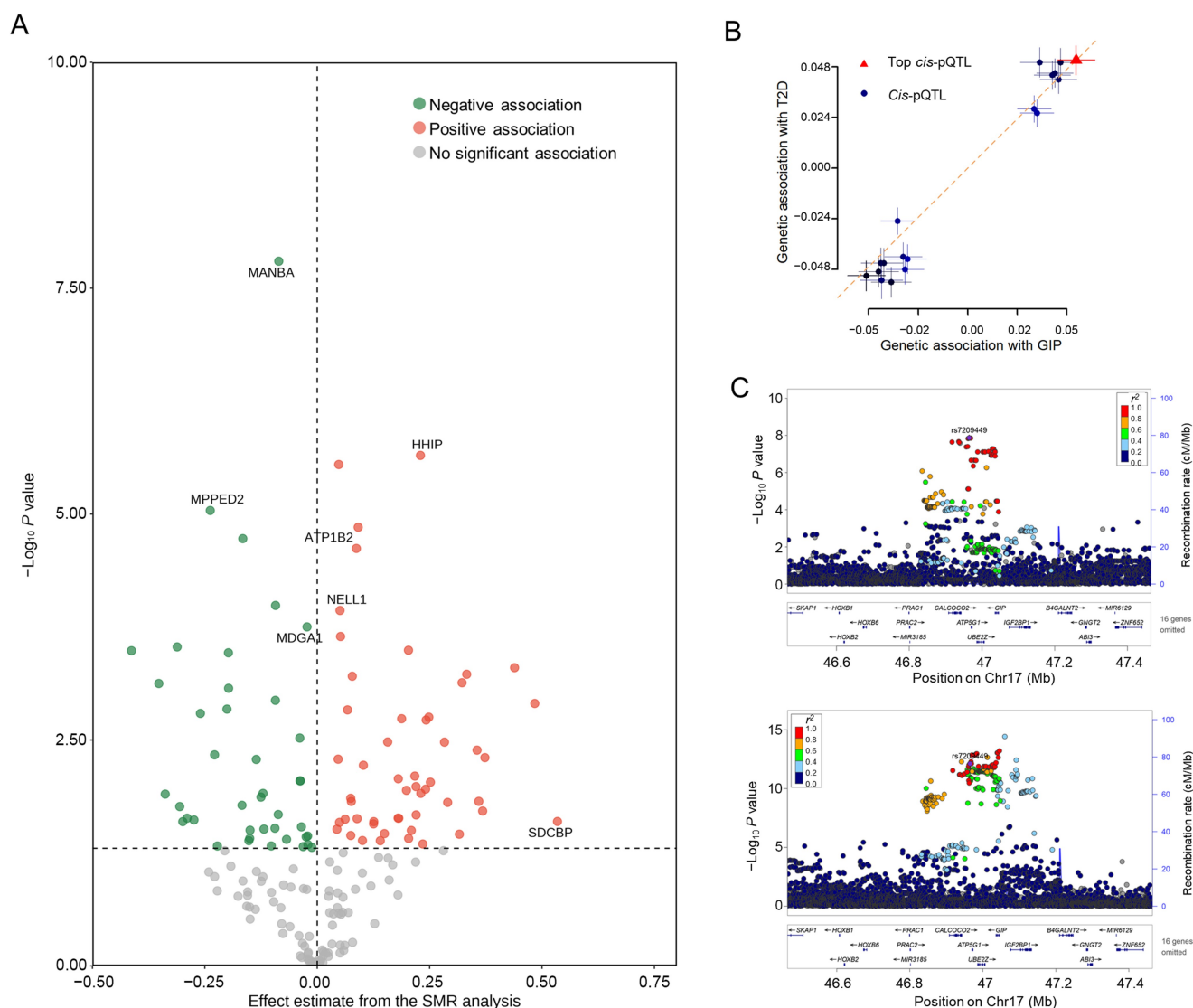


Figure 4 Link between genetically proxied statin use and proteomic pattern prioritizes potential therapeutic targets

A. Proteins nominally associated with statins. Red dots denote proteins positively associated with T2D risk; green dots signify proteins negatively associated with T2D risk. Proteins that passed the HEIDI test and received support from the *coloc* method were highlighted. **B.** Associations of GIP with T2D. **C.** Colocalization analysis of GIP with T2D. The upper panel shows the GWAS results for GIP protein abundance within the GIP gene region on Chr17, while the lower panel displays the T2D GWAS results. The posterior probability of a shared causal variant for GIP and T2D (PP_4) is 0.74, suggesting that GIP may mediate statin-induced T2D. Chr, chromosome; GIP, gastric inhibitory polypeptide; pQTL, protein quantitative trait locus.

expression levels and the clinical trait. Eight out of the twelve proteins successfully passed the test ($HEIDI P > 0.05$). Notably, gastric inhibitory polypeptide (GIP) emerged as a protein with converging evidence from both MR analyses and two colocalization methods (Table S9).

Genetic evidence from colocalization analyses suggested GIP as a putative risk factor for T2D ($OR = 2.75$, $SMR\text{-}multi P = 7.5 \times 10^{-6}$, $HEIDI P = 0.31$, $PP_4 = 0.74$) (Figure 4B and C), aligning with previous observational studies [23]. Furthermore, we sought to validate whether the associations between genetically proxied statin use and GIP levels were consistent across different populations, and confirmed positive associations using GWAS summary data of East Asians ($n = 262$) from the UK Biobank study ($Beta = 1.69$, $95\% CI = -0.27-3.66$, $P = 0.09$). Given the limited sample size, we further validated these findings within *in vitro* experiments. Leveraging published data from the Gene Expression Omnibus database (GEO: GSE8686) [24],

we observed that atorvastatin treatment in human microvascular endothelial cells induced a twofold increase in GIP expression compared to untreated cells (two-sided $P = 0.01$), suggesting that statin treatment enhances GIP expression levels. The addition of mevalonate reversed the atorvastatin-induced change (atorvastatin + mevalonate-treated *vs.* vehicle-treated, $P = 0.59$), consistent with statin effects mediated via the inhibition of the mevalonate pathway [24,25]. In summary, both genetic evidence across diverse ancestries and *in vitro* experiments strongly suggest that statin use may enhance GIP levels. Our analyses, in conjunction with previous studies, affirm that circulating GIP abundance is a putative risk factor in T2D progression.

Reverse MR analysis reveals the effect of T2D on molecular signatures

To explore how disease progression may influence molecular phenotypes and lead to reverse causal effects, we investigated

the potential causal effects of genetically determined T2D risk on molecular signatures by reverse MR analysis. The genetic instruments ($LD < 0.05$, $P < 5 \times 10^{-8}$) were selected from T2D GWAS studies, including the DIAMANTE meta-analysis (East Asians: 74,124 cases and 824,006 controls) and the AGEN-T2D study (Europeans: 77,418 cases and 356,122 controls) [16,17]. Molecular signature data were derived from the same datasets used in the statin-related analyses. We applied the generalized summary-data-based Mendelian randomization (GSMR) method and the HEIDI test to assess associations [26]. In East Asian and European populations, we identified 1 and 13 gut microbial features potentially affected by genetically determined T2D risk ($P < 0.05$) (Table S10), respectively; none of these features remained significant after FDR correction on microbial taxonomic levels. For metabolites, one in East Asian and seven in European populations were potentially influenced by genetically determined T2D risk ($P < 0.05$), among which six associations in Europeans remained significant after FDR correction. None of these six metabolites was associated with genetically proxied statin use in the MR analyses ($P > 0.05$) (Table S10). Moreover, genetically determined T2D risk was associated with 4 and 97 plasma proteins in East Asian and European populations, respectively. Among these, 88 associations in Europeans passed the FDR correction. Of the 88 proteins, 5 (APOL1, APOB, CNDP1, FCN3, and IGFALS) were also associated with genetically proxied statin use ($P < 0.05$) (Table S10).

Synergetic statin-induced multi-omics signatures enhance T2D prediction

To demonstrate the clinical relevance of the identified multi-omics signatures and prioritize their importance in predicting T2D, we employed a machine learning model (support vector machine), leveraging its robustness to overfitting and kernel-based capacity to model nonlinear relationships and complex correlations in high-dimensional data. Our analyses were conducted in the Guangzhou Nutrition and Health Study (GNHS), a community-based deep phenotyping cohort study in Guangzhou, China (Table S11). We included 130 T2D cases and 130 controls randomly selected from 719 controls. The model was trained using statin-altered features evaluated in the GNHS study ($P_{\text{Meta}} < 0.05$), including 7 microbial taxa, 7 metabolites, and 11 proteins, as well as traditional T2D risk factors including age, sex, body mass index (BMI), smoking status, LDL-C, high-density lipoprotein cholesterol (HDL-C), triglycerides (TG), systolic blood pressure (SBP), diastolic blood pressure (DBP), energy intake, family history of diabetes, education, and income. Compared with the model that only included traditional risk factors, the integration of statin-altered multi-omics features substantially improved prediction performance (Figure 5A). The result of the ten-fold cross-validation analysis showed higher accuracy of the integration model (mean value: 0.65 vs. 0.59) (Table S12) and higher area under the curve (AUC; mean value: 0.66 vs. 0.60).

To further demonstrate the associations between statin-induced signatures and T2D or glycemic traits, we performed a regression analysis based on cross-sectional data from the GNHS cohort. Compared with traditional risk factors in predicting T2D (Figure 5B), several identified features, such as creatine, adrenic acid (22:4n-6), and GC, were significantly associated with T2D risk (P values ranged from 2.3×10^{-6} to

0.02) (Figure 5C and D; Table S13), demonstrating a higher ranking. These findings may help interpret the statin-induced higher risk of T2D.

Discussion

In this study, omics-wide MR analyses on 273 signatures across East Asian and European populations revealed that the *HMGCR* gene region was associated with multiple molecular biomarkers that may play critical roles in linking statin use and higher risk of T2D. The protein-wide analysis in Europeans additionally identified GIP as a potential contributor to statin-induced higher risk of T2D. To our knowledge, this is the first systematic investigation of the genetically predicted effects of statin use on multi-omics signatures and the potential biological basis of statin-induced T2D.

Statin-related proteins have been identified through several epidemiological studies [27,28]. The sample size ($n = 135$ to 225) and the number of proteins ($n = 92$ to 350) in these prior studies, however, were limited. Our study substantially expanded the number of proteins and the sample size, identifying C1R, GC, GIP, and HEXB as candidates for preventing statin-induced T2D. C1R modulates the activation of the C1 complex of the classical complement pathway [29], suggesting that the immune system and inflammation may play a role in the development of statin-induced T2D. GC is mostly localized in the liver and alpha cells, and loss of GC leads to smaller, more hyperplastic alpha cells that produce less glucagon in response to low glucose levels [30]. GIP is a hormone secreted by enteroendocrine cells in the gastrointestinal tract [31,32]. Disruption of GIP may lead to dysregulation of the enteroinsular axis, further resulting in insulin resistance and obesity. Higher levels of GIP have been observed in diabetic populations [23], and GIP levels have also been positively associated with all-cause mortality [33]. Our *in vitro* experimental results suggest that the increased effect of statin therapy on GIP may be mediated by the mevalonate pathway, with cholesterol as one downstream product. We hypothesize that reducing cholesterol or other metabolites of the mevalonate pathway may decrease dipeptidyl-peptidase-4 (DPP-4) activity. DPP-4 is a widely expressed enzyme involved in GIP degradation; therefore, the reduction in DPP-4 activity could subsequently lead to an elevation in GIP levels. This hypothesis is partially supported by the observed negative correlation between GIP and LDL-C in a population study [34]. GIP also participates in the gut-bone axis and inhibits bone resorption [32,35], aligning with the protective effect of low-dose statin treatment on bone health [36]. Statin-induced changes in GIP levels may play a significant role in human health. Collectively, these results offer potential repurposing prospects for preventing T2D among patients with statin therapy.

We identified polyunsaturated fatty acids (PUFAs) as candidate mediators for statin-induced T2D. Our data showed that statin use might increase levels of adrenic acid (22:4n-6) and linoleic acid (18:2n-6) in the two ethnic populations. Consistent with our findings, previous studies have reported that patients taking statins exhibit lower plasma levels of n-3 fatty acids or a reduced n-3 to n-6 PUFA ratio [37,38]. The possible mechanism may be that *HMGCR* could regulate key regulatory enzymes involved in the endogenous synthesis of PUFAs, including fatty acid desaturases and fatty acid elongases encoded by *ELOVL* genes [39]. We also found that

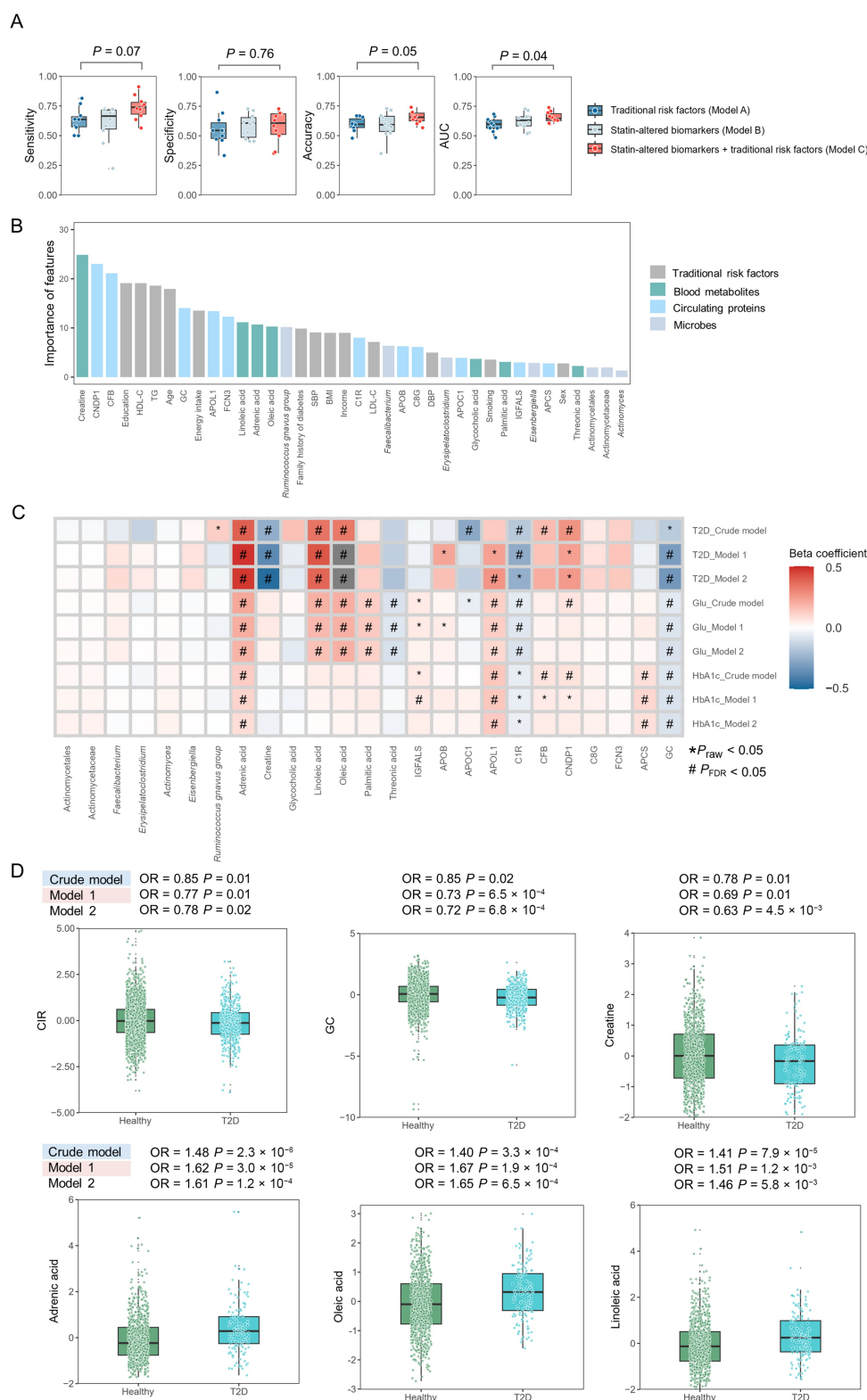


Figure 5 Prediction models integrating identified omics features and their contribution to T2D prediction

A. Performance of prediction models. The paired *t*-test provides statistical significance between the traditional risk factor model (Model A) and the model incorporating statin-altered multi-omics features (Model C). **B.** Contribution of multi-omics and traditional risk factors to the support vector machine model. **C.** Heatmap showing associations between statin-induced multi-omics signatures (microbes, blood metabolites, and circulating proteins) and T2D or glycemic traits in GNHS. Effect sizes are represented as the beta coefficient increase per SD or log-odds unit. **D.** Differences in the levels of identified proteins and metabolites between T2D patients and healthy participants from GNHS. Crude model includes age, sex, and BMI; model 1 adds alcohol drinking status, smoking, energy intake, SBP, DBP, family history of diabetes, education levels, income levels, physical activity, and use of medication for T2D or dyslipidemia; model 2 uses model 1 with the exclusion of participants who took medication for T2D or dyslipidemia. AUC, area under the curve; TG, triglycerides; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; Glu, fasting glucose; HbA1c, hemoglobin A1c; SD, standard deviation; FDR, false discovery rate.

genetically proxied statin use showed putative positive relationships with blood palmitic acid and oleic acid, but putative negative associations with taurodeoxycholic acid, threonic acid, creatine, and glycocholic acid. Palmitic acid and creatine have been linked to T2D or glycemic traits in previous human studies [40,41], supporting our findings. The negative association between threonic acid and glycemic traits was also found in a rodent model [42]. As threonic acid is a breakdown product of ascorbic acid, diminished threonic acid levels embody compromised ascorbic acid metabolism [43], which is suggestive of oxidative stress that plays an important role in the development of diabetes and diabetic complications [44]. Although the exact mechanism linking statins, fatty acid metabolism, and T2D remains unclear, these results highlight potential new metabolic pathways to interpret the statin–T2D association.

To address the potential reverse causation driven by disease progression, we investigated the effect of genetically determined T2D risk on molecular signatures. In the analysis of microbial taxa, metabolites, and plasma proteins, we identified five plasma proteins in the European population that were influenced by both genetically proxied statin use and genetically determined T2D risk ($P < 0.05$). Notably, the directions of effect for genetically determined T2D risk and these molecular features were opposite to those observed for genetically proxied statin use. These findings do not support the hypothesis that associations between genetically proxied statin use and these molecular traits are confounded by T2D progression.

In this study, we identified a wide range of multi-omics features associated with genetically proxied statin use. These signatures may serve as putative intervention targets to prevent T2D among statin users. Our machine learning results emphasize the importance of these biomarkers in predicting T2D, outperforming traditional risk factors. These findings, based on multi-omics data from East Asians and Europeans, reveal a potentially shared etiology for the statin–T2D association across ethnic populations. Given the high global prevalence of hyperlipidemia and the cost-effectiveness of this class of lipid-lowering medications, even modest reductions in T2D risk by intervening on these biomarkers could help alleviate the T2D pandemic.

There are several limitations to the present study. First, we use *HMGCR* variants to mimic the effect of statins, which may not be a perfect proxy, although this has been widely used previously [5,11,14]. Second, we did a cross-ethnic comparison for the effect of genetically proxied statin use on multi-omics signatures, while these signatures are not fully comparable between datasets from the two ethnic groups, with only a small proportion of overlapping biomarkers, which may bias results toward overlapping data. Third, due to the limited number of cases, we used a cross-validation strategy to assess the prediction performance of the multi-omics biomarkers within the GNHS cohort, limiting the clinical relevance of the model to predict T2D risk. External datasets with similar data structures are needed to establish generalizability. Therefore, the generalizability of our prediction model remains to be tested in future studies.

Conclusion

In summary, the cross-ethnic evidence suggests that statin therapy affects molecular signatures, which may potentially

mediate the positive association between statin use and T2D risk. These multi-omics signatures represent potential key intervention targets to prevent statin-induced T2D in future clinical settings.

Materials and methods

Study design and populations

We conducted a two-sample MR analysis to investigate multi-omics signatures of statin therapy. The genetic variant–LDL associations were drawn from the Global Lipids Genetics Consortium (GLGC) [15]. We analyzed 373 overlapping features (102 microbial features, 40 blood metabolites, and 131 circulating proteins) available and shared across the two cohorts of East Asian and European populations. The outcome datasets for MR analyses conducted in Europeans were derived from the following studies: the MiBioGen consortium for gut microbiome ($n = 18,340$) [18], the TwinsUK and KORA datasets for serum metabolites ($n = 7824$) [19], the Icelandic Cancer Project and deCODE genetics for plasma proteomics ($n = 35,559$) [45], and the DIAMANTE meta-analysis (74,124 cases and 824,006 controls) for T2D [16]. For East Asians, the summary statistics of serum proteomics were derived from our published GWAS data [46]. Our in-house cohort, GNHS, provided GWAS summary statistics data for other multi-omics features ($n = 895$ for serum metabolites, $n = 2510$ for erythrocyte membrane fatty acids, and $n = 7935$ for gut microbiota), while the AGEN-T2D study offered publicly available GWAS summary statistics data on T2D (77,418 cases and 356,122 controls) [17].

Furthermore, based on the individual-level data collected in our GNHS study, we developed a machine learning model to prioritize multi-omics targets according to the feature importance in the prediction model.

Multi-omics measurement in the GNHS

The individual participant data were based on the GNHS, involving middle-aged and elderly participants living in Guangzhou, China [47]. Collection of biological samples and questionnaires from each participant was completed upon their entry into this study (2008–2013) and during follow-up visits (every three years after their recruitment). The total energy intake was estimated using data from the food frequency questionnaire (FFQ) based on habitual dietary intake over the past 12 months [48] and calculated using the Chinese Food Consumption Table (2002 edition) [49]. For analysis, energy intake was categorized into quartiles. All participants were required to fast overnight before providing their whole blood samples. Multi-omics measurements (metabolome, proteome, genome-wide genotyping, and gut microbiome) are described in File S1.

MR analysis

We first assessed the genetically determined associations of statin therapy and two other lipid-lowering treatments (targeting LDLR and PCSK9) with T2D risk. We applied the IVW method, accounting for LD among SNPs, to investigate the effects of genetically predicted lipid-lowering treatments on T2D risk. The genetic instruments were drawn from the GLGC [15] and clumped at $r^2 < 0.3$. LD was estimated in 494 unrelated East Asians and 495 unrelated Europeans from the 1000 Genomes Project Phase 3 [50]. The effect size

estimates for lipid-related genetic variants were scaled to a one SD decrease in LDL-C. The MR analyses were performed using the “correl” option in the MendelianRandomization package [51].

HMGCR, the target of statins, was selected for future investigations to mimic the effects of statin treatment on multi-omics signatures. As described above, we used the genetic variants in the *HMGCR* gene region to represent the genetically proxied statin therapy. The IVW approach, integrating the LD matrix when estimating the coefficients, was used to assess the effect of genetically predicted statin use on multi-omics signatures. To avoid violating the MR assumption that instrumental variables are not directly associated with outcomes, we removed the genetic instruments that were significantly associated with outcomes ($P < 5 \times 10^{-8}$). In addition, we used the MR-Egger regression to detect the pleiotropic effects or to assess whether the genetic instruments were related to unmeasured confounders of the exposure–outcome association. Cochran’s Q statistic was used to assess the heterogeneity among genetic variants. Further power calculations for MR analyses were conducted with previously published methods [52].

Putative causal associations between statin (*HMGCR*)-related molecular signatures and T2D risk

For the identified statin-associated proteins, previously reported protein quantitative trait loci (pQTLs) were used as instrumental variables [20,45]. Specifically, for East Asian populations, *cis*-pQTLs were clumped at a significance threshold of 5×10^{-8} and *trans*-pQTLs at 1.6×10^{-10} (i.e., $5 \times 10^{-8}/304$ proteins). Considering the biological plausibility of *cis*-regulatory loci, we used the genome-wide significance threshold for *cis*-acting loci. For European populations, all pQTLs were clumped at a significance threshold of 1.8×10^{-9} . For statin-related blood metabolites, adrenic acid (22:4n-6) had the genetic instruments reported by Hu et al. in East Asians [53] and Guan et al. in Europeans [54]. For statin-related gut microbes, there were no available valid instrumental variables for the downstream analysis. We used the Wald ratio method to estimate the effect of the metabolites or proteins on T2D. We also performed a multivariable analysis with the adjustment of the genetic variants in the *HMGCR* gene region. The coefficient of indirect effect was estimated with the “product of coefficients” [55], and the standard errors were derived from the formula $\sigma_{\alpha\beta} = \sqrt{\alpha^2\sigma_\beta^2 + \beta^2\sigma_\alpha^2 - \sigma_\alpha^2\sigma_\beta^2}$.

Reverse MR analysis

We clumped the T2D-related genetic instruments ($LD < 0.05$, $P < 5 \times 10^{-8}$) from the following studies: the DIAMANTE meta-analysis (74,124 cases and 824,006 controls) for East Asians [16] and the AGEN-T2D study (77,418 cases and 356,122 controls) for Europeans [17]. The outcome datasets were the same as those used in the molecular trait analyses related to statin use. To improve computational efficiency and account for pleiotropic effects, we employed the GSMR method and HEIDI test for large-scale reverse MR analyses [26]. We reported results with $P < 0.05$ and highlighted associations with FDR < 0.05 .

Integration of pQTLs and T2D

We used two colocalization methods to identify putative proteins mediating the relationship between statins and T2D risk. The *coloc* method was performed to test whether the

proteins and T2D shared a common genetic architecture [56]. For the colocalization analysis, we set the prior probability that a SNP is associated with protein levels of T2D risk at 1×10^{-4} , and the prior probability that a SNP is associated with both traits at 1×10^{-5} . Furthermore, we used the SMR and HEIDI tests to investigate the potential causal relationships [21].

Machine learning model for T2D prediction

The analysis was based on 260 GNHS participants (130 T2D cases and 130 randomly selected controls). The model was trained with the traditional T2D risk factors, including age, sex, BMI, smoking, LDL-C, HDL-C, TG, education, SBP, DBP, total energy intake, family history of diabetes, and income, as well as statin-altered signatures measured in the GNHS, including 7 taxa, 7 metabolites, and 11 proteins. Missing values in the proteomic data were imputed using half the minimum observed value. To ensure normality, the statin-altered signature data were rank-based inverse normal transformed (RINT). The support vector machine method with ten-fold cross-validation was used to compare the multi-omics features with the traditional risk factors and capture the important features contributing to T2D risk, according to the average importance in the ten-fold cross-validation. All analyses were conducted using the R software (v4.3.3) with the *e1071* package (v1.7-14).

Observational associations between identified molecular signatures and T2D risk

To characterize the direction of associations, based on the GNHS study ($n = 971$ to 1943), we conducted logistic regression to assess the relationships between the identified multi-omics features and T2D, and used linear regression to estimate their associations with HbA1c and fasting glucose, adjusting for potential confounding factors. We fitted three models: crude model included age, sex, and BMI; model 1 added alcohol drinking status, smoking, SBP, DBP, energy intake, family history of diabetes, education, income, physical activity, and medication for T2D or dyslipidemia; model 2 was model 1 with the exclusion of participants who took medication for T2D or dyslipidemia.

Ethical statement

The study protocol was approved by the Ethics Committee of the School of Public Health at Sun Yat-sen University (Approval No. 2018048) and the Ethics Committee of Westlake University (Approval No. 20190114ZJS003), China. All participants provided written informed consent.

Data availability

A website developed for querying MR estimates of the effects of statin use on molecular phenotypes is available at <https://omics.lab.westlake.edu.cn/resource/statint2d.html>. The raw 16S rRNA sequencing data for each cohort have been deposited in the Genome Sequence Archive [57] at the National Genomics Data Center (NGDC), China National Center for Bioinformation (CNCB) (GSA: CRA006769), and are publicly accessible at <https://ngdc.cncb.ac.cn/gsa>. The raw data for serum proteomes have been uploaded to iProX (ProteomeXchange: PXD039236, PXD039231, and PXD038253), and are publicly accessible at <https://www.iprox.cn>. Other datasets generated and analyzed in this study

are available upon reasonable request by bona fide researchers for specified scientific purposes via contacting the corresponding authors.

CRedit author statement

Fengzhe Xu: Methodology, Investigation, Formal analysis, Writing – original draft. **Min Yang:** Methodology, Investigation, Formal analysis, Writing – original draft. **Wei Hu:** Data curation. **Shuai Yuan:** Writing – review & editing. **Xue Cai:** Data collection. **Wanglong Gou:** Data collection. **Zelei Miao:** Data curation. **Bang-yan Li:** Data curation. **Liang Yue:** Data collection. **Zhangzhi Xue:** Data collection. **Menglei Shuai:** Data collection. **Luqi Shen:** Writing – review & editing. **Yuanqing Fu:** Writing – review & editing. **Tiannan Guo:** Conceptualization. **Yu-ming Chen:** Conceptualization. **Ju-Sheng Zheng:** Conceptualization, Supervision, Writing – review & editing. All authors have read and approved the final manuscript.

Competing interests

The authors have declared no competing interests.

Supplementary material

Supplementary material is available at *Genomics, Proteomics & Bioinformatics* online (<https://doi.org/10.1093/gpbjnl/qzaf101>).

Acknowledgments

We thank the High-Performance Computing Center and the Research Center for Industries of the Future at Westlake University for assistance in computing. This study was funded by the National Natural Science Foundation of China (Grant Nos. 82073529, 81903316, 82073546, and 81773416), the Zhejiang Ten-thousand Talents Program (Grant No. 2019R52039), the 5010 Program for Clinical Researches (Grant No. 2007032) of the Sun Yat-sen University (Guangzhou, China), and the National Key R&D Program of China (Grant Nos. 2021YFA1301602, 2021YFA1301601, 2021YFA1301603, and 2020YFE0202200). The funders had no role in collecting data, study design, data interpretation, or the decision to submit the manuscript for publication.

ORCID

0000-0002-2686-1445 (Fengzhe Xu)
0009-0000-5429-9253 (Min Yang)
0000-0001-9332-9670 (Wei Hu)
0000-0001-5055-5627 (Shuai Yuan)
0000-0002-7427-2100 (Xue Cai)
0000-0002-4256-9452 (Wanglong Gou)
0000-0001-9478-1933 (Zelei Miao)
0000-0001-7446-2505 (Bang-yan Li)
0000-0001-6478-7306 (Liang Yue)
0000-0001-7311-6221 (Zhangzhi Xue)
0000-0001-9983-1477 (Menglei Shuai)
0009-0003-4349-4063 (Luqi Shen)
0000-0002-3955-9376 (Yuanqing Fu)
0000-0003-3869-7651 (Tiannan Guo)
0000-0003-1658-5528 (Yu-ming Chen)
0000-0001-6560-4890 (Ju-Sheng Zheng)

References

- [1] Ramkumar S, Raghunath A, Raghunath S. Statin therapy: review of safety and potential side effects. *Acta Cardiol Sin* 2016; 32:631–39.
- [2] Marcus ME, Manne-Goehler J, Theilmann M, Farzadfar F, Moghaddam SS, Keykhaei M, et al. Use of statins for the prevention of cardiovascular disease in 41 low-income and middle-income countries: a cross-sectional study of nationally representative, individual-level data. *Lancet Glob Health* 2022;10:e369–79.
- [3] Cederberg H, Stančáková A, Yaluri N, Modi S, Kuusisto J, Laakso M. Increased risk of diabetes with statin treatment is associated with impaired insulin sensitivity and insulin secretion: a 6 year follow-up study of the METSIM cohort. *Diabetologia* 2015; 58:1109–17.
- [4] Kim J, Lee HS, Lee KY. Effect of statins on fasting glucose in non-diabetic individuals: nationwide population-based health examination in Korea. *Cardiovasc Diabetol* 2018;17:155.
- [5] Liu G, Shi M, Mosley JD, Weng C, Zhang Y, Lee MTM, et al. A mendelian randomization approach using 3-HMG-coenzyme-A reductase gene variation to evaluate the association of statin-induced low-density lipoprotein cholesterol lowering with non-cardiovascular disease phenotypes. *JAMA Netw Open* 2021; 4:e2112820.
- [6] Sattar N, Preiss D, Murray HM, Welsh P, Buckley BM, de Craen AJM, et al. Statins and risk of incident diabetes: a collaborative meta-analysis of randomised statin trials. *Lancet* 2010; 375:735–42.
- [7] Caparrós-Martín JA, Lareu RR, Ramsay JP, Peplies J, Reen FJ, Headlam HA, et al. Statin therapy causes gut dysbiosis in mice through a PXR-dependent mechanism. *Microbiome* 2017;5:95.
- [8] Ryan PM, London LEE, Bjorndahl TC, Mandal R, Murphy K, Fitzgerald GF, et al. Microbiome and metabolome modifying effects of several cardiovascular disease interventions in apo-E^{−/−} mice. *Microbiome* 2017;5:30.
- [9] Khan TJ, Ahmed YM, Zamzami MA, Siddiqui AM, Khan I, Baothman OAS, et al. Atorvastatin treatment modulates the gut microbiota of the hypercholesterolemic patients. *OMICS* 2018; 22:154–63.
- [10] Nozue T, Michishita I. Statin treatment alters serum n-3 to n-6 polyunsaturated fatty acids ratio in patients with dyslipidemia. *Lipids Health Dis* 2015;14:67.
- [11] Ference BA, Ray KK, Catapano AL, Ference TB, Burgess S, Neff DR, et al. Mendelian randomization study of ACLY and cardiovascular disease. *N Engl J Med* 2019;380:1033–42.
- [12] Zheng J, Haberland V, Baird D, Walker V, Haycock PC, Hurler MR, et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat Genet* 2020;52:1122–31.
- [13] Gill D, Georgakis MK, Walker VM, Schmidt AF, Gkatzionis A, Freitag DF, et al. Mendelian randomization for studying the effects of perturbing drug targets. *Wellcome Open Res* 2021; 6:16.
- [14] Carter P, Vithayathil M, Kar S, Potluri R, Mason AM, Larsson SC, et al. Predicting the effect of statins on cancer risk using genetic variants from a Mendelian randomization study in the UK Biobank. *Elife* 2020;9:e57191.
- [15] Graham SE, Clarke SL, Wu KH, Kanoni S, Zajac GJM, Ramdas S, et al. The power of genetic diversity in genome-wide association studies of lipids. *Nature* 2021;600:675–9.
- [16] Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet* 2018;50:1505–13.
- [17] Spracklen CN, Horikoshi M, Kim YJ, Lin K, Bragg F, Moon S, et al. Identification of type 2 diabetes loci in 433,540 East Asian individuals. *Nature* 2020;582:240–5.
- [18] Kurilshikov A, Medina-Gomez C, Bacigalupe R, Radjabzadeh D, Wang J, Demirkan A, et al. Large-scale association analyses

- identify host factors influencing human gut microbiome composition. *Nat Genet* 2021;53:156–65.
- [19] Shin SY, Fauman EB, Petersen AK, Krumsiek J, Santos R, Huang J, et al. An atlas of genetic influences on human blood metabolites. *Nat Genet* 2014;46:543–50.
 - [20] Xu F, Yu EY, Cai X, Yue L, Jing L, Liang X, et al. Genome-wide genotype-serum proteome mapping provides insights into the cross-ancestry differences in cardiometabolic disease susceptibility. *Nat Commun* 2023;14:896.
 - [21] Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* 2016;48:481–7.
 - [22] Wu Y, Zeng J, Zhang F, Zhu Z, Qi T, Zheng Z, et al. Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nat Commun* 2018;9:918.
 - [23] Crockett SE, Mazzaferri EL, Cataland S. Gastric inhibitory polypeptide (GIP) in maturity-onset diabetes mellitus. *Diabetes* 1976;25:931–5.
 - [24] Boerma M, Fu Q, Wang J, Loose DS, Bartolozzi A, Ellis JL, et al. Comparative gene expression profiling in three primary human cell lines after treatment with a novel inhibitor of Rho kinase or atorvastatin. *Blood Coagul Fibrinolysis* 2008;19:709–18.
 - [25] Boerma M, Burton GR, Wang J, Fink LM, McGehee RE, Hauer-Jensen M. Comparative expression profiling in primary and immortalized endothelial cells: changes in gene expression in response to hydroxy methylglutaryl-coenzyme A reductase inhibition. *Blood Coagul Fibrinolysis* 2006;17:173–80.
 - [26] Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R, et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat Commun* 2018;9:224.
 - [27] Toribio M, Fitch KV, Stone L, Zanni MV, Lo J, de Filippi C, et al. Assessing statin effects on cardiovascular pathways in HIV using a novel proteomics approach: analysis of data from INTREPID, a randomized controlled trial. *EBioMedicine* 2018;35:58–66.
 - [28] deFilippi C, Toribio M, Wong LP, Sadreyev R, Grundberg I, Fitch KV, et al. Differential plasma protein regulation and statin effects in human immunodeficiency virus (HIV)-infected and non-HIV-infected patients utilizing a proteomics approach. *J Infect Dis* 2020;222:929–39.
 - [29] Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 2000;28:27–30.
 - [30] Vioria K, Hewison M, Hodson DJ. Vitamin D binding protein/GC-globulin: a novel regulator of alpha cell function and glucagon secretion. *J Physiol* 2022;600:1119–33.
 - [31] Stensen S, Gasbjerg LS, Krogh LL, Skov-Jepsen K, Sparre-Ulrich AH, Jensen MH, et al. Effects of endogenous GIP in patients with type 2 diabetes. *Eur J Endocrinol* 2021;185:33–45.
 - [32] Holst JJ, Rosenkilde MM. GIP as a therapeutic target in diabetes and obesity: insight from incretin co-agonists. *J Clin Endocrinol Metab* 2020;105:e2710–6.
 - [33] Juić A, Atabaki-Pasdar N, Nilsson PM, Almgren P, Hakaste L, Tuomi T, et al. Glucose-dependent insulinotropic peptide and risk of cardiovascular events and mortality: a prospective study. *Diabetologia* 2020;63:1043–54.
 - [34] Møller CL, Vistisen D, Færch K, Johansen NB, Witte DR, Jonsson A, et al. Glucose-dependent insulinotropic polypeptide is associated with lower low-density lipoprotein but unhealthy fat distribution, independent of insulin: the ADDITION-PRO study. *J Clin Endocrinol Metab* 2016;101:485–93.
 - [35] Gasbjerg LS, Hartmann B, Christensen MB, Lanng AR, Vilsbøll T, Jørgensen NR, et al. GIP's effect on bone metabolism is reduced by the selective GIP receptor antagonist GIP(3–30)NH₂. *Bone* 2020;130:115079.
 - [36] Leutner M, Matzhold C, Bellach L, Deisinger C, Harreiter J, Thurner S, et al. Diagnosis of osteoporosis in statin-treated patients is dose-dependent. *Ann Rheum Dis* 2019;78:1706–11.
 - [37] Nozue T, Yamamoto S, Tohyama S, Fukui K, Umezawa S, Onishi Y, et al. Comparison of effects of serum n-3 to n-6 polyunsaturated fatty acid ratios on coronary atherosclerosis in patients treated with pitavastatin or pravastatin undergoing percutaneous coronary intervention. *Am J Cardiol* 2013;111:1570–5.
 - [38] Kurisu S, Ishibashi K, Kato Y, Mitsuba N, Dohi Y, Nishioka K, et al. Effects of lipid-lowering therapy with strong statin on serum polyunsaturated fatty acid levels in patients with coronary artery disease. *Heart Vessels* 2013;28:34–8.
 - [39] Ishihara N, Suzuki S, Tanaka S, Watanabe Y, Nagayama D, Saiki A, et al. Atorvastatin increases *Fads1*, *Fads2* and *Elovl5* gene expression via the geranylgeranyl pyrophosphate-dependent Rho kinase pathway in 3T3-L1 cells. *Mol Med Rep* 2017;16:4756–62.
 - [40] Gualano B, DE Salles Pannelli V, Roschel H, Artioli GG, Neves M Jr, De Sá Pinto AL, et al. Creatine in type 2 diabetes: a randomized, double-blind, placebo-controlled trial. *Med Sci Sports Exerc* 2011;43:770–8.
 - [41] Forouhi NG, Koulman A, Sharp SJ, Imamura F, Kröger J, Schulze MB, et al. Differences in the prospective association between individual plasma phospholipid saturated fatty acids and incident type 2 diabetes: the EPIC-InterAct case-cohort study. *Lancet Diabetes Endocrinol* 2014;2:810–8.
 - [42] Piccolo BD, Graham JL, Stanhope KL, Fiehn O, Havel PJ, Adams SH. Plasma amino acid and metabolite signatures tracking diabetes progression in the UCD-T2DM rat model. *Am J Physiol Endocrinol Metab* 2016;310:E958–69.
 - [43] Fahrman J, Grapov D, Yang J, Hammock B, Fiehn O, Bell GI, et al. Systemic alterations in the metabolome of diabetic NOD mice delineate increased oxidative stress accompanied by reduced inflammation and hypertriglyceremia. *Am J Physiol Endocrinol Metab* 2015;308:E978–89.
 - [44] Giacco F, Brownlee M. Oxidative stress and diabetic complications. *Circ Res* 2010;107:1058–70.
 - [45] Ferkingstad E, Sulem P, Atlason BA, Sveinbjornsson G, Magnusson MI, Styrismisdottir EL, et al. Large-scale integration of the plasma proteome with genetics and disease. *Nat Genet* 2021;53:1712–21.
 - [46] Xu F, Yu EY, Cai X, Yue L, Jing L, Liang X, et al. Genome-wide genotype-serum proteome mapping provides insights into the cross-ancestry differences in cardiometabolic disease susceptibility. *Nat Commun* 2023;14:896.
 - [47] Cao Y, Wang C, Guan K, Xu Y, Su YX, Chen YM. Association of magnesium in serum and urine with carotid intima-media thickness and serum lipids in middle-aged and elderly Chinese: a community-based cross-sectional study. *Eur J Nutr* 2016;55:219–26.
 - [48] Zhang CX, Ho SC. Validity and reproducibility of a food frequency questionnaire among Chinese women in Guangdong province. *Asia Pac J Clin Nutr* 2009;18:240–50.
 - [49] Liu B, Woo J, Tang N, Ng K, Ip R, Yu A. Assessment of total energy expenditure in a Chinese population by a physical activity questionnaire: examination of validity. *Int J Food Sci Nutr* 2001;52:269–82.
 - [50] Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. *Nature* 2015;526:68–74.
 - [51] Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int J Epidemiol* 2017;46:1734–9.
 - [52] Brion MJ, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. *Int J Epidemiol* 2013;42:1497–501.
 - [53] Hu Y, Li H, Lu L, Manichaikul A, Zhu J, Chen YD, et al. Genome-wide meta-analyses identify novel loci associated with n-3 and n-6 polyunsaturated fatty acid levels in Chinese and European-ancestry populations. *Hum Mol Genet* 2016;25:1215–24.
 - [54] Guan W, Steffen BT, Lemaitre RN, Wu JHY, Tanaka T, Manichaikul A, et al. Genome-wide association study of plasma n6 polyunsaturated fatty acids within the cohorts for heart and

- aging research in Genomic Epidemiology Consortium. *Circ Cardiovasc Genet* 2014;7:321–31.
- [55] VanderWeele TJ. Mediation analysis: a practitioner's guide. *Annu Rev Public Health* 2016;37:17–32.
- [56] Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* 2014;10:e1004383.
- [57] Zhang S, Chen X, Jin E, Wang A, Chen T, Zhang X, et al. The GSA Family in 2025: a broadened sharing platform for multi-omics and multimodal data. *Genomics Proteomics Bioinformatics* 2025;23:qzaf072.