



OPEN Identification of causal plasma metabolite biomarkers for ischemic stroke using Mendelian randomization and mediation analysis

Jian Lyu^{1✉}, Fumei Liu², Yan Chai³, Xiting Wang^{4✉}, Yi Liu^{2✉} & Yanming Xie^{2✉}

The Global Burden of Disease Study 2021 reports that stroke remains a leading cause of death, with ischemic stroke (IS) presenting significant challenges in screening, prevention, and treatment. We explored the causal effects of 1,400 plasma metabolites on IS outcomes using a two-sample Mendelian randomization (MR) framework. We assessed causal relationships between IS and 11 common clinical risk factors and further examined these relationships for metabolites. Mediation analysis identified mechanisms for metabolites affecting both IS and its risk factors. Finally, a phenome-wide association study (PheWAS) MR analysis evaluated the side effects and additional indications of IS-associated metabolites across 3,948 phenotypes from the UKBB GWAS. Nineteen metabolites showed a causal relationship with IS. MR analysis confirmed body mass index (BMI), high-density lipoprotein (HDL), systolic blood pressure (SBP), diastolic blood pressure (DBP), and type 2 diabetes (T2D) as risk factors for IS. Among 136 metabolites associated with at least one IS risk factor, 132 were linked to risk factors but not directly to IS. BMI, DBP, and coffee intake mediated the causal relationship between IS and the levels of 1-stearoyl-GPG (18:0), 1-oleoyl-2-linoleoyl-GPE (18:1/18:2), Octadecadienedioate (C18:2-DC), and X-24,951. Phe-MR analysis indicated that these metabolites were protective and affected other indications similarly to IS. Our findings reveal causal pathways and identify four potential biomarkers for IS, providing new insights for its screening, prevention, and treatment.

Keywords Ischemic stroke, Human plasma metabolomics, Mendelian randomization, Biomarkers, Mediating effects

Abbreviations

IS	Ischemic stroke
MR	Mendelian randomization
PheWAS	Phenome-wide association study
BMI	Body mass index
LDL	Low-density lipoprotein
HDL	High-density lipoprotein
non-HDL	Non-high-density lipoprotein
SBP	Systolic blood pressure
DBP	Diastolic blood pressure
T2D	Type 2 diabetes
CHD	Coronary heart disease

¹NMPA Key Laboratory for Clinical Research and Evaluation of Traditional Chinese Medicine & National Clinical Research Center for Chinese Medicine Cardiology, XiYuan Hospital, China Academy of Chinese Medical Sciences, No.1 Xiyuan playground Road, Haidian District, Beijing 100091, PR China. ²Institute of Basic Research in Clinical Medicine, China Academy of Chinese Medical Sciences, No.16 Nanxiaojie, Inner Dongzhimen, Dongcheng District, Beijing 100700, PR China. ³Department of Epidemiology, University of California, 405 Hilgard Avenue, Los Angeles 90095, CA, USA. ⁴Academy of Mathematics and Systems Science, Chinese Academy of Sciences, No. 55 Zhongguancun East Road, Beijing 100190, China. ✉email: dr.jianlyu-tcm@outlook.com; wangxiting@amss.ac.cn; dr.liuyi93@gmail.com; ktzu2018@163.com

GBD	Global burden of disease
DALYs	Disability-adjusted life years
CIRI	Cerebral ischemia-reperfusion injury
CKB	China Kadoorie Biobank
GWAS	Genome-wide association studies
CLSA	Canadian longitudinal study on aging
SNPs	Single nucleotide polymorphisms
MR-PRESSO	MR pleiotropy residual sum and outlier
PheWAS	Phenome-wide association study
STROBE-MR	Strengthening the reporting of observational studies in epidemiology using mendelian randomization
IEU	Integrative epidemiology unit
FDR	False discovery rate
SE	Standard error
CI	Confidence interval
IVW	Inverse-variance weighted
Phe-MR	Phenome-wide MR
ETS	Environmental tobacco smoke
ACS	Acute coronary syndrome

Ischemic stroke (IS) remains a leading cause of death globally, accounting for 70–80% of all stroke cases^{1–3}. According to the Global Burden of Disease Study (GBD 2021), approximately 7.63 million individuals were diagnosed with IS in 2019, leading to an economic loss of approximately \$96.451 billion, placing a significant burden on the global economy^{4,5}. The “China Stroke Report 2020” identified stroke as the primary cause of premature death in China, with a notable increase in IS cases. Furthermore, 41% and 17% of IS patients experience recurrence within five years and one year, respectively^{6–8}. According to the 2019 GBD data, stroke is the leading cause of disability-adjusted life years (DALYs) in China, surpassing heart disease and cancers of the respiratory and digestive systems^{9,10}. In 2019, China recorded 4.33 million hospitalizations for IS, with hospital costs quadrupling since 2010, posing a significant health burden and prevention challenge¹¹. Given the high burden of IS and the limitations of secondary prevention, identifying modifiable risk factors and upstream molecular mechanisms has become increasingly important.

Key modifiable risk factors for IS prevention include hypertension, diabetes, dyslipidemia, heart disease, smoking, alcohol intake, diet, overweight/obesity, physical inactivity, and psychological factors^{9,12,13}. Hypertension is the most significant risk factor for stroke, with persistent hypertension after IS increasing the risk of poor outcomes and recurrence within one year^{14–16}. Diabetes affects approximately 14% of adult stroke patients in China and significantly increases stroke-related mortality^{17,18}. Dyslipidemia independently contributes to stroke risk; each 1 mmol/L increase in total cholesterol and low-density lipoprotein cholesterol significantly raises IS risk^{19,20}. A study with a 34-year follow-up found that age-adjusted stroke incidence was over double for patients with coronary heart disease compared to those without²¹. Smoking increases stroke occurrence and worsens prognosis, whereas quitting smoking substantially reduces recurrence and mortality²². Alcohol consumption proportionally increases stroke-related mortality²³. Excess weight and obesity significantly elevate stroke risk, with metabolic abnormalities further amplifying this risk^{24,25}. Clarifying the genetic underpinnings of these modifiable risk factors is essential for identifying precise biomarkers and therapeutic targets, promoting personalized prevention and clinical intervention strategies.

The rapidly evolving field of metabolomics plays a pivotal role in elucidating IS mechanisms, facilitating the systematic identification of metabolites and pathways essential in IS prevention and treatment, and highlighting potential therapeutic targets. Compared to proteomics or transcriptomics, metabolomics uniquely detects small-molecule metabolites capable of crossing the blood-brain barrier, offering distinct advantages for IS biomarkers discovery²⁶. Metabolites, as intermediate or final products of metabolic reactions, influence disease risk and therapeutic targets. Mendelian randomization (MR), which leverages genetic variation from genome-wide association studies (GWAS), investigates causal relationships between exposures and diseases while minimizing confounding and reverse causality^{27,28,32}. In 2023, a large-scale GWAS published in *Nature Genetics* analysed 1,091 metabolites and 309 metabolite ratios from 8,299 individuals in the Canadian Longitudinal Study on Aging (CLSA), providing insights into their genetic architecture and relevance to IS, presenting opportunities for therapeutic target discovery²⁹. A systematic search revealed only one study exploring the correlation between 1,400 plasma metabolites and IS³⁰. No study has systematically explored potential metabolite targets demonstrating a causal relationship with IS and IS-related risk factors among these 1,400 plasma metabolites or conducted mediation effects and MR analysis across the entire phenome. To systematically investigate the causal pathways and potential therapeutic targets between the human plasma metabolome and IS, we conducted a two-sample MR analysis involving 1,400 plasma metabolites. Genetic instrumental were derived from GWAS data from 8,299 participants, and summary statistics for IS and 11 established risk factors were obtained from the Integrated Epidemiology Unit database. We further examined the mediating roles of these risk factors in the metabolite-IS associations and explored the disease implication of IS-associated metabolites through a phenome-wide MR analysis across 3,948 phenotypes from the UKBB GWAS³¹.

Materials and methods

Study design

In this study, we initially conducted a two-sample MR analysis on 1,400 metabolites and IS to investigate their causal relationship. Subsequently, we employed two-sample MR to identify the causal relationships between

IS and 11 common clinical risk factors: body mass index (BMI), low-density lipoprotein (LDL), high-density lipoprotein (HDL), non-high-density lipoprotein (non-HDL), systolic blood pressure (SBP), diastolic blood pressure (DBP), smoking, alcohol consumption, coffee intake, type 2 diabetes (T2D), and coronary heart disease (CHD). Additionally, we performed two-sample MR analyses to determine the causal relationships between metabolites and each positive risk factor. We conducted mediation analysis on metabolites with causal relationships to both IS and IS-related risk factors, aiming to elucidate the underlying mechanisms through which these metabolites affect IS by influencing IS risk factors. Finally, we carried out a PheWAS MR analysis to comprehensively explore metabolites with causal regulatory relationships with IS, assessing potential side effects and additional indications for IS-associated metabolites across 3,948 phenotypes. The overall study design is illustrated in Fig. 1. The STROBE-MR (Strengthening the Reporting of Observational Studies in Epidemiology using Mendelian Randomization) checklist was completed for this observational study^{32,33}.

Data sources

We obtained summary data for 1,091 plasma metabolites and 309 metabolite ratios from a large genome-wide association study (GWAS) involving 8,299 European individuals in the Canadian Longitudinal Study on Aging (CLSA) cohort²⁹. The data is accessible on the GWAS Summary Statistics. For ischemic stroke (34,217 cases and 406,111 controls) and 11 risk factors (BMI, LDL, HDL, non-HDL, SBP, DBP, smoking, alcohol consumption, coffee intake, T2D, and CHD), we retrieved GWAS data (ebi-a-GCST005843) from the Integrative Epidemiology Unit (IEU) database (<https://gwas.mrcieu.ac.uk/>). The search keywords included: ischemic stroke/IS, body mass index/BMI, low-density lipoprotein/LDL, high-density lipoprotein/HDL, non-high-density lipoprotein/non-HDL, systolic blood pressure/SBP, diastolic blood pressure/DBP, smoking, alcohol, coffee, type 2 diabetes, and coronary heart disease/CHD. The search covered the database's entire duration up to February 9, 2024. Samples for MR analysis were prioritized based on their sample size. We utilized summary data from the UK Biobank's online repository for PheWAS analysis. The data, curated in the IEU database, includes 3,948 phenotypes encompassing diagnoses, current health status, treatment records, biochemical analyses, body and anthropometric measurements, family history, lifestyle factors, and mental health. Detailed information about the sources, original questionnaires, or measurements can be found on the UK Biobank. Detailed information on the GWAS datasets used for ischemic stroke and its risk factors is summarized in Table 1.

Identification of causal plasma metabolites for IS

We extracted exposure data for 1,400 plasma metabolites and conducted SNP selection based on a genome-wide significance threshold of $p < 1e-05$. SNPs with linkage disequilibrium ($kb = 10,000$, $r^2 = 0.001$) and weak instruments ($F\text{-statistic} < 10$) were removed. The causal effects of each metabolite on IS were evaluated using five MR methods: MR Egger, weighted median, inverse variance weighted, simple mode, and weighted mode, as described in the Statistical Analysis section. To ensure the directionality of the causal relationship, we further performed reverse MR analysis by treating IS as the exposure and positive metabolites as the outcome to rule out reverse causality.

Identification of causal risk factors for IS

Based on stroke risk factors outlined in the 2021 Guideline for the Prevention of Stroke in Patients With Stroke and Transient Ischemic Attack⁹, we selected 11 common clinical risk factors for IS: BMI, LDL, HDL, non-HDL, SBP, DBP, smoking, alcohol consumption, coffee intake, T2D, and CHD, to establish causal inferences with IS. GWAS summary data for these exposure factors were obtained, and SNPs were selected using $p < 5 \times 10^{-8}$ as the significance threshold. Instrumental variables with linkage disequilibrium or weak instrument strength were excluded. MR analysis was performed to assess the causal effects of each risk factor on IS, following the procedures detailed in the Statistical Analysis section.

Identification of causal metabolites for IS risk factors

We conducted MR analysis to explore the causal effects of 1,400 metabolites on the 11 IS-related risk factors. SNPs were selected using the same criteria as described above. MR analysis procedures and filtering criteria were consistent with those described in the Statistical Analysis section.

Mediation analysis for metabolites, risk factors, and IS

For metabolites causally associated with both IS and risk factors, we performed mediation analysis to quantify indirect through the identified risk factors. Inclusion criteria were: (a) positive MR results for metabolites and risk factors; (b) positive MR results for risk factors and IS; (c) positive causal association between metabolites and IS without reverse causation. To distinguish between direct and indirect effects, we employed two-step MR results. The Product method was used to estimate the beta of the indirect effect, while the Delta method was used to calculate the standard error (SE) and confidence interval (CI)³⁴.

Phenome-wide MR (Phe-MR) analysis of 3948 phenotypes for metabolites causally associated with IS

Phenome-Wide Association Study (PheWAS) is a method that examines associations between a specific SNP or phenotype and all phenotypes within a phenome. To extend the exploration of side effects and additional indications for the nineteen IS-associated metabolites to non-IS phenotypes, we conducted Phe-MR analyses across a broad spectrum of diseases. Using metabolites with positive MR results for IS and leveraging summary data from the IEU database encompassing 3,948 phenotypes in the UK Biobank, we performed PheWAS-MR analyses. Genetic instruments for the 19 significant metabolites were selected using the same criteria as described

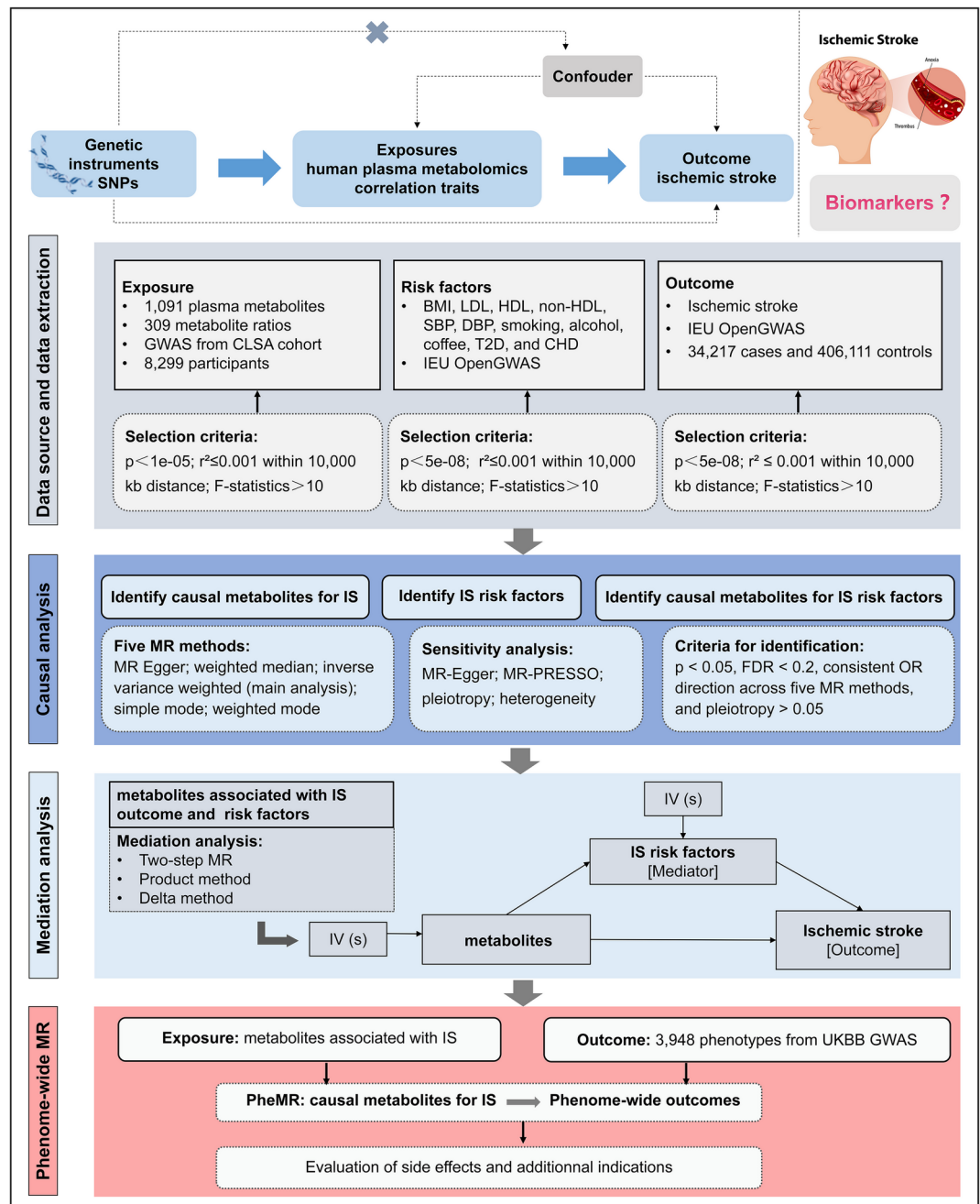


Fig. 1. Overall study design. IS ischemic stroke, MR Mendelian randomization, SNPs single nucleotide polymorphisms, GWAS genome-wide association studies, CLSA Canadian Longitudinal Study on Aging, BMI body mass index, LDL low-density lipoprotein, HDL high-density lipoprotein, non-HDL non-high-density lipoprotein, SBP systolic blood pressure, DBP diastolic blood pressure, T2D type 2 diabetes, CHD coronary heart disease, MR-PRESSO MR pleiotropy residual sum and outlier, FDR false discovery rate, OR odds ratio, IVs instrumental variables, PheMR phenome-wide MR.

previously. MR analyses were performed as detailed in the Statistical Analysis section. Associations with $p < 0.05$ were considered suggestive.

Statistical analysis

MR estimates were based on five methods (MR Egger, weighted median, inverse variance weighted, simple mode, and weighted mode). The primary MR estimates were calculated using the inverse-variance weighted (IVW) method under multiplicative random effects. Instrumental variables (SNPs) were selected using genome-wide significance thresholds ($p < 1 \times 10^{-5}$ for metabolites and $p < 5 \times 10^{-8}$ for risk factors), with linkage disequilibrium (LD) pruning ($kb = 10,000$, $r^2 = 0.001$) and exclusion of weak instruments (F-statistic < 10). The Benjamini-Hochberg method that controls the FDR was applied to correct for multiple testing. The association

Trait / Exposure	Abbreviation	GWAS source	Sample size (Cases / Controls or total)	Ancestry	ID
Ischemic stroke	IS	IEU/GWAS catalog	34,217 / 406,111	European	ebi-a-GCST005843
Body mass index	BMI	GIANT / IEU	681,275	European	ieu-b-40
LDL cholesterol	LDL	IEU	431,167	European	ebi-a-GCST90002412
HDL cholesterol	HDL	UK Biobank / IEU	403,943	European	ieu-b-109
Non-HDL cholesterol	non-HDL	IEU	115,078	European	met-d-Remnant_C
Systolic BP	SBP	IEU/GWAS catalog	810,865	European	ebi-a-GCST90000062
Diastolic BP	DBP	IEU/GWAS catalog	810,865	European	ebi-a-GCST90000059
Smoking status	Smoking	Within family GWAS consortium / IEU	99,996	European	ieu-b-4858
Alcohol consumption	Alcohol	UK Biobank / IEU	112,117	European	ieu-a-1283
Coffee intake	Coffee	MRC-IEU / IEU	428,860	European	ukb-b-5237
Type 2 diabetes	T2D	IEU/GWAS catalog	1,178 / 61,714	European	ebi-a-GCST006867
Coronary heart disease	CHD	GWAS catalog	8,192	European	GCST90199621

Table 1. Summary of GWAS datasets for ischemic stroke and its clinical risk factors.

with a Benjamini–Hochberg adjusted p -value < 0.2 ^{35,36} was deemed statistically significant. The MR results were filtered based on $p < 0.05$, FDR < 0.2 , consistent OR direction across five MR methods, and pleiotropy > 0.05 . Sensitivity analysis was conducted to validate the robustness of MR results, including MR-Egger, MR-PRESSO, pleiotropy, and heterogeneity. The mediation effect was calculated by the formula: 1– direct effect (the estimate after adjusting for the mediator)/total effect (the estimate in the univariable MR analysis). All analyses were two-sided and performed using the Two Sample MR, Mendelian Randomization, and MRPRESSO R packages in R software 4.2.1.

Results
Identification of IS-associated metabolites

We obtained GWAS data for 1,400 metabolites and IS GWAS data (ebi-a-GCST005843) from the CLSA cohort and IEU databases²⁹. We conducted an association analysis on the metabolite data, excluding SNPs in linkage disequilibrium. Five MR methods were applied to determine the causal relationships between these metabolites and IS, estimating ORs, heterogeneity, and pleiotropy. Additionally, scatter plots, forest plots, funnel plots, and leave-one-out sensitivity analysis forest plots were generated. We summarized the overall results in a circular clustering heatmap (Fig. 2). We identified 19 IS-associated metabolites based on p -value, FDR, and OR values, excluding those with pleiotropy. Using MR-PRESSO, we identified 19 metabolites causally associated with IS, including levels of 1-linoleoylglycerol (18:2), 1-stearoyl-GPG (18:0), S-methylcysteine sulfoxide, 4-methylcatechol sulfate, 1-oleoyl-GPG (18:1), glycodeoxycholate 3-sulfate, 1-linoleoyl-GPG (18:2), 1-oleoyl-2-linoleoyl-GPE (18:1/18:2), 1-linoleoyl-2-linolenoyl-GPC (18:2/18:3), octadecenedioate (C18:1-DC), octadecadienedioate (C18:2-DC), N-succinyl-phenylalanine, 1-palmitoyl-2-linoleoyl-GPC (16:0/18:2), X-11,299, X-24,546, X-24,951, the arginine to phosphate ratio, the aspartate to mannose ratio, and the cholesterol to linoleoyl-arachidonoyl-glycerol (18:2 to 20:4) ratio. The integrated results are presented in a forest plot (Fig. 3). Among these metabolites, levels of 1-linoleoylglycerol (18:2), S-methylcysteine sulfoxide, and the aspartate to mannose ratio showed positive associations with IS, indicating them as risk factors. The remaining metabolites were negatively associated with IS, suggesting their roles as protective factors.

Identification of likely causal IS risk factors

We investigated the association between SNPs from 11 risk factors (BMI, LDL, HDL, non-HDL, SBP, DBP, smoking, alcohol, coffee, T2D, and CHD), accounting for linkage disequilibrium to obtain instrumental variables. Five MR methods were used to analyze the causal relationships between these 11 risk factors and IS. The results are depicted in forest plots, funnel plots, scatter plots, and leave-one-out sensitivity analysis forest plots. The analysis suggests a causal relationship between all 11 risk factors and IS. BMI, HDL, SBP, DBP, and T2D showed positive correlations with IS, identifying them as risk factors (Fig. 4). In contrast, LDL, non-HDL, smoking, alcohol, coffee, and CHD demonstrated negative correlations with IS, indicating their roles as protective factors.

Identification of IS risk factors associated metabolites

We conducted MR analysis on all 1,400 plasma metabolites with the eleven identified IS risk factors, calculating OR values, heterogeneity, and pleiotropy. The overall results are visualized with circular cluster heatmaps. Metabolites associated with the eleven risk factors were filtered based on p -values, FDR, OR values, and pleiotropy. After multiple testing correction, 136 metabolites were found to be associated with at least one IS risk factor.

We identified 44 metabolites associated with BMI, with 28 showing positive correlations and 16 showing negative correlations. For LDL-associated metabolites, 24 were identified, with 8 showing positive correlations and 16 showing negative correlations. For HDL, 42 metabolites were associated, with 17 positive correlations and 25 negative correlations. Twenty metabolites were associated with non-HDL, with an equal split of 10 positive and 10 negative correlations. Five metabolites were linked to SBP, with 3 positive correlations and 2 negative correlations. For DBP, 25 metabolites were associated, with 5 positive correlations and 20 negative correlations.

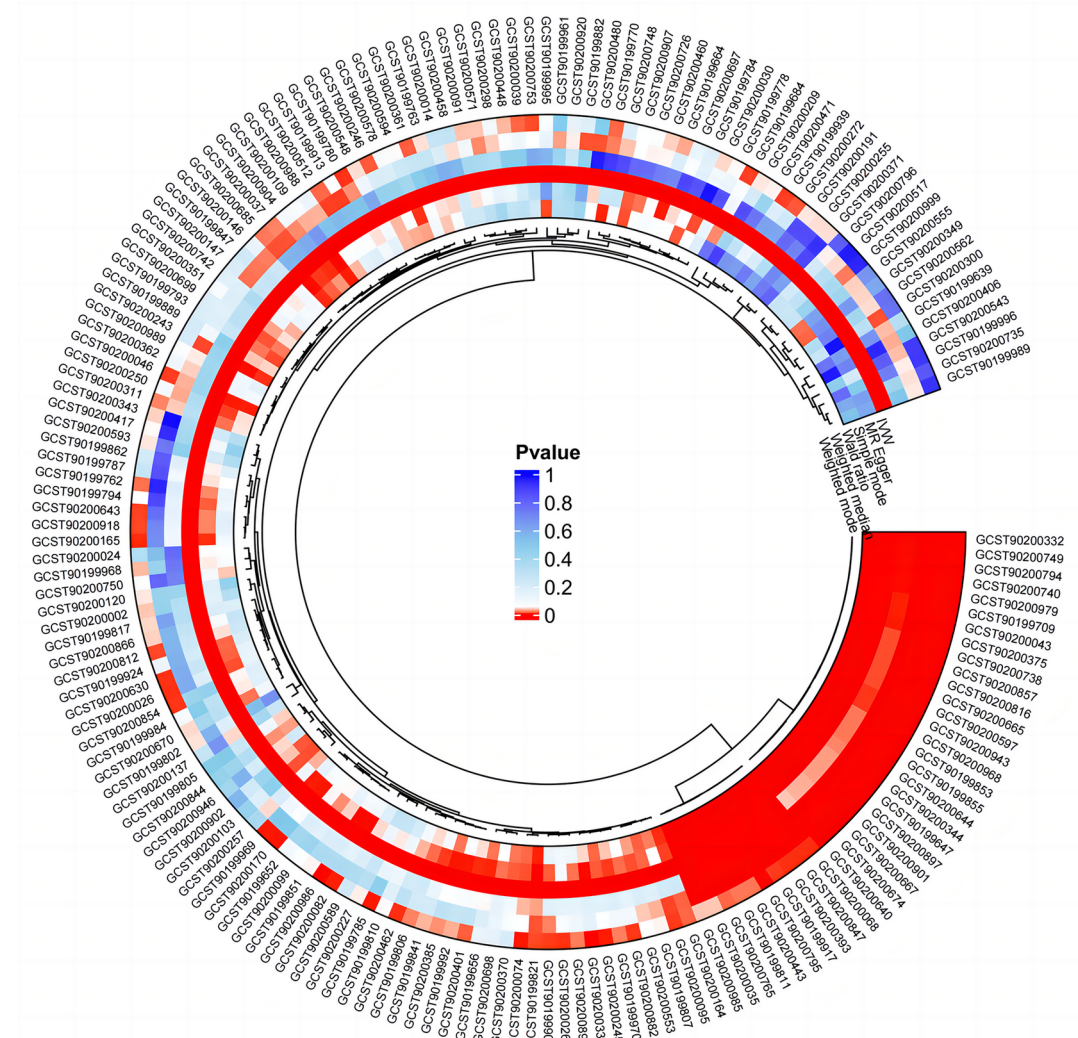


Fig. 2. Circular clustering heatmap of the causal relationship between 1,400 metabolites and IS based on five MR methods. MR Mendelian randomization, IVW inverse variance weighted. GWAS data for 1400 metabolites from the CLSA cohort (Supplementary Table S1). IS GWAS data from the IEU databases (ebi-a-GCST005843).

Smoking was associated with 2 metabolites, both showing positive correlations. Only one metabolite was associated with alcohol, showing a negative correlation. Coffee was linked to 9 metabolites, with 5 positive correlations and 4 negative correlations. For T2D, 10 associated metabolites were found, with 6 positive and 4 negative correlations. Lastly, 5 metabolites were associated with CHD, with 3 positive correlations and 2 negative correlations. There was no evidence of horizontal pleiotropy, and sensitivity analyses provided consistent causal effect estimates. Among the 19 IS-associated metabolites, 4 were identified as being associated with one of the IS risk factors (Fig. 5). Of the 136 metabolites associated with at least one IS risk factor, 132 were associated with the risk factors but not with the IS outcome.

Mediation effect of four metabolites on IS outcomes via risk factors

To investigate the indirect effects of metabolites on IS outcomes via risk factors, we conducted a mediation analysis using effect estimates from two-step MR and the total effect from primary MR. This analysis focused on four metabolites that showed evidence of an effect in both MR analyses on risk factors and IS outcomes: 1-stearoyl-GPG (18:0), 1-oleoyl-2-linoleoyl-GPE (18:1/18:2), octadecadienedioate (C18:2-DC), and X-24,951. Indirect effects were estimated using the product method, with standard errors (SE) and confidence intervals (CI) calculated using the delta method. The mediation effect of 1-stearoyl-GPG (18:0) via DBP accounted for 9.82% (Fig. 6a), while that of 1-oleoyl-2-linoleoyl-GPE (18:1/18:2) via DBP accounted for -12.4% (Fig. 6b). The indirect effect of octadecadienedioate (C18:2-DC) through coffee intake on IS risk accounted for -2.97% of the total effect (Fig. 6c). Similarly, the mediation effect of X-24,951 through BMI on IS was -2.85% (Fig. 6d).

exposure	nsnp	method	pval		OR(95% CI)	heterogeneity_Q_pval	pleiotropy_pval
1-linoleoylglycerol (18:2) levels	13	MR Egger	0.062		1.233 (1.012 to 1.503)	0.85042910	0.2917683
1-linoleoylglycerol (18:2) levels	13	Inverse variance weighted	0.005		1.112 (1.032 to 1.199)	0.81874258	0.2917683
1-stearoyl-GPG (18:0) levels	18	MR Egger	0.022		0.837 (0.729 to 0.960)	0.13268314	0.1706455
1-stearoyl-GPG (18:0) levels	18	Inverse variance weighted	0.004		0.917 (0.865 to 0.973)	0.09013318	0.1706455
S-methylcysteine sulfoxide levels	16	MR Egger	0.019		1.246 (1.059 to 1.465)	0.81381665	0.1109043
S-methylcysteine sulfoxide levels	16	Inverse variance weighted	0.005		1.094 (1.027 to 1.164)	0.66694363	0.1109043
4-methylcatechol sulfate levels	17	MR Egger	0.172		0.844 (0.669 to 1.064)	0.61838502	0.4659682
4-methylcatechol sulfate levels	17	Inverse variance weighted	0.006		0.919 (0.866 to 0.976)	0.64691706	0.4659682
1-oleoyl-GPG (18:1) levels	3	MR Egger	0.507		0.960 (0.885 to 1.042)	0.74835022	0.4655338
1-oleoyl-GPG (18:1) levels	3	Inverse variance weighted	0.001		0.925 (0.882 to 0.969)	0.51036852	0.4655338
Glycodeoxycholate 3-sulfate levels	3	MR Egger	0.595		0.957 (0.851 to 1.076)	0.20921963	0.8544872
Glycodeoxycholate 3-sulfate levels	3	Inverse variance weighted	0.003		0.945 (0.909 to 0.981)	0.43557971	0.8544872
1-linoleoyl-GPG (18:2) levels	14	MR Egger	0.059		0.921 (0.852 to 0.995)	0.21926608	0.7912082
1-linoleoyl-GPG (18:2) levels	14	Inverse variance weighted	0.001		0.929 (0.889 to 0.971)	0.27636033	0.7912082
1-oleoyl-2-linoleoyl-GPE (18:1/18:2) levels	3	MR Egger	0.350		0.896 (0.785 to 1.022)	0.21926608	0.5763587
1-oleoyl-2-linoleoyl-GPE (18:1/18:2) levels	3	Inverse variance weighted	0.002		0.942 (0.908 to 0.978)	0.27636033	0.5763587
1-linoleoyl-2-linoleoyl-GPC (18:2/18:3) levels	17	MR Egger	0.006		0.883 (0.818 to 0.953)	0.41199975	0.5763587
1-linoleoyl-2-linoleoyl-GPC (18:2/18:3) levels	17	Inverse variance weighted	0.005		0.943 (0.905 to 0.982)	0.52490867	0.5763587
Octadecenedioate (C18:1-DC) levels	3	MR Egger	0.882		0.970 (0.868 to 1.083)	0.30699159	0.6213972
Octadecenedioate (C18:1-DC) levels	3	Inverse variance weighted	0.003		0.936 (0.897 to 0.978)	0.46742761	0.6213972
Octadecadienedioate (C18:2-DC) levels	4	MR Egger	0.495		0.969 (0.898 to 1.045)	0.59504798	0.4972319
Octadecadienedioate (C18:2-DC) levels	4	Inverse variance weighted	<0.001		0.942 (0.909 to 0.976)	0.63364859	0.4972319
N-succinyl-phenylalanine levels	23	MR Egger	0.069		0.944 (0.889 to 1.001)	0.78834380	0.6408596
N-succinyl-phenylalanine levels	23	Inverse variance weighted	0.004		0.955 (0.926 to 0.985)	0.82151234	0.6408596
1-palmitoyl-2-linoleoyl-gpc (16:0/18:2) levels	4	MR Egger	0.140		0.874 (0.782 to 0.976)	0.99839642	0.6872128
1-palmitoyl-2-linoleoyl-gpc (16:0/18:2) levels	4	Inverse variance weighted	<0.001		0.894 (0.846 to 0.945)	0.97428086	0.6872128
X-11299 levels	13	MR Egger	0.293		0.924 (0.803 to 1.063)	0.49687927	0.8053095
X-11299 levels	13	Inverse variance weighted	0.003		0.909 (0.853 to 0.969)	0.57740325	0.8053095
X-24546 levels	3	MR Egger	0.590		0.946 (0.817 to 1.094)	0.21192402	0.7668777
X-24546 levels	3	Inverse variance weighted	0.004		0.922 (0.873 to 0.974)	0.40914173	0.7668777
X-24951 levels	3	MR Egger	0.779		0.892 (0.479 to 1.660)	0.18265348	0.7788717
X-24951 levels	3	Inverse variance weighted	0.006		0.800 (0.683 to 0.937)	0.36629090	0.7788717
Arginine to phosphate ratio	17	MR Egger	0.420		1.062 (0.921 to 1.224)	0.52438195	0.8448350
Arginine to phosphate ratio	17	Inverse variance weighted	0.004		1.076 (1.023 to 1.132)	0.59463252	0.8448350
Aspartate to mannose ratio	17	MR Egger	0.092		1.190 (0.985 to 1.438)	0.13654088	0.3867345
Aspartate to mannose ratio	17	Inverse variance weighted	0.004		1.098 (1.030 to 1.170)	0.13928907	0.3867345
Cholesterol to linoleoyl-arachidonoyl-glycerol (18:2 to 20:4) [2] ratio	3	MR Egger	0.343		0.867 (0.733 to 1.025)	0.59483653	0.7356641
Cholesterol to linoleoyl-arachidonoyl-glycerol (18:2 to 20:4) [2] ratio	3	Inverse variance weighted	<0.001		0.898 (0.848 to 0.951)	0.78772914	0.7356641

Fig. 3. Effects of 19 potential causal metabolites on IS outcomes. Primary MR estimates were calculated using the inverse-variance weighted method under multiplicative random effects. Using MR-PRESSO, we identified 19 metabolites that were causally associated with IS.

Phenome-wide MR (Phe-MR) analysis of IS-associated metabolites

To evaluate the effects of the nineteen IS-associated metabolites on other conditions, we conducted an extensive MR screen of 3,948 diseases and traits in the UK Biobank. Metabolites are classified as “deleterious” if their effect direction on a disease aligns with their effect on IS, and “beneficial” if the direction is opposite.

Notably, several metabolites showed significant associations with diseases known to be comorbid or mechanistically linked with IS. For instance, 1-stearoyl-GPG (18:0), a metabolite positively associated with IS, also showed risk associations with coronary atherosclerosis, heart disease, and hypercholesterolemia, indicating shared vascular-metabolic pathways. Similarly, 1-linoleoylglycerol (18:2) was positively associated with heart failure, asthma, and type 2 diabetes—all recognized IS risk factors—supporting its deleterious role in cerebrovascular health. In contrast, Octadecadienedioate (C18:2-DC) exhibited protective effects against hypertension, heart disease, and type 2 diabetes, suggesting its potential as a systemic cardiometabolic modulator. X-24,951 was linked to hypertrophic cardiomyopathy and depression, implying a role in heart-brain axis pathology, while also being inversely associated with hypercholesterolemia and valvular disease. Furthermore, 1-oleoyl-2-linoleoyl-GPE (18:1/18:2) demonstrated risk associations with hypertension and gastric ulcers but protected against respiratory conditions like bronchitis and allergic rhinitis. Finally, S-methylcysteine sulfoxide, although associated with digestive diseases, showed inverse associations with coronary atherosclerosis, asthma, and musculoskeletal conditions. Together, these findings highlight that IS-associated metabolites often influence diseases beyond stroke, particularly within cardiovascular, metabolic, and neuropsychiatric domains, offering insights into comorbidity mechanisms and translational implications.

Discussion

This study is the first large-scale analysis employing GWAS data on 1,400 plasma to systematically identify IS biomarkers. Using GWAS data on 1,091 plasma metabolites and 309 metabolite ratios, we provide robust evidence of nineteen causal metabolites for IS. Elevated levels of 1-linoleoylglycerol (18:2), S-methylcysteine sulfoxide, and the aspartate to mannose ratio were positively associated with IS, indicating risk factors. The remaining sixteen metabolites show negative associations, suggesting they act as protective factors. A prior study reported thirteen causal metabolites that did not overlap with ours, despite also focusing on a European population, which may reflect differences in recruitment, outcome definitions, and overall study design. In addition, we employed a lower significance threshold for instrument selection, enabling the detection of more subtle associations that might have been excluded under stricter criteria. Importantly, all 19 metabolites in

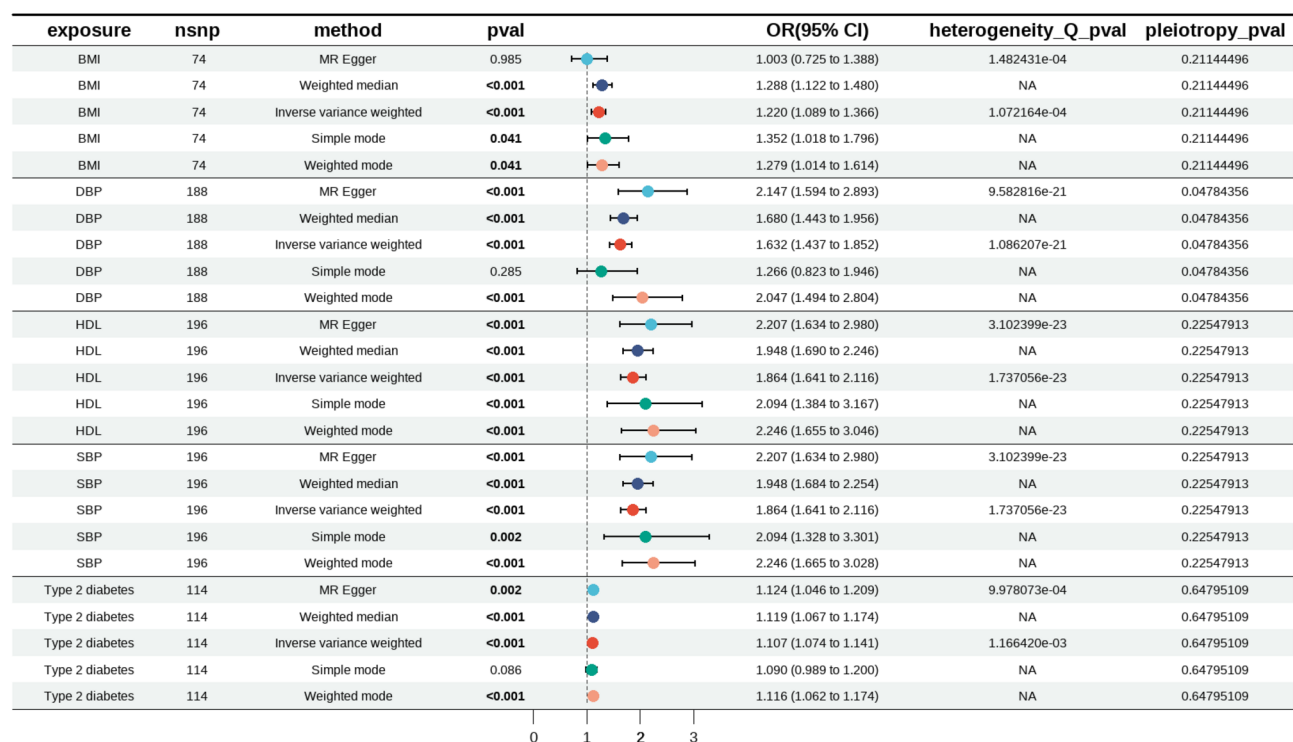


Fig. 4. Causal effects of five risk factors on IS outcomes. MR analyses of the effect of risk factors on stroke outcomes. GWAS data of IS risk factors from the IEU databases. Primary MR estimates were calculated using the inverse-variance weighted method under multiplicative random effects. BMI, HDL, SBP, DBP, and T2D demonstrated positive correlations with IS, identifying them as risk factors. *BMI* body mass index, *HDL* high-density lipoprotein, *SBP* systolic blood pressure, *DBP* diastolic blood pressure, *T2D* type 2 diabetes.

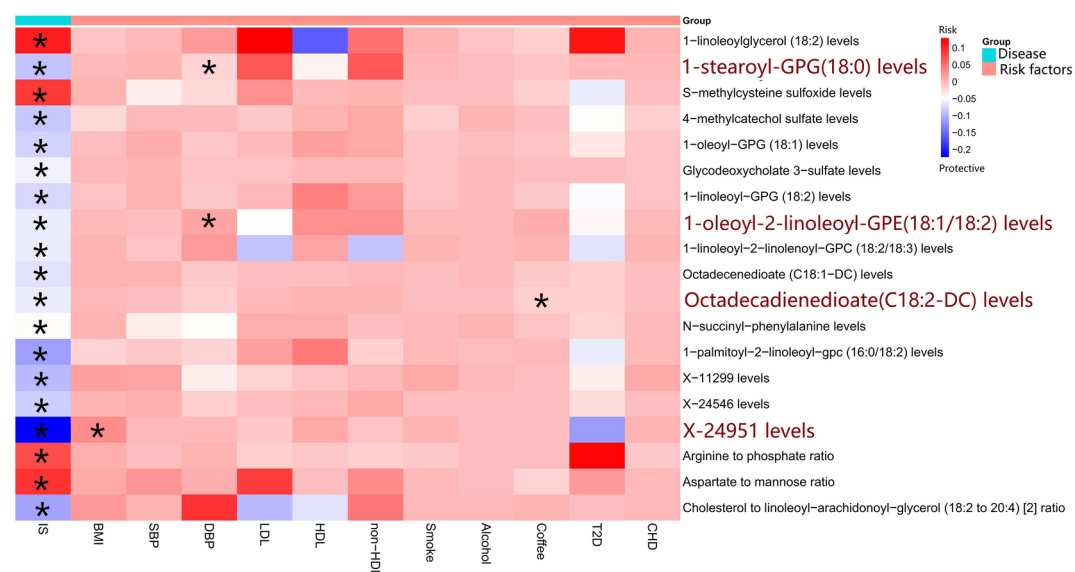


Fig. 5. Among the nineteen IS-associated metabolites, four were identified as being associated with one of the IS risk factors. MR analyses of the effect of metabolites on IS and IS risk factors. The leftmost column shows 19 IS-associated metabolites. X-24,951 was identified as being associated with BMI. 1-stearoyl-GPG (18:0) and 1-oleoyl-2-linoleoyl-GPE (18:1/18:2) were identified as being associated with DBP. Octadecadienedioate (C18:2-DC) was identified as being associated with coffee intake. *Indicates that the causal association is significant, which passed $p < 0.05$, false discovery rate (FDR) < 0.2 , consistent OR direction across five MR methods, and pleiotropy > 0.05 . *BMI* body mass index, *LDL* low-density lipoprotein, *HDL* high-density lipoprotein, *non-HDL* non-high-density lipoprotein, *SBP* systolic blood pressure, *DBP* diastolic blood pressure, *T2D* type 2 diabetes, *CHD* coronary heart disease.

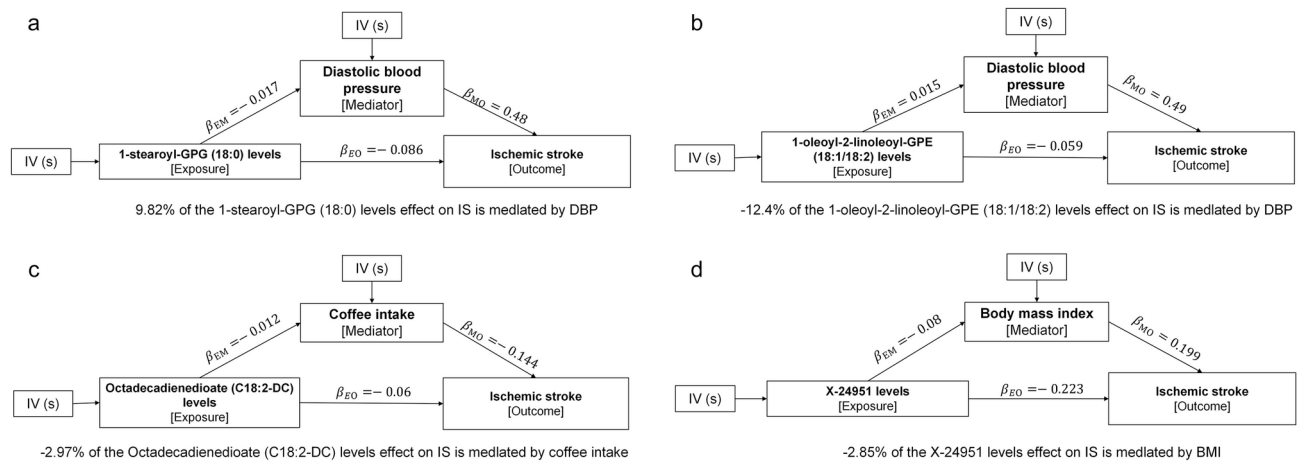


Fig. 6. Mediation effects of metabolites on IS via risk factors. Mediation analyses to quantify the effects of four metabolites on IS outcomes via risk factors. (a) 1-stearoyl-GPG (18:0) levels effect on IS mediated by DBP. (b) 1-oleoyl-2-linoleoyl-GPE (18:1/18:2) effect on IS mediated by DBP. (c) Octadecadienedioate (C18:2-DC) effect on IS mediated by coffee intake. (d) X-24,951 effect on IS mediated by BMI. β_{EM} : effects of exposure on mediator, β_{MO} : effects of mediator on outcome, β_{EO} : effects of exposure on outcome. IVs instrumental variables.

our study met the screening criteria of $p < 0.05$, $FDR < 0.2$, consistent OR direction in five MR methods, and pleiotropy > 0.05 , underscoring the reliability of our findings. We identified eleven risk factors causally related to IS, including BMI, HDL, SBP, DBP, and T2D, indicating their crucial role in IS pathogenesis, consistent with classical epidemiological data^{9,37–39}. Conversely, LDL, non-HDL, smoking, alcohol consumption, coffee intake, and CHD were negatively associated with IS, acting as protective factors, which contrasts with epidemiological data. Our analysis indicates that the associations between 1-stearoyl-GPG (18:0), 1-oleoyl-2-linoleoyl-GPE (18:1/18:2), octadecadienedioate (C18:2-DC), and X-24,951 levels with IS may be mediated by risk factors like BMI, DBP, and coffee intake. Additionally, we identified 132 other metabolites with causal relationships to these risk factors. The Phe-MR analysis highlighted the bidirectional effects of the 19 IS-associated metabolites on other conditions.

BMI, HDL, SBP, DBP, and T2D were causally associated with increased IS risk. Previous studies have shown that increased BMI significantly elevates stroke risk. Reducing BMI through lifestyle changes and dietary control can significantly lower this risk⁴⁰. Changes in HDL composition appear to correlate with the severity and outcomes of acute ischemic stroke, potentially serving as biomarkers for risk stratification and management⁴¹. The China Hypertension Cohort Study revealed that elevated blood pressure (DBP/SBP) is positively associated with stroke incidence, supporting our results⁴². Other studies also support the causal relationship between T2D and IS risk, indicating that genetic susceptibility to T2D and higher HbA1c levels are linked to increased risk of large-artery and small-vessel ischemic strokes⁴³. Conversely, LDL, non-HDL, smoking, alcohol consumption, coffee intake, and CHD were negatively associated with IS. Studies have shown that LDL is negatively associated with IS⁴⁴, consistent with our findings. Most previous studies investigating the relationship between elevated non-HDL cholesterol levels and IS risk have found correlations, suggesting that elevated non-HDL cholesterol might be a better marker for IS risk^{45,46}. In this study, we found for the first time that non-HDL cholesterol is negatively associated with IS risk, a result warranting further investigation. Smoking is a well-established risk factor for stroke occurrence and poor prognosis. The correlation between a family history of stroke and IS risk is more pronounced among smokers but is not observed among those who have quit smoking for over ten years or never smoked⁴⁷. Recent data indicate that active smoking or exposure to environmental tobacco smoke (ETS) is linked to increased risks for all types of strokes, including their major pathological and etiological subtypes⁴⁸. Despite our contrary findings showing that smoking is negatively associated with IS, we still encourage younger individuals to avoid smoking, promote quitting, and support smoke-free environments. Traditional observational epidemiological studies have found that, compared to non-drinkers, moderate alcohol consumers have a slightly reduced risk of IS, aligning with our findings. However, a recent study indicated that increased alcohol consumption consistently raises blood pressure levels and subsequently increases stroke risk⁴⁹. This research, derived from the CKB project, included prospective follow-up and genetic data from 160,000 adults, suggesting that moderate alcohol consumption does not protect against stroke and that even low levels of alcohol intake may increase stroke risk. Large-scale cohort studies have previously found that drinking coffee may be linked to reduced risks of stroke, dementia, and post-stroke dementia⁵⁰. This study showed that compared to non-coffee drinkers, people who consume 2–3 cups of coffee daily have a 32% reduced risk of stroke and a 28% reduced risk of dementia. Coffee contains polyphenols and other bioactive compounds with potentially beneficial health effects, such as neuroprotection, antioxidative stress, anti-inflammation, inhibition of β -amyloid accumulation, and anti-apoptotic properties⁵⁰. Previous research has demonstrated that age-adjusted stroke incidence rates more than doubled among patients with CHD, including acute coronary syndrome (ACS), compared to those without CHD⁵¹. This contrasts with our findings. Our results partially contradict the existing literature. Possible

reasons for identifying LDL and smoking as “protective factors” include the broad phenotype definitions in GWAS, which may overlook details such as smoking intensity or LDL subtypes; a “survivor bias” that excludes high-risk individuals from the final sample; unaddressed heterogeneity among stroke subtypes; and the inability to fully eliminate residual pleiotropy. Moreover, MR captures lifetime genetic exposure, differing from short-term or behavior-based effects. Future research employing larger cohorts, delineated stroke subtypes, and functional validation will be crucial to reconciling these discrepancies and strengthening causal inferences.

Metabolomics represents a highly promising approach for biomarker identification. Unlike other analytical techniques, its advantage lies in providing a comprehensive spectrum of low molecular weight metabolites rather than focusing on a single molecular profile. Alterations in metabolite concentrations during cerebral ischemia can profoundly affect primary neuronal function⁵². MR analysis of 1,400 plasma metabolites with IS revealed that 1-linoleoylglycerol (18:2), S-methylcysteine sulfoxide, and the aspartate-to-mannose ratio are positively correlated with IS, serving as risk factors. In contrast, the following metabolites were negatively associated with IS, acting as protective factors: 1-stearoyl-GPG (18:0), 4-methylcatechol sulfate, 1-oleoyl-GPG (18:1), glycodeoxycholate 3-sulfate, 1-linoleoyl-GPG (18:2), 1-oleoyl-2-linoleoyl-GPE (18:1/18:2), 1-linoleoyl-2-linolenoyl-GPC (18:2/18:3), octadecenedioate (C18:1-DC), octadecadienedioate (C18:2-DC), N-succinyl-phenylalanine, 1-palmitoyl-2-linoleoyl-GPC (16:0/18:2), X-11,299, X-24,546, X-24,951, arginine to phosphate ratio, and cholesterol-to-linoleoyl-arachidonoyl-glycerol (18:2 to 20:4) [2]. Among the 19 metabolites causally associated with IS, we identified two—1-stearoyl-GPG (18:0) and 1-oleoyl-2-linoleoyl-GPE (18:1/18:2)—that are potentially negatively associated with IS risk by modulating DBP. Notably, 1-stearoyl-GPG (18:0) was also identified in a study distinguishing elite and non-elite athletes, indicating genetic tendencies for elite athletic performance⁵³. This metabolite might reduce IS risk by influencing DBP after intense exercise. 1-oleoyl-2-linoleoyl-GPE (18:1/18:2), known as PE (18:1(9Z)/18:2(9Z,12Z)), plays an essential regulatory role in central nervous system diseases like Parkinson’s and Alzheimer’s diseases, with DBP being a known risk factor for these conditions^{54,55}. Additionally, we identified that octadecadienedioate (C18:2-DC) and X-24,951 are potentially negatively associated with IS risk by regulating coffee intake and BMI, respectively. Further research is needed to understand the relationship and mechanisms of these metabolites with IS. It is noteworthy that some analyses revealed a “negative mediation effect,” meaning the metabolite’s influence on the mediator opposes the mediator’s effect on IS. Rather than amplifying risk, this suggests the mediator may buffer or suppress the causal pathway. Such findings imply complex biological processes, possibly involving pleiotropy or compensatory responses. Further research, including functional assays or refined stratification by IS subtype, could clarify these mechanisms and inform targeted interventions. For metabolites identified in the MR analysis as positively associated with IS, we performed a multitrait colocalization analysis using the Hyprcoloc package⁵⁶. Unfortunately, the positive metabolites identified through our MR analysis weren’t replicated through colocalization. Recent research indicates that a lack of colocalization evidence does not invalidate findings, as colocalization methods have a high false-negative rate, often around 60%⁵⁷. Therefore, despite the absence of the same positive metabolites in our multitrait colocalization analysis, it does not undermine the robustness of our study results. However, we acknowledge that our colocalization findings require validation in independent datasets, limiting the robustness of causal interpretations. Current methods (e.g., coloc, Hyprcoloc) assume a single causal variant and rely on accurate LD estimation, potentially introducing biases in regions with complex LD or small samples. Future work should confirm these results in larger cohorts, employ multiple-signal tools (e.g., SuSiE, eCAVIAR), and integrate functional assays to strengthen causal inferences.

The results of the phenome-wide MR analysis underscore the systemic influence of IS-associated metabolites, revealing interconnected pathways that extend beyond the cerebrovascular domain. Several metabolites, including 1-stearoyl-GPG (18:0) and 1-linoleoylglycerol (18:2), demonstrated consistent risk profiles for both IS and other cardiovascular, metabolic, or respiratory conditions, suggesting shared pathophysiological mechanisms. Conversely, metabolites such as octadecadienedioate (C18:2-DC) displayed protective effects against conditions frequently comorbid with IS, including hypertension and type 2 diabetes. These cross-disease associations emphasize the need to view IS not as an isolated entity but as part of a broader cardiometabolic continuum, wherein specific metabolites could serve as key mediators. In addition, the linkage of certain metabolites, like X-24,951, with both cardiovascular and neuropsychiatric conditions signals a heart-brain interplay that may inform future research on integrative preventive strategies. Protective or deleterious roles observed across multiple disease phenotypes highlight potential therapeutic targets, enabling a more personalized approach to intervention. Overall, this Phe-MR analysis advances our understanding of how IS-related metabolites influence a wide spectrum of disorders, providing valuable insights into shared molecular pathways and offering avenues for novel diagnostic, prognostic, and therapeutic developments.

Conclusion

In conclusion, we identified four human plasma metabolites with causal relationships to both IS and its risk factors. Our analysis revealed their mediating effects on IS through one or more risk factors. Using Phe-MR analysis of 3,948 phenotypes associated with the target metabolites, we found that these four metabolites influenced other conditions in the same direction as IS, all of which were protective. These findings provide new insights into screening, prevention, and treatment of IS using these metabolites as biomarkers. However, it is important to acknowledge the current lack of sufficient evidence supporting the associations between these metabolites and IS risk. We advocate for rigorously designed, large-sample, prospective cohort studies to confirm whether these metabolites can serve as definitive biomarkers for IS. Future research should focus on elucidating the pathways and biological processes involving these biomarkers to better understand the specific mechanisms by which they mediate IS.

Data availability

The datasets used in this study are publicly available summary datasets and can be found in cited papers, in the IEU OpenGWAS Project repository [<https://gwas.mrcieu.ac.uk/>], in the GWAS Catalogue repository [http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/], or in the UK Biobank website [<https://biobank.ndph.ox.ac.uk/showcase/search.cgi>]. There are no restrictions on data availability other than those imposed by the corresponding data committee.

Received: 23 August 2024; Accepted: 5 May 2025

Published online: 14 May 2025

References

1. GBD 2021 Diseases and Injuries Collaborators. Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy life expectancy (HALE) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990–2021: a systematic analysis for the global burden of disease study 2021. *Lancet* **403**, 2133–2161 (2024).
2. Saini, V., Guada, L. & Yavagal, D. R. Global Epidemiology of Stroke and Access to Acute Ischemic Stroke Interventions. *Neurology* **97**, S6–S16 (2021).
3. Roth, G. A. et al. Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study. *J. Am. Coll. Cardiol.* **76**, 2982–3021 (2020).
4. Gerstl, J. V. E. et al. Global, regional, and National economic consequences of stroke. *Stroke* **54**, 2380–2389 (2023).
5. Ding, Q. et al. Global, regional, and National burden of ischemic stroke, 1990–2019. *Neurology* **98**, e279–e290 (2022).
6. Wang, Y. J. et al. China Stroke Report 2020 (Chinese version) (1). *Chin. Stroke J.* **17**, 433–447. https://kns.cnki.net/kcms2/article/abstract?v=m2RMPZxbF1LgbYp4rLQmSx7EGy1MWJn18acAjjfXSLOuNJA9DfC8wiYnncUX0v-EF3K2An0nLze89rv7YgdzRPt1UajdGggvIAzKnBeZYLXn6iObKKmxE_4bo-hoTPs1dhlRoq9N2I=&uniplatform=NZKPT&language=CHS (2022).
7. Jia, J. et al. Application of metabolomics to the discovery of biomarkers for ischemic stroke in the murine model: a comparison with the clinical results. *Mol. Neurobiol.* **58**, 6415–6426 (2021).
8. Chen, Y. et al. Mortality and recurrent vascular events after first incident stroke: a 9-year community-based study of 0.5 million Chinese adults. *Lancet Glob Health.* **8**, e580–e590 (2020).
9. Kleindorfer, D. O. et al. 2021 Guideline for the prevention of stroke in patients with stroke and transient ischemic attack: A guideline from the American heart association/American stroke association. *Stroke* **52**, e364–e467 (2021).
10. GBD 2019 Stroke Collaborators. Global, regional, and National burden of stroke and its risk factors, 1990–2019: a systematic analysis for the global burden of disease study 2019. *Lancet Neurol.* **20**, 795–820 (2021).
11. Tu, W. J., Wang, L. D. & Special Writing Group of China Stroke Surveillance Report. China stroke surveillance report 2021. *Mil Med. Res.* **10**, 33 (2023).
12. Report on stroke prevention and treatment in China Writing Group. Summary of the China stroke prevention and control report 2019. *Chin. J. Cerebrovasc. Dis.* **17**, 272–281 (2020).
13. Boehme, A. K., Esenwa, C. & Elkind, M. S. Stroke risk factors, genetics, and prevention. *Circ. Res.* **120**, 472–495 (2017).
14. Xia, X. et al. Prevalence and risk factors of stroke in the elderly in Northern China: data from the National stroke screening survey. *J. Neurol.* **266**, 1449–1458 (2019).
15. Xie, X. et al. Assessing the applicability of 2017 ACC/AHA hypertension guidelines for secondary stroke prevention in the BOSS study. *J. Clin. Hypertens.* **21**, 1534–1541 (2019).
16. Dai, L. et al. Different contribution of SBP and DBP variability to vascular events in patients with stroke. *Stroke Vasc Neurol.* **5**, 110–115 (2020).
17. Wang, W. et al. Prevalence, incidence, and mortality of stroke in China: results from a nationwide Population-Based survey of 480,687 adults. *Circulation* **135**, 759–771 (2017).
18. Bragg, F. et al. Association between diabetes and Cause-Specific mortality in rural and urban areas of China. *JAMA* **317**, 280–289 (2017).
19. Wang, A. et al. Visit-to-Visit variability of lipids measurements and the risk of stroke and stroke types: A prospective cohort study. *J. Stroke.* **22**, 119–129 (2020).
20. Gu, X. et al. Association of lipids with ischemic and hemorrhagic stroke: A prospective cohort study among 267,500 Chinese. *Stroke* **50**, 3376–3384 (2019).
21. Wolf, P. A., Abbott, R. D. & Kannel, W. B. Atrial fibrillation as an independent risk factor for stroke: the Framingham study. *Stroke* **22**, 983–988 (1991).
22. Epstein, K. A. et al. Smoking cessation and outcome after ischemic stroke or TIA. *Neurology* **89**, 1723–1729 (2017).
23. Yang, L. et al. Alcohol drinking and overall and cause-specific mortality in China: nationally representative prospective study of 220,000 men with 15 years of follow-up. *Int. J. Epidemiol.* **41**, 1101–1113 (2012).
24. Liu, S. et al. Association of general and abdominal obesity and their changes with stroke in Chinese adults: results from an 11.8-year follow-up study. *Nutr. Metab. Cardiovasc. Dis.* **30**, 2001–2007 (2020).
25. Zhou, Y. et al. Increased stroke risk in metabolically abnormal normal weight: a 10-Year Follow-up of 102,037 participants in China. *Transl Stroke Res.* **12**, 725–734 (2021).
26. Su, Z. J. et al. Progress of metabolomic pathogenesis and Pharmacological intervention in ischaemic stroke. *J. Stroke Neurol. Dis.* **40**, 568–572 (2023).
27. Malik, R. et al. Multi-ancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* **50**, 524–537 (2018).
28. Verbanck, M., Chen, C. Y., Neale, B. & Do, R. Detection of widespread horizontal Pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **50**, 693–698 (2018).
29. Chen, Y. et al. Genomic atlas of the plasma metabolome prioritizes metabolites implicated in human diseases. *Nat. Genet.* **55**, 44–53 (2023).
30. He, M. et al. Causal relationship between human blood metabolites and risk of ischemic stroke: a Mendelian randomization study. *Front. Genet.* **15**, 1333454 (2024).
31. Denny, J. C. et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat. Biotechnol.* **31**, 1102–1110 (2013).
32. Skrivankova, V. W. et al. Strengthening the reporting of observational studies in epidemiology using Mendelian randomization: the STROBE-MR statement. *JAMA* **326**, 1614–1621 (2021).
33. Skrivankova, V. W. et al. Strengthening the reporting of observational studies in epidemiology using Mendelian randomisation (STROBE-MR): explanation and elaboration. *BMJ* **375**, n2233 (2021).
34. Carter, A. R. et al. Mendelian randomisation for mediation analysis: current methods and challenges for implementation. *Eur. J. Epidemiol.* **36**, 465–478 (2021).
35. Hao, X. et al. Association between circulating immune cells and the risk of prostate cancer: a Mendelian randomization study. *Front. Endocrinol. (Lausanne)* **15**, 1358416. (2024).

36. Whipp, A. M. et al. Protein associations and protein-metabolite interactions with depressive symptoms and the p-factor. *Transl Psychiatry*. **15**, 128 (2025).
37. Lawes, C. M., Bennett, D. A., Feigin, V. L. & Rodgers, A. Blood pressure and stroke: an overview of published reviews. *Stroke* **35**, 1024 (2004).
38. Mitchell, A. B. et al. Obesity increases risk of ischemic stroke in young adults. *Stroke* **46**, 1690–1692 (2015).
39. Kivimäki, M. et al. Overweight, obesity, and risk of cardiometabolic Multimorbidity: pooled analysis of individual-level data for 120,813 adults from 16 cohort studies from the USA and Europe. *Lancet Public Health*. **2**, e277–e285 (2017).
40. Shao, Y. et al. Link between triglyceride-glucose-body mass index and future stroke risk in middle-aged and elderly Chinese: a nationwide prospective cohort study. *Cardiovasc. Diabetol.* **23**, 81 (2024).
41. Papagiannis, A. et al. HDL cholesterol efflux capacity and phospholipid content are associated with the severity of acute ischemic stroke and predict its outcome. *Clin. Chim. Acta*. **540**, 117229 (2023).
42. Wu, S. et al. Blood pressure classification of 2017 associated with cardiovascular disease and mortality in young Chinese adults. *Hypertension* **76**, 251–258 (2020).
43. Georgakis, M. K. et al. Diabetes mellitus, glycemic traits, and cerebrovascular disease: A Mendelian randomization study. *Neurology* **96**, e1732–e1742 (2021).
44. Holmes, M. V. et al. Lipids, lipoproteins, and metabolites and risk of myocardial infarction and stroke. *J. Am. Coll. Cardiol.* **71**, 620–632 (2018).
45. Emerging Risk Factors Collaboration. Major lipids, apolipoproteins, and risk of vascular disease. *JAMA* **302**, 1993–2000 (2009).
46. O'Donnell, M. J. et al. Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. *Lancet* **376**, 112–123 (2010).
47. Fan, M. et al. Family history, tobacco smoking, and risk of ischemic stroke. *J. Stroke*. **21**, 175–183 (2019).
48. Wang, X. et al. Tobacco use and risk of acute stroke in 32 countries in the INTERSTROKE study: a case-control study. *EClinicalMedicine* **70**, 102515 (2024).
49. Millwood, I. Y. et al. Conventional and genetic evidence on alcohol and vascular disease aetiology: a prospective study of 500,000 men and women in China. *Lancet* **393**, 1831–1842 (2019).
50. Zhang, Y. et al. Consumption of coffee and tea and risk of developing stroke, dementia, and poststroke dementia: A cohort study in the UK biobank. *PLoS Med.* **18**, e1003830 (2021).
51. Chinese Medical Association Geriatrics Branch Neurology Group, Beijing Neurology Association Vascular Neurology Committee, Arteriosclerosis Ischemic Stroke/Transient Ischemic Attack Combined with Coronary Heart Disease Diagnosis and Treatment China Expert Consensus Writing Group. Expert consensus on the diagnosis and treatment of atherosclerotic ischemic stroke/transient ischemic attack combined with coronary heart disease in China (2022). *Chin. Med. J.* **102**, 3569–3580 (2022).
52. Chumachenko, M. S., Waseem, T. V. & Fedorovich, S. V. Metabolomics and metabolites in ischemic stroke. *Rev. Neurosci.* **33**, 181–205 (2021).
53. Al-Khelaifi, F. et al. Metabolic GWAS of elite athletes reveals novel genetically-influenced metabolites associated with athletic performance. *Sci. Rep.* **9**, 19889 (2019).
54. Chang, W. et al. The relationship between phospholipids and insulin resistance: from clinical to experimental studies. *J. Cell. Mol. Med.* **23**, 702–710 (2019).
55. Calzada, E., Onguka, O. & Claypool, S. M. Phosphatidylethanolamine metabolism in health and disease. *Int. Rev. Cell. Mol. Biol.* **321**, 29–88 (2016).
56. Foley, C. N. et al. A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *Nat. Commun.* **12**, 764 (2021).
57. Hukku, A. et al. Probabilistic colocalization of genetic variants from complex and molecular traits: promise and limitations. *Am. J. Hum. Genet.* **108**, 25–35 (2021).

Acknowledgements

We gratefully acknowledge the participants and investigators of all GWAS from which we used summary statistics data.

Author contributions

All authors read and approved the final manuscript and author contributions statement using CRediT with the degree of contribution: Jian Lyu: Conceptualization, Funding acquisition, Data curation, Investigation, Formal analysis, Methodology, Project administration, Software, Visualization, Writing – original draft. Fumei Liu: Data curation, Formal analysis, Methodology, Software, Visualization. Yan Chai: Data curation, Formal analysis, Methodology, Software, Visualization. Xiting Wang: Methodology, Project administration, Resources, Supervision, Validation, Visualization. Yi Liu: Data curation, Formal analysis, Methodology, Resources, Software, Visualization, Writing – review & editing. Yanming Xie: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – review & editing. All data were generated in-house, and no paper mill was used. All authors agree to be accountable for all aspects of work ensuring integrity and accuracy.

Funding

This work was supported by grants from the National Natural Science Foundation of China (grant number 82205101), Fundamental Research Funds for the Central public welfare research institutes (ZZ15-YQ-016), Special Fund for Health Development Research of the Capital (2024-4-4176), National Key Research and Development Program of China (2018YFC1707400), Special Research Fund of the National Research Center for Traditional Chinese Medicine in Cardiovascular Diseases (CMC2022012), and Capacity Enhancement Project at Xiyuan Hospital of China Academy of Chinese Medical Sciences (XYZX0404-11).

Declarations

Competing interests

The authors declare no competing interests.

Ethics statement

We used summary-level data from publicly available GWAS studies which have received ethical approval from their respective institutional review boards and informed consent from all participants. Additionally, since this research does not involve any personal data, it is not necessary to apply for new ethical review approval. No administrative permissions were required to access the data.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-01329-z>.

Correspondence and requests for materials should be addressed to J.L., X.W., Y.L. or Y.X.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025