Cardiometabolic Biomarkers in Blood and the Risk of Cerebrovascular Disease

Nuffield Department of Population Health

Yu-Jie Chiou, MD, MSc
St Cross College, University of Oxford

A thesis submitted for the degree of Doctor of Philosophy
Trinity Term 2025



Abstract

Background: Studies have reported inconsistent associations, with little causal evidence, between metabolites and ischaemic stroke. The aim of this DPhil project is to understand the molecular mechanisms underlying ischaemic stroke and to identify predictive and causal biomarkers based on the sociodemographic, lifestyle, clinical, molecular, and genetic data in UK Biobank (UKB).

Methods: I studied 249 plasma metabolites using the nuclear magnetic resonance platform from UKB (N = 274,355, median 13.8 follow-up years). First, I studied the association of each of these metabolites individually with ischaemic stroke using Cox proportional hazards regression adjusting for age, sex, ethnicity, lipid modifying medication, fasting time, spectrometers, assessment centres, smoking status, alcohol intake, dietary supplement use, 13 dietary variables, physical activity, body mass index, prevalent diabetes, and systolic blood pressure. Second, I applied extreme gradient boosting model (XGBoost) and random survival forests (RSF) to assess the additional predictive values of metabolites over traditional risk factors. The Boruta algorithm was used to procure the relevant metabolites related to ischaemic stroke. Finally, bidirectional two-sample Mendelian randomisation (MR) was used to evaluate the causal relationship of identified metabolites with ischaemic stroke. Genome-wide association studies were conducted to determine genetic instruments for selected metabolites in white British ancestry of UKB (N = 230,197, more than twice the participants of any previous study), and the largest summary-level data for ischaemic stroke were from the GIGASTROKE Consortium in populations of European descent (62,100 cases and 1,234,808 controls). The false discovery rate correction was used to adjust for multiple testing.

Results: After adjusting for multiple testing and potential confounders, 40 metabolites were associated with incident ischaemic stroke in Cox regression, including albumin, phenylalanine, very low-density lipoproteins (VLDLs), low-density lipoproteins, and high-density lipoproteins (e.g. albumin: hazard ratios 0.87 per 1 standard deviation increment, 95% CI 0.84-0.91; phenylalanine: 1.05, 1.02-1.10). The metabolites did not materially add predictive information over conventional risk factors for incident ischaemic stroke by XGBoost and RSF in the whole population, older adults, men, and women separately. Eight relevant metabolites with ischaemic stroke were selected using the Boruta algorithm, and albumin overlapped with the 40 metabolites from the previous analysis. Across the 47 metabolites, a total of 6,137 independent genome-wide significant associations were detected, including 3,025 novel variants (49.3%). The genetically predicted level of three VLDL ratios exhibited consistent direction of effect across univariable MR methods and cholesteryl esters to total lipids ratio in chylomicrons and extremely large VLDL further demonstrated a statistically significant direct effect on ischaemic stroke in the multivariable MR. Additionally, genetic liability to ischaemic stroke had little effect on the metabolites.

Conclusion: These findings substantiate the importance of metabolites as potential modifiable risk factors for ischaemic stroke. Compared with the conventional risk factors, the metabolites did not improve the risk prediction of ischaemic stroke. The causal relationships between three VLDL ratios and ischaemic stroke indicated that indirect measurement of lipoprotein concentrations may be of greater importance than their direct measurement. This thesis makes several original contributions. First, it is the first to combine large-scale prospective analyses, high-dimensional metabolomic profiling, machine learning, and advanced causal inference methods (including the recently developed multivariable MR framework) in a single, unified investigation of

ischaemic stroke. Second, it introduces an expanded GWAS with more than double the previous sample size to generate stronger genetic instruments, thereby enabling more robust Mendelian randomisation analyses of causal effects. Third, it establishes novel genetic associations for key metabolites and provides new insights into their biological architecture and potential pathways underlying ischaemic stroke risk. Future studies should focus on causality and prediction using a wider range of metabolites on ischaemic stroke subtypes in more diverse populations.

Acknowledgements

I wholeheartedly thank Professor Maria Quigley and Associate Professor Mark Sheehan for their uninterrupted support and advice. I must express my heartfelt gratitude to my parents, Jin-feng Shie and Fu-Chiuan Chiou, my wife, Chuan-yi, my daughter, Shiu-yun, St Cross College advisors Dr Nick Jones, Dr Luke Treadwell, and Sharon Durno, and University advisors Maria Kouroumali and Dr James Cones for their backing and warmth.

I would like to thank my supervisors, Professor Sarah Lewington, Associate Professor Benjamin Lacey, Dr Eirini Trichia, and Dr Neil Wright for their guidance and support during my DPhil. Their feedback, help, and encouragement have been invaluable. I could not have wished for better supervisors.

I would like to express my great appreciation to Emeritus Professor Robert Clarke, Associate Professor Iona Millwood, Dr Lei Clifton, and Nia Roberts for their useful questions and advice at several points during the project.

I extend my thanks to Dr Jun Liu, Dr Brittany Ulm, Dr Austin Argentieri, Sihao Xiao, Peien Zhou, Yaqing Gao, Dr Jiang Chang, Dr Shahzad Ahmad, Dr Siyu Chen, Dr Danyao Jin, Dr Lazaros Belbasis, Dr Alankar Atreya, Dr Pallavi Kaushik, Sahba Seddighi, David Shaw, Valirie Ndip, Dr Wenyu Liu, Bowen Liu, Tianshu Liu, Yunhe Wang, Hanyu Wang, Dr Jason Torres, Dr James Liu, Dr Julie Burrett, Dr Andrew Browne, Dr Shing Chan, Dr Peter KaHung Chan, and Dr Pek Kei Im for companionship in Big Data Institute. I have found the experience enjoyable and rewarding. And thank you to my friends, Chia-Yi Lin, Pin-Yuan Lee, Yi-Jhih Huang, Bryan Ng, Ta-Ying Cheng, Shuo-Fang Liang, Wei-Hsuan Lan, Yi-Chuan Chiu, Hong-Po Hsieh, Chiao-Li Ou, Sheng Yang, Chia-Yu Liu, Pai-Jui Yeh, and Jon Chang, in the committee of Oxford

University Taiwanese Student Society, for all of your cooperation on holding fresher's events.

My sincere appreciation also extends to the research training allowance from my department (Oxford Population Health [Nuffield Department of Population Health]), the participants of UK Biobank and the supporting staff for their contribution to the resource.

Finally I would like to thank Kaohsiung Chang Gung Memorial Hospital for providing the grant to enable me to undertake this DPhil.

Statement of contribution to work

I am responsible for conceiving and developing the research questions addressed in this thesis. I performed all aspects of the literature reviews and developed the analysis plans. I conducted all statistical analyses with advice from my supervisors Professor Sarah Lewington, Associate Professor Benjamin Lacey, Dr Eirini Trichia, and Dr Neil Wright, and produced all the tables and figures presented. To complete some of the analyses and figures in Chapters 3, 4 and 6, I revised R scripts available on the UK Biobank Research Analysis Platform, and those developed by colleagues at Oxford Population Health, namely Dr Eirini Trichia, Dr Jun Liu, Dr Shahzad Ahmad, Dr Diego Aguilar Ramirez, and Tin Oreskovic. I interpreted the findings and wrote-up all aspects of the study. This thesis used the UK Biobank resource (application number 61054), which includes health outcome data provided by patients and collected by the NHS as part of their care and support.

Table of contents

Abstract ... 2
Acknowledgements ... 5
Statement of contribution to work ... 7
Table of contents ... 8
List of tables ... 16
List of figures ... 16
Table of abbreviations ... 17
Chapter 1. Introduction ... 19
1.1 Epidemiology of stroke ... 20
1.2 Risk factors for stroke ... 21
1.3 Risk prediction ... 22
1.4 Potential novel risk factors and predictors ... 24
1.5 Integration of metabolomics and genetic data ... 26
1.6 Thesis objectives ... 27
Chapter 2. Literature review ... 30
2.1 Literature review for the first objective of the thesis ... 31
2.1.1 Review methodology ... 31
2.1.2 Results ... 32
2.1.2.1 Biochemical and immunoassay measures ... 32
2.1.2.2 Mass spectrometry ... 40
2.1.2.3 Nuclear magnetic resonance ... 41
2.1.3 Summary and gaps of the first objective: Associations of metabolites with ischaemic stroke risk ... 44
2.2 Literature review for the second objective of the thesis ... 45
2.2.1 Review methodology ... 45
2.2.2 Results ... 46
2.2.3 Summary and gaps of the second objective ... 50
2.3 Literature review for the third objective of the thesis ... 50
2.3.1 Review methodology ... 51
2.3.2 Results ... 51
2.3.3 Summary and gaps of the third objective ... 55
Chapter 3. Data and Methods ... 57
3.1 Study design and participants ... 58
3.2 Measurement of metabolites using the Nightingale platform ... 60

3.3 Genomic data processing 61
3.4 Definitions of outcome 62
3.5 Definitions of covariates and baseline data 62
3.6 Imputation of missing values 68
3.7 Description of NMR metabolites by key socio-demographic and lifestyle characteristics 68
3.7.1 Background 68
3.7.2 Methods 69
3.7.3 Comparison of metabolic measurements between NMR and clinical chemistry 70
3.7.4 Metabolic biomarkers at baseline 72
3.7.4.1 Cross-correlations 72
3.7.4.2 Mean NMR biomarkers by age and sex 78
3.7.4.3 Possible confounders in the association of ischaemic stroke 80
3.7.4.4 Metabolic biomarkers and selected baseline characteristics as possible confounders 82
3.7.4.4.1 Ethnicity 82
3.7.4.4.2 Smoking Status 89
3.7.4.4.3 Alcohol intake 89
3.7.4.4.4 Other baseline characteristics 90
3.7.5 Discussion 91
3.7.5.1 Comparison of correlations with chemistry biomarkers 91
3.7.5.2 Cross-correlations of metabolic biomarkers at baseline 92
3.7.5.3 Mean NMR biomarkers by age and sex 93
3.7.5.4 Metabolic biomarkers and selected baseline characteristics as possible confounders 94
3.7.6 Conclusion 95
Chapter 4. Epidemiological associations of individual metabolites with incident ischaemic stroke 96
4.1 Background 97
4.2 Methods 97
4.2.1 Summary of baseline characteristics 98
4.2.2 Associations of NMR metabolites with incident ischaemic stroke 100
4.2.3 Secondary analyses 103
4.2.4 Comparison of the UK Biobank findings with other studies 104
4.3 Results 105
4.3.1 Baseline characteristics 105
4.3.2 Associations of the NMR metabolites with incident ischaemic stroke 111
4.3.3 Secondary analyses 117

4.3.4 Comparison of the UK Biobank findings with other studies....117
4.4 Discussion....121
4.4.1 Association of the metabolites with incident ischaemic stroke....121
4.4.2 Comparison with other studies....125
4.4.3 Strengths and limitations....126
4.5 Conclusion....127
Chapter 5. Predictive value of NMR metabolites for ischaemic stroke....128
5.1 Background....129
5.2 Methods....131
5.2.1 Variable selection and model building....132
5.2.2 Extreme gradient boosting model....134
5.2.3 Boruta algorithm....135
5.2.4 Random survival forests....136
5.2.5 Validation methods....136
5.2.6 The Shapley additive explanation....138
5.3 Results....138
5.3.1 Predictive value of metabolites for ischaemic stroke risk in the overall population....138
5.3.2 Predictive value of metabolites for ischaemic stroke risk among those aged ≥ 50 years....150
5.3.3 Predictive value of metabolites for ischaemic stroke risk by sex....153
5.4 Discussion....156
5.5 Conclusion....159
Chapter 6. Assessment of the causality of the associations between metabolites and ischaemic stroke: A Mendelian randomisation study....161
6.1 Background....163
6.2 Methods....165
6.2.1 Genome-wide association study of metabolites....166
6.2.2 Mendelian randomisation....172
6.2.3 Multivariable Mendelian randomisation....176
6.2.4 Secondary analyses....177
6.2.5 Framework for determining causality between metabolites and ischaemic stroke....177
6.3 Results....178
6.3.1 Genome-wide association study in UK Biobank....178
6.3.2 Genetic instruments for metabolites in UK Biobank....183
6.3.3 Mendelian randomisation....186
6.3.4 Multivariable Mendelian randomisation....191

6.3.5 Secondary analyses ... 192
6.4 Discussion ... 192
6.5 Conclusion ... 198
Chapter 7. Discussion ... 200
7.1 Main findings ... 201
7.1.1 Description of NMR metabolites by key socio-demographic and lifestyle characteristics ... 202
7.1.2 Epidemiological associations of individual metabolites with incident ischaemic stroke ... 202
7.1.3 Predictive value of NMR metabolites for ischaemic stroke ... 203
7.1.4 Causal relationships between the metabolites and ischaemic stroke in Mendelian randomisation analysis ... 203
7.2 Strengths and limitations ... 204
7.3 Implications for population health ... 206
7.4 Future research ... 208
7.5 Conclusion ... 209
Reference ... 210
Appendices ... 227
Supplementary Table 3.1 Description of 249 metabolites ... 227
Supplementary Table 3.2 Variables for imputation of missing data ... 236
Supplementary Table 3.3 The original dataset and the imputed dataset ... 237
Supplementary Table 4.1 Association of individual metabolites with incident ischaemic stroke in Cox proportional hazards regression in model 1 and 2 (n = 184,564) ... 242
Supplementary Table 4.2 Association of individual metabolites with incident ischaemic stroke in Cox proportional hazards regression in model 3 and 4 (n = 184,564) ... 251
Supplementary Table 4.3 Association of individual metabolites with incident ischaemic stroke in Cox proportional hazards regression in model 5 and 6 (n = 184,564) ... 257
Supplementary Table 4.4 Association of individual metabolites with incident ischaemic stroke in Cox proportional hazards regression in model 7 (n = 184,564) ... 263
Supplementary Table 4.5 Association of individual metabolites with incident ischaemic stroke in the lipid modifying agent nonusers and users in model 7 ... 269
Supplementary Table 4.6 Association of individual metabolites with incident ischaemic stroke in competing risks regression in model 7 (n = 184,564) ... 277
Supplementary Table 4.7 Association of individual metabolites with incident ischaemic stroke after exclusion of the first 3 years of follow-up in model 7 (n = 183,189) ... 286
Supplementary Table 4.8 Association of individual metabolites with incident ischaemic stroke in the White ethnic group and the others in model 7 ... 297
Supplementary Table 4.9 Association of individual metabolites with incident ischaemic stroke in model 7 with additional adjustment for blood pressure medication ... 305

Supplementary Table 4.10 Association of individual metabolites with incident ischaemic stroke in Cox regression in UK Biobank when adjusting for the same confounding factors as other cohorts (n = 184,564)..... 315

Supplementary Table 5.1 Baseline characteristics of participants in the training and testing dataset for machine learning..... 322

Supplementary Table 5.2 Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) checklist..... 325

Supplementary Table 5.3 Details of hyperparameter optimization and model performance in the training dataset in XGBoost model..... 326

Supplementary Table 5.4 Details of hyperparameter optimization and evaluation metrics in the older group (age ≥ 50 years) in XGBoost model..... 327

Supplementary Table 5.5 Details of hyperparameter optimization and evaluation metrics in the male group in XGBoost model..... 328

Supplementary Table 5.6 Details of hyperparameter optimization and evaluation metrics in the female group in XGBoost model..... 329

Supplementary Table 6.1 STROBE-MR checklist of recommended items to address in reports of Mendelian randomization studies..... 330

Supplementary Table 6.2 Genome-wide significant results for 47 metabolites..... 334

Supplementary Table 6.3 Number of genome-wide significant single nucleotide polymorphisms identified for 47 metabolites..... 343

Supplementary Table 6.4 Number of unique genome-wide significant single nucleotide polymorphisms identified for 47 metabolites by chromosome..... 345

Supplementary Table 6.5 Top 20 single nucleotide polymorphisms with the largest number of associated metabolites..... 346

Supplementary Table 6.6 All results from Mendelian randomisation analyses of association of metabolites with ischaemic stroke..... 349

Supplementary Table 6.7 Detailed results from the inverse variance-weighted method and Mendelian randomisation-Egger method of association of metabolites with ischaemic stroke..... 357

Supplementary Table 6.8 Results from Mendelian randomisation pleiotropy residual sum and outlier (MR-PRESSO) global test for association of metabolites with ischaemic stroke..... 361

Supplementary Table 6.9 Results from Mendelian randomisation pleiotropy residual sum and outlier (MR-PRESSO) distortion test for association of metabolites with ischaemic stroke..... 363

Supplementary Table 6.10 Mendelian randomisation pleiotropy residual sum and outlier (MR-PRESSO) outlier-corrected results for association of metabolites with ischaemic stroke..... 367

Supplementary Table 6.11 Results of posterior inclusion probabilities estimated via the iterative Sum of Single Effects (SuSiE) algorithm..... 370

Supplementary Table 6.12 The results from multivariable Mendelian randomisation using inverse-variance weighted estimation..... 372

Supplementary Table 6.13 All results from reverse Mendelian randomisation analyses of association of ischaemic stroke with metabolites..... 373

Supplementary Table 6.14 Reverse Mendelian randomisation pleiotropy residual sum and outlier (MR-PRESSO) outlier-corrected results for association of ischaemic stroke with metabolites..... 378

Supplementary Figure 1.1 Metabolic measures quantified by nuclear magnetic resonance spectroscopy..... 382

Supplementary Figure 2.1 Flowchart of article selection and inclusion in literature review for the first objective of the thesis..... 383

Supplementary Figure 2.2 Flowchart of article selection and inclusion in literature review for the second objective of the thesis..... 384

Supplementary Figure 2.3 Flowchart of article selection and inclusion in literature review for the third objective of the thesis..... 385

Supplementary Figure 3.1 The distribution of NMR metabolites by age and sex (n = 184,564)..... 390

Supplementary Figure 3.2 Associations of each NMR metabolite with selected baseline characteristics..... 414

Supplementary Figure 4.1 Associations of incident ischaemic stroke with quintiles of levels of individual metabolites (n = 184,564)..... 419

Supplementary Figure 4.2 Results of metabolome-wide association analysis..... 421

Supplementary Figure 4.3 Associations of individual metabolites with incident ischaemic stroke in model 7 (n = 184,564)..... 425

Supplementary Figure 4.4 Comparison of associations of metabolic biomarkers with incident ischaemic stroke in the lipid modifying agent nonusers with users..... 426

Supplementary Figure 4.5 Comparison of associations of metabolic biomarkers with risk of incident ischaemic stroke between Cox regression and competing risks regression 428

Supplementary Figure 4.6 Comparison of associations of metabolic biomarkers with incident ischaemic stroke in the overall population with those excluding the first three years of follow-up..... 430

Supplementary Figure 4.7 Comparison of associations of metabolic biomarkers with incident ischaemic stroke in the White ethnic group with the others..... 432

Supplementary Figure 4.8 Comparison of associations of metabolic biomarkers with incident ischaemic stroke in the overall population with the White ethnic group..... 434

Supplementary Figure 4.9 Comparison of associations of metabolic biomarkers with incident ischaemic stroke in the model 7 with additional adjustment for blood pressure medication..... 436

Supplementary Figure 5.1 Predictive values on ischaemic stroke by extreme gradient boosting model and random survival forests..... 439

Supplementary Figure 5.2 SHAP dependence plots of the extreme gradient boosting model with age, sex, and significant metabolites..... 448

Supplementary Figure 5.3 Extreme gradient boosting model calibration..... 449

Supplementary Figure 5.4 The SHAP summary plot of extreme gradient boosting model with risk factors in Framingham Stroke Risk Profile..... 453

Supplementary Figure 5.5 SHAP dependence plots of the extreme gradient boosting model with risk factors in Framingham Stroke Risk Profile..... 456

Supplementary Figure 5.6 SHAP dependence plots and interaction plots of the extreme gradient boosting model with age, sex and 8 metabolites..... 460

Supplementary Figure 6.1 A framework for assessing causal relationships between metabolites and ischaemic stroke..... 461

Supplementary Figure 6.2 Manhattan plots of 47 metabolites and single nucleotide polymorphisms (N = 230,197)..... 485

Supplementary Figure 6.3 Quantile-Quantile (QQ) Plot of p values from Genome-Wide Association Study (GWAS) with genomic control lambda (λ_{GC})..... 486

Supplementary Figure 6.4 Comparison of association of ischaemic stroke with metabolites from observational study and Mendelian randomisation..... 491

Supplementary Figure 6.5 Scatter plots of Mendelian randomisation results..... 517

Supplementary Figure 6.6 Comparison of Mendelian randomisation results from the current thesis (updated) and previous study..... 519

Supplementary Figure 6.7 Reverse Mendelian randomisation of ischaemic stroke with metabolites..... 521

List of tables

Table 2.1 Literature review for the first objective of the thesis: (1) meta-analyses

Table 2.2 Literature review for the second objective of the thesis

Table 2.3 Literature review for the third objective of the thesis

Table 3.1 The definitions of outcome and the covariates in the study

Table 3.2 Association of metabolic biomarkers with selected baseline characteristics

Table 4.1 Baseline characteristics of UKB participants included in the epidemiological analyses

Table 4.2 Overall association results between metabolites and incident ischaemic stroke

Table 4.3 Association of significant individual metabolites with incident ischaemic stroke in model 7 (n = 184,564)

Table 5.1 Prediction of ischaemic stroke and components of each model in the overall population (n = 184,564)

Table 5.2 Prediction of ischaemic stroke and components of each model among those aged ≥ 50 years (n = 137,310)

Table 5.3 Prediction of ischaemic stroke and components of each model in men (n = 83,236)

Table 5.4 Prediction of ischaemic stroke and components of each model in women (n = 101,328)

Table 6.1 Characteristics of participants in genome-wide association analyses in UK Biobank (n = 230,197)

Table 6.2 Genomic control lambda (λ_{GC}), Linkage Disequilibrium Score Regression intercept, and heritability (h^2) estimates for 47 metabolites

Table 6.3 The proportions of variance in metabolites explained by the genetic instruments (R^2) and F-statistics

List of figures

Figure 3.1 A flowchart describing the general framework

Figure 3.2 Comparisons of the NMR measurements with clinical chemistry

Figure 3.3 The correlation matrix of the metabolites in the analyses dataset (n = 184,567)

Figure 4.1 A flowchart describing the general framework of the epidemiological analysis (current Chapter) and the prediction modelling (Chapter 5)

Figure 4.2 Comparison of significant associations of incident ischaemic stroke and metabolites between UK Biobank and China Kadoorie Biobank

Figure 4.3 Comparison of significant associations of incident ischaemic stroke and metabolites between UK Biobank and the consortium of 7 studies on metabolomics and stroke

Figure 5.1 The SHAP summary plot of extreme gradient boosting model with age, sex, and metabolites selected by Boruta algorithm

Figure 5.2 The SHAP summary plot of extreme gradient boosting model with age, sex, and significant metabolites in the training dataset.

Figure 5.3 SHAP dependence plots were shown for the XGBoost-model 2 with age, sex, and metabolites selected by Boruta algorithm

Figure 6.1 A flowchart describing the genetic study.

Table of abbreviations

ApoA1	Apolipoprotein A1
ApoB	Apolipoprotein B
BMI	Body mass index
CI	Confidence interval
CKB	China Kadoorie Biobank
cML	Constrained maximum likelihood
CVD	Cardiovascular disease
DHA	Docosahexaenoic acid
EPIC	European Prospective Investigation into Cancer and Nutrition
FA	Fatty acids
FDR	False discovery rate
FSRP	Framingham Stroke Risk Profile
GWAS	Genome-wide association study
HDL	High-density lipoprotein
HES	Hospital Episode Statistics
HRC	Haplotype Reference Consortium
IDI	Integrated discrimination improvement
IRR	Incidence rate ratio
LA	Linoleic acid
LD	Linkage disequilibrium
LDL	Low-density lipoprotein
LOCO	Leave-one-chromosome-out
MAC	Minor allele count
MAF	Minor allele frequency
MR	Mendelian randomisation
MR-PRESSO	Mendelian randomisation pleiotropy residual sum and outlier
MS	Mass-spectrometry
MUFA	Monounsaturated fatty acids
MVMR	Multivariable Mendelian randomisation
MVMR-cML-SuSiE	Multivariable Mendelian Randomisation with Constrained Maximum Likelihood and Sum of Single Effects
NMR	Nuclear magnetic resonance
OR	Odds ratio
PEDW	Patient Episode Database for Wales
PIPs	posterior inclusion probabilities
PUFA	Polyunsaturated fatty acids
RCT	Randomised controlled trial
SD	Standard deviation
SFA	Saturated fatty acids
SHAP	Shapley additive explanation
SMR	Scottish Morbidity Record

SNP	Single nucleotide polymorphism
SuSiE	Sum of Single Effects
Triglyceride	TG
TRIPOD	Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis
TRIPOD	The transparent reporting of a multivariable prediction model for individual prognosis or diagnosis
UKB	UK Biobank
UVMR-cML-BIC	Univariable Mendelian Randomization using Constrained Maximum Likelihood with Bayesian Information Criterion
UVMR-cML-MA	Univariable Mendelian randomisation constrained maximum likelihood with model averaging
XGBoost	Extreme gradient boosting model
λGC	Genomic inflation factor

Chapter 1. Introduction

Chapter summary

Stroke is a leading cause of death and disability worldwide, with ischaemic stroke accounting for the majority of new cases. Although established risk prediction tools such as the Framingham Stroke Risk Profile (FSRP) and QStroke are widely used, their predictive performance is modest, highlighting the need for new approaches. Metabolomics offers an opportunity to identify novel biomarkers by capturing downstream effects of genetic, proteomic, and environmental influences. Nuclear magnetic resonance (NMR) spectroscopy enables cost-effective, large-scale profiling of lipoproteins, fatty acids, and other metabolites, though its molecular coverage is narrower than that of mass spectrometry.

Combining metabolomics with genetic approaches such as Mendelian randomisation (MR) can help distinguish correlation from causation, providing mechanistic insights into the role of metabolites in ischaemic stroke. The recent expansion of metabolomic data in UK Biobank (N ~ 274,000) offers improved statistical power to examine these relationships compared with earlier studies.

The aim of this thesis is to investigate predictive and causal metabolic biomarkers of ischaemic stroke using a multidisciplinary framework. Specifically, it seeks to identify associations through prospective analyses, evaluate predictive value using machine learning, and assess causality with MR. Together, these objectives form a coherent approach to advancing understanding of the molecular mechanisms and predictive signatures of ischaemic stroke.

1.1 Epidemiology of stroke

Stroke was the third-leading cause of death and disability combined (5.7% of total disability-adjusted life-years) and the second-leading cause of mortality (11.6% of total deaths) globally in 2019.[1] It is defined as an acute episode of focal dysfunction of the brain, spinal cord or retina persisting for over 24 hours, or of any span if autopsy or imaging (computed tomography or magnetic resonance imaging) reveal haemorrhage or focal infarction related to the symptoms.[2] It is categorised into ischaemic stroke if a neurological dysfunction caused by focal cerebral infarction occurs, and haemorrhagic stroke if an episode of bleeding into the brain by the break of a blood vessel occurs.

Three major pathophysiological subtypes of ischaemic stroke include large artery atherosclerotic stroke, cardioembolic stroke, and cerebral small vessel disease.[3] Large artery atherosclerotic stroke occurs when a atherosclerotic plaque develops in a major brain artery or branch cortical artery, leading to luminal narrowing, plaque rupture, and subsequent atherosclerosis that blocks cerebral blood flow.[4] Cardioembolic stroke results from emboli that originate in the heart and travel to the cerebral circulation.[4] The most common sources include atrial fibrillation, valvular heart disease, and impaired ventricular function. Cerebral small vessel disease develops due to structural and functional changes in the small penetrating arterioles, capillaries, and venules, typically measuring less than 1 millimetre in diameter, located deep within the brain.[4,5] These alterations impair microvascular perfusion, leading to chronic hypoxia, tissue injury, and damage to surrounding brain cells.[5]

The clinical symptoms of stroke comprise the sudden onset of a focal clinical deficit, referable to a particular location in the central nervous system.[6] Weakness of the upper or lower limbs on one side of the body, impaired balance, nausea, abrupt severe headache of an unknown cause, speech impairment (aphasia), visual impairment, facial

numbness and weakness are the most typical symptoms of stroke.[6] Both men and women equally report these common symptoms, but light headedness and loss of consciousness are more frequently present in women.[7] Accurate and speedy diagnosis of stroke is pivotal for choosing proper acute stroke treatment, such as endovascular mechanical thrombectomy treatment and intravenous tissue plasminogen activator.[8] The differential diagnoses of stroke involve metabolic disturbances, functional disorders, vestibular disturbances, seizures, and migraine.[6] Furthermore, intracerebral haemorrhage requires differentiation from ischaemic stroke. However, clinical symptoms alone are often insufficient to accurately distinguish between stroke subtypes or differentiate stroke from its mimics, and neuroimaging is the key to diagnosis. Brain imaging commonly contains computed tomography or magnetic resonance imaging.[6]

Of the newly diagnosed patients with stroke in 2019 globally, it has been estimated that 62% were ischaemic and 38% were haemorrhagic strokes.[1,2] The overall worldwide disability-adjusted life-years lost due to ischaemic stroke increased from 40,503,990 (95% confidence interval (CI) 37,572,940-43,726,500) in 1990 to 63,478,270 (95% CI 57,827,410-68,986,520) in 2019.[9]

1.2 Risk factors for stroke

The INTERSTROKE Study, a case-control study based in 32 countries[10,11] indicated that over 90% of all strokes are preventable and attributable to 10 modifiable risk factors, including cardiac causes (atrial fibrillation, myocardial infarction, rheumatic valve disease, and prosthetic valve disease), history of hypertension, history of diabetes mellitus, current smoking, alcohol intake, regular physical activity, diet, waist-to-hip ratio, psychosocial stress and depression, and ratio of apolipoprotein B (ApoB) to apolipoprotein A1 (ApoA1). Among these, blood pressure is considered a particularly important modifiable risk factor for stroke, accounting for two-thirds in

low-to-middle income countries and one-third of stroke in high-income countries.[11] The World Health Organization also emphasised that hypertension and smoking are major modifiable risk factors.[12] Four out of 10 deaths due to stroke could have been prevented by well-controlled hypertension.[12] For every 5 people aged under 65 who die of stroke, two are linked to smoking.[12] Furthermore, reduced levels of HDL cholesterol[13,14] and increased levels of triglycerides[13,14] and LDL cholesterol[15] were also identified as risk factors for stroke. HDL cholesterol may boost the reverse transport of LDL cholesterol from peripheral tissues to the liver where degradation happens, or HDL cholesterol may inhibit the oxidation of LDL cholesterol, which has been believed to play a central role in atherogenesis.[16] It has also been suggested that HDL cholesterol makes LDL cholesterol less susceptible to oxidation within the arterial wall through transporting antioxidants to LDL cholesterol.[17] Once inside the arterial wall, LDL cholesterol can become oxidatively modified and is then taken up by macrophages through scavenger receptors, leading to the formation of lipid-laden foam cells.[18] This oxidatively modified LDL cholesterol also promotes local inflammation and attracts circulating monocytes, thereby amplifying plaque formation and instability, which are critical processes in the pathogenesis of stroke.[18]

1.3 Risk prediction

There are currently two well-established tools to predict stroke risk[19] – the QStroke algorithms[20] and the Framingham Stroke Risk Profile (FSRP).[21] The QStroke algorithms have been designed to identify future stroke cases in the UK primary care population free of stroke or transient ischaemic attack or oral anticoagulants. It comprises age, sex, ethnicity, systolic blood pressure (SBP), Townsend deprivation score, smoking status, family history of coronary heart disease in first degree relative under 60 years, coronary heart disease, atrial fibrillation, congestive cardiac failure, type

1 diabetes, type 2 diabetes, treated hypertension, rheumatoid arthritis, chronic renal disease, valvular heart disease, and ratio of total cholesterol to high-density lipoprotein (HDL) cholesterol concentrations. It is noteworthy that discrimination between stroke cases and non-cases expressed with the area under the receiver operating characteristics (ROC) curves (AUC) for 5-year stroke risks from QStroke in males was markedly lower (0.60, 95% CI 0.57-0.62) compared to those from FSRP (0.75, 0.73-0.78).[19] There was also a tendency to overestimate female stroke risk at higher levels of predicted risk.[20]

In contrast, the FSRP has been validated with an AUC of over 0.66.[21] It has been the most extensively used prediction tool and recommended by the American Heart Association for physicians to predict individual risk of patients and to outline treatment plans.[22] The individual stroke risk factors and a hazard ratio (HR) for incident stroke in FSRP were derived from US adults recruited in the late 1960s and 1970s and prospective follow-up over the 1970s and 1980s (N_{total} = 5,734, $N_{stroke\ events}$ = 472, follow-up time = 10 years). A 10-year probability of incident stroke could be estimated by the integration of age, sex, use of antihypertensive medications, systolic blood pressure, presence or absence of left ventricular hypertrophy on electrocardiography, prevalent cardiovascular disease (CVD), current smoking status, prevalent diabetes mellitus and prevalent atrial fibrillation. A new version of FSRP was described in 2017, and left ventricular hypertrophy was omitted because of its low prevalence (1.6%) in the Framingham Heart Study sample. Although the new FSRP better predicted the risk of stroke in two external cohorts (N_{total} = 2,949, N_{cases} = 262, follow-up time = 10 years in the “3 Cities” study, AUC = 0.70 in men and 0.72 in women; N_{total} = 23,691, N_{cases} = 629, follow-up time = 5 years in the “Reasons for Geographic and Racial Differences in Stroke” study, AUC = 0.66 in men and 0.71 in women), its predictive ability for stroke risk remains to be improved.[22] Abnormal levels of serum lipids, such as lower HDL or

higher triglycerides (TG), have been found to be associated with stroke risk,[23] but the FSRP did not contain any metabolites. For example, the potential for improvement has been previously shown for prediction algorithms for type 2 diabetes risk. By adding 143 metabolic biomarkers to a model with the conventional risk factors of type 2 diabetes, the discriminatory ability of the model was slightly improved as assessed with the concordance index which increased from 0.80 (95% CI 0.79-0.81) to 0.83 (0.82-0.84).[24] Although this difference is unlikely to be clinically meaningful, there is potential for metabolic biomarkers to add more predictive ability under other scenarios (e.g. other diseases).

1.4 Potential novel risk factors and predictors

Development in high-throughput technologies has introduced new opportunities to understand the pathophysiology of various diseases, and multi-omics data have produced large amounts of information at different molecular levels, including genes (genomics), ribonucleic acids (transcriptomics), proteins (proteomics), and metabolites (metabolomics).[25] Metabolites are small molecules (generally less than 1500 Da) from entities such as amino acids, carbohydrates, lipids, and nucleotides,[25] which are the products and substrates of metabolism that propel vital cellular functions (e.g. signal transduction and apoptosis, energy storage and production).[26]

Metabolomics is the thorough study of metabolic profiling (various metabolites from diverse molecular pathways) to identify multiple classes of small-molecule metabolites from a variety of sample types (e.g. serum, urine, culture media, and cell extracts).[27,28] The usefulness of metabolomics lies with the concept that metabolites fall downstream of genetic, transcriptomic, proteomic, and environmental variation, therefore, capturing the most dynamic and integrated information of medical condition and phenotypic changes.[28] Furthermore, progress in technology facilitates

high-throughput measurements based on large screening processes to prevent selection bias and provides broad lists of molecules for assessment as biomarkers.[25]

The human metabolome includes all metabolites that are driving our metabolism. The metabolome is regulated on the one hand by the genome and proteome, and on the other hand by the exposome, that is the aggregate of all exposures of a person including diet, lifestyle, environment and medication.[29] Despite the technical advancements in analytical platforms, it has been proven to be difficult to capture the entire metabolome, which comprises 217,920 compounds in The Human Metabolome Database 5.0 in 2022.[30] The two most commonly used approaches to characterise the metabolome are nuclear magnetic resonance (NMR) and mass spectrometry (MS).[28] NMR is quantitative, highly reproducible and is thus a time-effective approach for large-scale studies.[31] The disadvantage of NMR is the low number of compounds that can be identified.[32,33] Unlike MS, which is more sensitive, NMR cannot detect compounds with very low molecular weights. The signals of many compounds overlap, resulting in complex peaks on the NMR spectra involving multiple compounds that are difficult to disaggregate into the chemical molecules that underlie the peak. MS is based on mass/charge ratio[28] and may cover a wider spectrum of metabolites than NMR. However, although the number of compounds identified is still rapidly increasing and outnumber that of NMR, many compounds are not yet identified. MS is more time consuming and thus more costly and less easy to scale up to high throughput compared to NMR.[34] A potential challenge of MS is that quantification of metabolite concentrations is estimated instead of directly quantified. The selection of analytical platforms is a trade-off between bigger sample size (NMR) and wider metabolite coverage (MS). In this thesis, I used data from UK Biobank (UKB), which included metabolite measurements from the Nightingale Health platform, an example of a

broadly used NMR platform to quantify various lipoproteins and other metabolites in a cost-effective manner.[35,36] In contrast to the clinical chemistry measurement, the NMR platform yields quantification of a wide range of low-molecular-weight metabolites together with very detailed lipoprotein subclass profiling, some abundant proteins, and many fatty-acid levels in a single experiment (**Supplementary Figure 1.1**).[36,37] The NMR platform technically differs from the traditional biochemical and immunoassay measurement in its minimal sample preparation, non-destructive approach, capability to high throughput, and detecting compounds hard to derivatize.[33] The accuracy of metabolic quantification and pricing by the NMR platform is comparable to what is ordinarily reached by assays used in clinical chemistry.[35,36,38]

1.5 Integration of metabolomics and genetic data

Research on human genetics is one way to understand the role of metabolites in stroke. Observational studies are liable to various biases, including residual confounding by various diseases, medication, and lifestyle factors, which restrained their ability from inferring if metabolic effects are causal, rather than just predictive for stroke. Elucidating the causal relationship of metabolites with stroke could contribute to intervention targets for therapies. Considering the high heritability of metabolic levels (proportion of metabolic variance due to genetic factors $h^2_{479\ \text{significant metabolite loci}} = 0.47$),[29] Mendelian randomisation (MR), an instrumental variable method that uses genetic variants related to exposures to estimate their unbiased effects on outcomes,[39] has been used to evaluate evidence of causality for the relationship between lipid levels and type 2 diabetes,[40] coronary artery disease,[40] dementia,[41] and stroke[41] in UKB. The impact of genetic liability on metabolic traits in these studies[40,41] has been explored in the initial release of the NMR metabolomics in UKB (N~120,000) that the MRC Integrative Epidemiology Unit OpenGWAS platform used for the genome-wide association study

(GWAS) of NMR metabolomics.[42] Using this initial data release, Harshfield et al.[41] did not find any evidence of a causal relationship between metabolites with stroke and ischaemic stroke after accounting for multiple testing. However, the updated release with almost double the sample size might provide greater power for such analyses.[43] This second release includes data from N~274,000 participants and provides greater power for causal inference analyses.[43] Because genetic instruments in MR analyses typically explain only a small proportion of the variance in the exposure, MR studies generally have lower statistical power than observational studies of the same size. Therefore, a doubling of the sample size is particularly valuable for MR analyses, as it enhances the ability to detect causal effects.[44]

1.6 Thesis objectives

The primary aim of my DPhil project is to understand the molecular mechanisms underlying stroke and to identify predictive and causal biomarkers that may ultimately improve early risk stratification, prevention strategies, and therapeutic development for ischaemic stroke. This work is driven by the clinical need for more precise and mechanistic insights into stroke pathogenesis, with the goal of informing future research, guiding health policy, and improving clinical outcomes. The motivation for the work is based on a clear clinical need to identify new insights into ischaemic stroke and to investigate the causal association between metabolites and ischaemic stroke, independent of vascular risk factors. Previous researchers studied the incidence of overall stroke comprising a heterogeneous group of patients with different subtypes that were not always classified correctly in UKB data sources.[45] I have focused on patients with incident ischaemic stroke because the positive predictive value for incident ischaemic stroke codes in UKB is over 80%, while for incident intracerebral haemorrhage codes it is below 50%.[45] Furthermore, previous literature found that

metabolites were more relevant to incident ischaemic stroke than haemorrhagic stroke.[46,47] The specific objectives of this thesis are:

(1) To evaluate the association of individual blood metabolic levels with incident ischaemic stroke in UKB using conventional statistical methods, and to assess the consistency of these associations by comparing the findings with those from a nested case-control study within the China Kadoorie Biobank (CKB) and a consortium of seven studies on metabolomics and stroke. This cross-cohort comparison will help identify robust and replicable metabolic markers across populations with different genetic, lifestyle, and environmental backgrounds.

(2) To assess to what extent metabolites can jointly predict incident ischaemic stroke with machine learning algorithms and evaluate the additive predictive value of metabolites over risk factors in FSRP.

(3) To ascertain the causality of the relationships examined above using MR. This approach enables a distinction between associations and causation, identifying which associations are likely to reflect underlying biological mechanisms rather than confounding or reverse causation.

Each objective contributes to the overarching aim of elucidating the molecular mechanisms and predictive signatures of ischaemic stroke. Objective 1 helps identify candidate metabolites through statistical association and external validation; Objective 2 evaluates their potential utility in risk prediction; and Objective 3 uses genetic instruments to assess whether these associations are likely to be causal, thus providing mechanistic insights. Together, these objectives form a coherent pipeline from biomarker discovery to evaluation and causal inference.

The structure of this thesis reflects the sequence of the research objectives and methods. **Chapter 1.** introduces the background, rationale, and existing knowledge gaps. **Chapter 2.** provides a literature review, summarising existing evidence on metabolites, ischaemic stroke, and related epidemiological and genetic studies. **Chapter 3.** outlines the study methodology, including data preprocessing, variable definitions, and the statistical approaches applied in subsequent analyses. **Chapter 4.** addresses Objective 1 by evaluating associations between individual metabolites and incident ischaemic stroke in UKB, followed by comparison with results from CKB and a consortium of seven studies on metabolomics and stroke. **Chapter 5.** addresses Objective 2 by applying machine learning algorithms to assess the additional predictive value of metabolites beyond conventional risk factors. Predicting the risk of ischaemic stroke specifically, rather than composite cardiovascular disease, is clinically valuable because ischaemic stroke has distinct pathophysiology, higher long-term disability burden, unique preventive strategies, and requires subtype-specific interventions. **Chapter 6.** addresses Objective 3 through Mendelian randomisation analyses to examine the causal relationship of metabolites to ischaemic stroke. Finally, **Chapter 7.** integrates the key findings across all analyses, discusses their implications for biological understanding, prediction, and prevention of ischaemic stroke, and outlines potential directions for future research and clinical translation.

Chapter 2. Literature review

Chapter summary

This chapter reviews evidence underpinning the three aims of this thesis:

1. associations between circulating metabolites and incident ischaemic stroke;

2. incremental predictive value of metabolites beyond conventional risk factors;

3. causal inference using Mendelian randomisation (MR).

For Aim 1, a structured search identified 16 eligible studies (meta-analyses, cohorts, nested case control, and cross sectional) across biochemical or immunoassay, mass spectrometry, and nuclear magnetic resonance (NMR) platforms. Replicable associations were limited. Differences in confounder adjustment, platform coverage and quantification, and study design hindered broader replication.

For Aim 2, only two prospective NMR based prediction studies met criteria. Both showed little or no improvement in discrimination when metabolites were added to clinical predictors. These findings highlight collinearity, nonlinearity, and interaction challenges that motivate the machine learning approaches evaluated in this thesis.

For Aim 3, eight MR studies were included. Many relied on small exposure samples (often fewer than 8,000 participants) or on standard MR methods that are vulnerable to pleiotropy. Reported signals rarely survived multiple testing or robustness checks. These limitations support the use in this thesis of expanded UK Biobank NMR data (about 274,000 participants), stronger instruments from new GWAS, and modern frameworks such as multivariable MR to address correlation and pleiotropy. Overall, the review identifies key evidence gaps and informs the analytic choices used in later chapters.

2.1 Literature review for the first objective of the thesis

The first objective of my thesis is to evaluate the association of individual blood metabolic levels with incident ischaemic stroke in UKB using conventional statistical methods.

2.1.1 Review methodology

The search terms included ((("ischaemic stroke"[All Fields] OR "ischemic stroke"[All Fields]) AND ("metabolite*"[All Fields] OR "metabolomic*"[All Fields])) OR ((metabolomic*[Text Word] OR metabolite*[Text Word] OR "metabolic level*"[Text Word] OR "metabolic profile*"[Text Word] OR "metabolic biomarker*"[Text Word] OR "metabolic marker*"[Text Word] OR "blood metabol*"[Text Word] OR "plasma metabol*"[Text Word] OR "serum metabol*"[Text Word]) AND ("uk biobank*")). Inclusion criteria comprised: adult humans; estimating the relationship between exposures and outcome; a wide range of metabolites or any specific type of metabolites in blood as the exposures; ischaemic stroke as the outcome; and cross-sectional study or case control study or prospective study or randomised trials (e.g. treatment on hyperlipidaemia). Exclusion criteria contained: animal studies; cohorts of selected groups (such as those with a specific disease); and studies only focusing on the predictive values of metabolites in a prediction model. The studies included in the literature review table were identified from the PubMed databases from their initiation to December 2024. After exporting the initial search results, titles and abstracts were then screened based on the pre-specified inclusion and exclusion criteria. Studies deemed potentially relevant were retrieved in full text and assessed again for final inclusion. Studies that met all eligibility criteria were extracted and included in **Table 2.1**. To ensure comprehensive coverage, the reference lists of three relevant

published systematic literature reviews were also manually screened for additional eligible studies.[25,33,48]

2.1.2 Results

Out of 1074 reports and three published systematic literature reviews,[25,33,48] 962 studies were excluded after title screening (**Supplementary Figure 2.1**). Among the 112 articles selected for abstract review, 73 were excluded due to non-eligible population (n = 22), exposure (n = 7), outcome (n = 21), or study design (n = 23). Of the 39 full-text articles assessed, 28 were excluded, comprising three due to the non-eligible population, three due to the non-eligible exposure, seven due to the non-eligible outcome, and 15 due to the non-eligible study design. In addition, five relevant articles were identified by screening the reference lists of three published systematic reviews. Sixteen independent studies were finally included (**Table 2.1**), including three meta-analyses,[15,47,49] seven prospective cohort studies,[41,50-56] three nested case-control studies,[46,57,58] and three cross-sectional studies.[59-61] Regarding analytical platforms, there were four reports using biochemical and immunoassay testing,[15,49,50,52] six reports using mass spectrometry (MS),[51,53,54,57,60,61] and six reports using nuclear magnetic resonance (NMR).[41,46,47,55,56,59] All the articles had an overall sample size larger than 100. The largest study was a meta-analysis of prospective cohort studies with 302,430 participants and 2534 incident ischaemic stroke cases, focusing on lipids and apolipoproteins using biochemical and immunoassay measures.[49] The smallest study was a cross-sectional study with 101 participants and 54 incident ischaemic stroke cases.[59]

2.1.2.1 Biochemical and immunoassay measures

Of the 16 selected studies, 2 meta-analyses[15,49] and 2 independent prospective cohort studies[50,52] examined associations of biochemical and immunoassay measures with incident ischaemic stroke.

A meta-analysis[49] of 68 prospective studies (N_{total} = 302,430, N_{cases} = 2,534, median follow-up = 8.1 years) revealed that non-HDL cholesterol (HR per 1 standard deviation (SD) increment = 1.12, 95% CI 1.04-1.20), the ratio of non-HDL cholesterol to HDL cholesterol (1.14, 1.05-1.24), the ratio of ApoB to ApoA1 (1.13, 1.05-1.21), and ApoB (1.19, 1.05-1.34) were positively associated with incident ischaemic stroke, after adjusting for age, sex, SBP, smoking status, history of diabetes, body mass index (BMI), and lipid measures. However, this meta-analysis provided little evidence that TG (1.02, 0.94-1.11), HDL cholesterol (0.93, 0.84-1.02), or ApoA1 (0.97, 0.88-1.08) were associated with incident ischaemic stroke. In this meta-analysis, LDL cholesterol was directly measured in a subset of eight studies comprising 44,234 participants, and the analysis of these data was restricted to coronary heart disease rather than ischaemic stroke.

Table 2.1 Literature review for the first objective of the thesis: (1) meta-analyses

The first objective of the thesis is to evaluate the association of individual blood metabolic levels with incident ischaemic stroke in UKB using conventional statistical methods.

Analysis platform and sample preparation	Author	Title	Covariates	Geographic location and ethnicity	Age, years (mean)	N (cases/total)	Sex (% male)	Follow-up, years	Key outcomes
Biochemical and immunoassay testing; Multiple sources.	The Emerging Risk Factors Collaboration[49] (2009)	Major lipids, apolipoproteins, and risk of vascular disease	Age, sex, SBP, smoking status, history of diabetes, BMI, and lipid measures.	Worldwide White, Asian, Black and other	59	2534/302430	57%	8.1 (median, total) 6.1 (median, cases)	Incident IS was positively associated with non-HDL-C, non-HDL-C/HDL-C, ratio of apolipoprotein B to apolipoprotein A1, and apolipoprotein B.
Biochemical and immunoassay testing; Multiple sources.	Cholesterol Treatment Trialists' Collaborators[15] (2012)	The effects of lowering LDL cholesterol with statin therapy in people at low risk of vascular disease: meta-analysis of individual data from 27 randomised trials	Age, gender, HDL-C, LDL-C, treatment for hypertension, SBP, DBP, creatinine, history of MI, history of other coronary heart disease but no MI, history of stroke, history of PAD, other/nonspecific vascular disease history, and history of DM.	Worldwide Multi-ethnic	62.8	3330/174149	73%	5.1 (median, total)	Reduction of LDL-C with statin therapy significantly reduced the risk of IS.
NMR; Multiple sources.	Vojinovic[47] (2021)	Association of circulating metabolites in plasma or serum and risk of stroke: Meta-analysis from seven prospective cohorts	Age, sex, BMI, lipid-lowering medication, smoking status, DM, and hypertension.	Europe Multi-ethnic	Mean age from 44.5 to 75.2 (healthy subjects) Mean age from 58.1 to 76.9 (cases)	1278/38797	36.7% to 74.6%	2 to 15 (mean, total)	Ischaemic stroke was positively correlated with glycoprotein acetyls, phenylalanine and pyruvate, and inversely correlated with cholesterol in large HDL, free cholesterol in large HDL, cholesterol in HDL 2 and histidine.

NMR, nuclear magnetic resonance. SBP, systolic blood pressure. BMI, body mass index. HDL-C, high-density lipoprotein cholesterol. LDL-C, low-density lipoprotein cholesterol. DBP, diastolic blood pressure. IS, ischaemic stroke. MI, myocardial infarction. PAD, peripheral arterial disease. DM, diabetes mellitus.

Table 2.1 (Continued) Literature review for the first objective of the thesis: (2) prospective cohort studies

Analysis platform and sample preparation	Author	Title	Covariates	Geographic location and ethnicity	Age, years (mean)	N (cases/total)	Sex (% male)	Follow-up, years	Key outcomes
Biochemical and immunoassay testing; Plasma.	Willey[50] (2009)	Lipid profile components and risk of ischemic stroke: the Northern Manhattan Study (NOMAS)	Age, sex, ethnicity, education, hypertension, DM, smoking, CAD, moderate alcohol consumption, and physical activity	US (Northern Manhattan Study) Hispanic, non-Hispanic black, and non-Hispanic white	68.8	160/2940	37%	7.5 (mean, total)	There is no evidence that IS was related to baseline levels of HDL-C, non-HDL-C, LDL-C, total cholesterol, ratio of TG to HDL-C, ratio of total cholesterol to HDL-C, and TG.
Biochemical and immunoassay testing; Multiple sources.	Marklund[52] (2019)	Biomarkers of dietary omega-6 fatty acids and incident cardiovascular disease and mortality	Age, sex, race, field center, BMI, education, smoking, physical activity, alcohol intake, prevalent DM, treated hypertension, treated hypercholesterolemia, regular aspirin use, levels of α-linolenic acid, EPA, sum of trans isomers of oleic acid, and sum of trans isomers of linoleic acid.	Worldwide Multi-ethnic	49 to 77	3705/68659	Not available	2.5 to 31.9 (median, total)	IS was inversely correlated with linoleic acid.
GC-MS and LC-MS; Serum.	Sun[51] (2019)	A prospective study of serum metabolites and risk of ischemic stroke	Age, sex, race, field center, batch, DM, hypertension, current smoking, BMI, and estimated GFR.	US (Atherosclerosis Risk In Communities) White, African American	53.6	346/3904	40%	26 (median, total) 22.5 (mean, healthy subjects) 11.6 (mean, cases)	Incident IS was correlated with hexadecanedioate and tetradecanedioate.
LC-MS; Serum.	Floegel[54] (2018)	Serum metabolites and risk of myocardial infarction and ischemic stroke: a targeted metabolomic approach in	age, sex, education, smoking, alcohol intake, physical activity, fasting status, WC, BMI, and prevalent hypertension.	German	49 (cohort 1) 51 (cohort	147/2361 121/891	38% 46%	7.8 (mean, total) 7.3 (mean, total)	None of the metabolites was consistently associated with IS.

Analysis platform and sample preparation	Author	Title	Covariates	Geographic location and ethnicity	Age, years (mean)	N (cases/total)	Sex (% male)	Follow-up, years	Key outcomes
		two German prospective cohorts			2)				
NMR; Plasma.	Julkunen[55] (2023)	Atlas of plasma NMR biomarkers for health and disease in 118,461 individuals from the UK Biobank	age, sex, UK Biobank recruitment center	UK White, Asian, Black and other	58	1608/117724	46%	Not available	Histidine, albumin, creatinine, total cholesterol, clinical LDL cholesterol, HDL cholesterol, apolipoprotein A1, PUFA, LA, DHA, omega-3%, omega-6%, PUFA%, MUFA%, LA%, DHA%, and PUFA/MUFA were associated with IS.
NMR; Plasma.	Guo[56] (2022)	Circulating metabolites associated with incident myocardial infarction and stroke: A prospective cohort study of 90438 participants	age, sex, ethnicity, deprivation, smoking, alcohol, physical activity, BMI, hypertension, DM, and spectrometer.	UK White, Asian, Black and other	56	833/90662	43%	9.0 (median, total)	Albumin was inversely associated with risks of IS.
NMR; Plasma.	Harshfield[41] (2023)	Association of baseline metabolomic profiles with incident stroke and dementia and with imaging markers of cerebral small vessel disease	age, sex, UK Biobank recruitment center, spectrometer, Townsend deprivation index, taking blood pressure medications or statins, BMI, smoking status, and type 2 DM status.	UK White, Asian, Black and other	56	2053/118021	46%	13.1 (median, total)	The results are not reliable because several associations between metabolites and incident ischaemic stroke are accompanied by implausible hazard ratios.

GC-MS, gas chromatography-mass spectrometry. LC-MS, liquid chromatography-mass spectrometry. NMR, nuclear magnetic resonance. PUFA, polyunsaturated fatty acids. DHA, docosahexaenoic acid. MUFA, monounsaturated fatty acids. LA, linoleic acid. CAD, coronary artery disease. DM, diabetes mellitus. EPA, eicosapentaenoic acid. BMI, body mass index. WC, waist circumference. GFR, glomerular filtration rate. HDL-C, high-density lipoprotein cholesterol. LDL-C, low-density lipoprotein cholesterol.

Table 2.1 (Continued) Literature review for the first objective of the thesis: (3) nested case-control studies

Analysis platform and sample preparation	Author	Title	Covariates	Geographic location and ethnicity	Age, years (mean)	N (cases/total)	Sex (% male)	Follow-up, years	Key outcomes
LC-MS; Plasma.	Ament[53] (2022)	Nucleosides associated with incident ischemic stroke in the REGARDS and JHS cohorts	age, sex, race, age by race interaction, current smoking status, SBP, hypertension, DM, CVD, left VH, and AF.	US non-Hispanic Black and White	68	1075/2043	52%	7.1 (mean, total)	Guanosine and pseudouridine were associated with incident ischemic stroke.
LC-MS; Plasma.	Niu[57] (2024)	Associations of plasma metabolites with risks of incident stroke and its subtypes in Chinese adults	age, BMI, smoking status, drinking status, regular exercise, hypertension, AF, left VH, fasting glucose, serum total cholesterol, HDL-C, TG, family history of stroke, and storage time of samples	China Asian	67	321/642	63%	5.3 (median, total)	Glutamate was associated with ischaemic stroke.
NMR; Plasma.	Holmes[46] (2018)	Lipids, lipoproteins, and metabolites and risk of myocardial infarction and stroke	Age, sex, fasting hours, region, smoking status, and educational attainment	China Asian	43.8	1146/2612	47%	Not available	IS was positively associated with low-, intermediate-, and very LDL particles, TGs within most lipoproteins, ApoB, ratio of ApoB to ApoA1, DHA, acetoacetate, beta hydroxybutyrate, glucose, and GlycA. IS was inversely associated with cholesterol in large HDL.

LC-MS, liquid chromatography-mass spectrometry. NMR, nuclear magnetic resonance. SBP, systolic blood pressure. DM, diabetes mellitus. CVD, cardiovascular disease. VH, ventricular hypertrophy. AF, atrial fibrillation. BMI, body mass index. HDL-C, high-density lipoprotein cholesterol. TG, triglycerides. LDL-C, low-density lipoprotein cholesterol. DBP, diastolic blood pressure. MI, myocardial infarction. PAD, peripheral arterial disease. IS, ischaemic stroke. ApoB, apolipoprotein B. Apolipoprotein A1, ApoA1. DHA, docosahexaenoic acid. GlycA, glycoprotein acetyls.

Table 2.1 (Continued) Literature review for the first objective of the thesis: (4) cross-sectional studies

Analysis platform and sample preparation	Author	Title	Covariates	Geographic location and ethnicity	Age, years (mean)	N (cases/total)	Sex (% male)	Follow-up, years	Key outcomes
GC-MS and LC-MS; Serum.	Liu[61] (2017)	Discovery of metabolite biomarkers for acute ischemic stroke progression	None	China Asian	58.1	66/129	57%	Not applicable	Prevalent ischaemic stroke was associated with raised levels of mannose, galactose, betaine, carnitine, acetylcarnitine, L-isoleucyl-L-proline, LysoPE(18:2), LysoPE(20:2), LysoPE(20:4), and LysoPE(20:5), and reduced levels of alanine, glycine, serine, isoleucine, threonine, aspartate, proline, ornithine, lysine, tricarballylic acid, trihydroxy palmitic acid, LysoPC(16:0), PA(18:3/0:0), PC(1:0/16:0), PC(5:0/5:0), and PI(22:2/0:0).
LC-MS; Serum.	Jiang[60] (2011)	A metabonomic approach applied to predict patients with cerebral infarction	None	China Asian	61.8	67/129	51%	Not applicable	Betanin, adenosine, sucrose 6-phosphate, folic acid, deoxocathasterone, tetrahydrofolate, and aldosterone were inversely associated with ischaemic stroke. Cysteine, hydroxyoctadecadienoic acid, S-adenosyl-homocysteine, oxidized glutathione, and hydroxyeicosatetraenoic acid were positively associated with ischaemic stroke.
NMR; Plasma.	Jung[59] (2011)	^{1}H-NMR-based metabolomics study of cerebral infarction	None	South Korea Asian	68.5	54/101	52%	Not applicable	Prevalent ischaemic stroke was related to increased pyruvate, lactate, formate, and glycolate, and decreased methanol, valine, and glutamine.

GC-MS, gas chromatography-mass spectrometry. LC-MS, liquid chromatography-mass spectrometry. NMR, nuclear magnetic resonance. PUFA, polyunsaturated fatty acids. DHA, docosahexaenoic acid. MUFA, monounsaturated fatty acids. LA, linoleic acid.

Another meta-analysis[15] of 27 randomised trials showed that there was a proportional reduction in ischaemic stroke of 21% per 1.0 mmol/L low-density lipoprotein (LDL) cholesterol reduction (N_{total} = 174,149, N_{cases} = 3,330; rate ratio = 0.79, 95% CI 0.74-0.85).

Willey et al.[50] evaluated plasma metabolites in Hispanic (53.1%), non-Hispanic black (24.3%) and non-Hispanic white (20.3%) populations in the prospective cohort Northern Manhattan Study (NOMAS; N_{total} = 2,940, N_{cases} = 160, mean follow-up = 7.5 years). They found no evidence that ischaemic stroke was related to baseline levels of HDL cholesterol, non-HDL cholesterol, LDL cholesterol, total cholesterol, TG, or of the ratios of TG to LDL cholesterol, or total cholesterol to HDL cholesterol.

Marklund et al.[52] used harmonised individual-level analyses in 21 prospective studies (N_{total} = 68,659, N_{cases} = 3,705, median study follow-up time = 2.5-31.9 years) and adjusted for age, sex, race, field center, BMI, education, smoking, physical activity, alcohol intake, prevalent diabetes mellitus, treated hypertension, treated hypercholesterolemia, regular aspirin use, levels of α-linolenic acid, eicosapentaenoic acid, sum of trans isomers of oleic acid, and sum of trans isomers of linoleic acid (LA). They found that ischaemic stroke was inversely associated with LA (HR per interquintile range = 0.88, 95% CI 0.79-0.98), but not associated with arachidonic acid (0.99, 0.90-1.10).

In summary, most of the available evidence showed that LDL cholesterol, non-HDL cholesterol, LA, the ratio of ApoB to ApoA1, and ApoB associated with ischaemic stroke, while little evidence of associations was found for other biomarkers.

2.1.2.2 Mass spectrometry

As mentioned earlier, the two leading analytical platforms for metabolomics are MS and NMR spectroscopy.[62] There were two prospective studies,[51,54] two nested case-control studies,[57] and two cross-sectional studies[60,61] that used MS spectrometry with sample sizes varying from 129 to 3904 and the numbers of cases from 66 to 346. Four studies[53,54,57,58,60] used liquid chromatography-mass spectrometry (LC-MS) only, and two[51,61] used both liquid LC-MS and gas chromatography-mass spectrometry (GC-MS).

Regarding the two prospective studies, Sun et al.[51] used GC-MS and LC-MS-based quantification protocol to measure 245 serum metabolites in a prospective, biracial population-based study (N_{total} = 3,904, N_{cases} = 346, median follow-up time = 26 years) and adjusted for age, sex, race, batch, field center, diabetes, hypertension, current smoking status, BMI, and estimated glomerular filtration rate. They found that higher levels of hexadecanedioate (HR per 1 SD increment = 1.12, 95% CI 1.07-1.17) and tetradecanedioate (1.11, 1.06-1.16) were associated with an increased risk of incident ischaemic stroke. In another prospective study, Floegel et al.[54] applied LC-MS to measure 161 serum metabolites in the European Prospective Investigation into Cancer and Nutrition (EPIC) Germany study (EPIC-Potsdam: N_{total} = 2,361, N_{cases} = 147, median time to incident ischaemic stroke = 7.8 years; EPIC-Heidelberg: N_{total} = 891, N_{cases} = 121, median time to incident ischaemic stroke = 7.3 years), but no metabolite was consistently associated with ischaemic stroke.

There were two nested case-control studies using LC-MS. Ament et al.[53] measured 162 plasma metabolites (N_{total} = 2,043, N_{cases} = 1075, median time to incident ischaemic stroke = 7.1 years) and adjusted for age, sex, race, age by race interaction, current smoking status, SBP, hypertension, diabetes mellitus, cardiovascular disease, left ventricular hypertrophy, and atrial fibrillation in Cox regression models. They

verified that guanosine (HR = 1.28, 95% CI 1.13-1.45) and pseudouridine (1.34, 1.18-1.53) were significantly associated with incident ischaemic stroke. Niu et al.[57] measured 55 plasma metabolites in middle- aged and elderly Chinese people (N_{total} = 642, N_{cases} = 321, median follow-up time = 5.3 years) and adjusted for age, BMI, smoking status, drinking status, regular exercise, hypertension, atrial fibrillation, left ventricular hypertrophy, fasting glucose, serum total cholesterol, HDL cholesterol, triglycerides, family history of stroke, and storage time of samples. They found that glutamate was positively associated with ischaemic stroke (odds ratio (OD) = 1.26, 95% CI 1.11-1.43).

The cross-sectional studies found 12 to 26 metabolites significantly associated with ischaemic stroke. Liu et al.[61] (N_{total} = 129, N_{cases} = 66) that used LC-MS and GC-MS in a Chinese population found 26 serum metabolites as potential biomarkers for prevalent ischaemic stroke without explicit consideration of any confounder. Jiang et al.[60] (N_{total} = 129, N_{cases} = 67) also found 12 metabolites associated with prevalent ischaemic stroke. These cross-sectional studies took few confounding factors into account and may be subject to reverse causation owing to the study design.

Although MS was consistently applied to analyse the sample in these seven studies, the diverse coverages of studied metabolites and sample size limitations led to limited replication.

2.1.2.3 Nuclear magnetic resonance

Compared to MS, NMR provides unparalleled advantages, including non-destructive approach, minimal sample preparation and capacity to high throughput.[62]

There were five studies[41,46,47,55,56] using NMR and incident ischaemic stroke outcomes and one[59] with prevalent ischaemic stroke as the outcome. The first study is a meta-analysis of seven prospective cohort studies[47] based on individual-participant data

with 1278 incident ischaemic stroke events among 38,797 individuals, including the Rotterdam Study (N_{total} = 2565, N_{cases} = 183, mean follow-up time = 6 years), Whitehall II study (N_{total} = 5989, N_{cases} = 126, mean follow-up time = 13 years), FINRISK study 1997 (FINRISK97, N_{total} = 6858, N_{cases} = 405, mean follow-up time = 15 years), Dietary, Lifestyle and Genetic determinants of Obesity and Metabolic syndrome (DILGOM, N_{total} = 4531, N_{cases} = 84, mean follow-up time = 7 years), PROspective Study of Pravastatin in the Elderly at Risk (PROSPER, N_{total} = 4824, mean follow-up time = 2 years), Estonian biobank (N_{total} = 10,576, N_{cases} = 261, mean follow-up time = 7 years), and Framingham Heart Study (N_{total} = 3454, N_{cases} = 219, mean follow-up time = 15 years). Risk of incident ischaemic stroke was positively associated with glycoprotein acetyls (HR = 1.13, 95% CI 1.06-1.20), phenylalanine (1.12, 1.05-1.19) and pyruvate (HR = 1.13, 95% CI 1.07-1.19), and inversely correlated with cholesterol in large HDL (0.89, 0.84-0.95), free cholesterol in large HDL (0.89, 0.82-0.95), and histidine (0.89, 0.84-0.95). The second study is a nested case-control study comprising 1146 incident patients with ischaemic stroke, 1138 incident patients with intracerebral haemorrhage and 1466 control subjects, in CKB.[46] This study found that docosahexaenoic acid (DHA), glucose, glycoprotein acetyls and ketone body levels were correlated to both ischaemic and haemorrhagic stroke.[46] Lipids and lipoproteins levels were associated with ischaemic stroke, but not with intracerebral haemorrhage.[46] Together, the two studies were also used as external validation and replication cohorts for the findings derived from the first objective of this thesis.

There were three studies[41,55,56] using the initial release of the NMR metabolomics in UKB (N~120,000) to evaluate the biomarkers of incident ischaemic stroke. Julkunen et al.[55] adjusted for limited and the same confounding factors (i.e. age, sex, and assessment centres) on over 700 disease endpoints without consideration of disease-

specific confounders, and claimed that histidine, albumin, creatinine, total cholesterol, clinical LDL cholesterol, HDL cholesterol, apolipoprotein A1, polyunsaturated fatty acids (PUFA), LA, DHA, omega_3_pct, omega_6_pct, PUFA_pct, ratio of monounsaturated fatty acids to total fatty acids (MUFA_pct), LA_pct, DHA_pct, and PUFA_by_MUFA were associated with incident ischaemic stroke. On the other hand, Guo et al.[56] found that only albumin was inversely associated with incident ischaemic stroke after adjusting for age, sex, ethnicity, deprivation, smoking, alcohol, physical activity, BMI, hypertension, diabetes, and spectrometer. In the third UKB study, Harshfield et al.[41] reported that numerous lipoprotein subclasses within small HDL and large HDL were associated with ischaemic stroke. However, several associations between metabolites and incident ischaemic stroke were accompanied by implausible HR (e.g., extremely large, small, or infinite HRs), as noted in their eTable 7[41], raising questions about the reliability of these estimates. The potential reasons are that estimates per 1 unit might be reported rather than per 1 SD as claimed while the standard errors vary by a lot for the same disease endpoint in the supplementary tables.[41]

Jung et al.[59] used ^{1}H-NMR spectroscopy to investigate metabolic profiles of plasma in a Korean population in a cross-sectional study (N_{total} = 101, N_{cases} = 54). They found that prevalent ischaemic stroke was related to increased pyruvate, lactate, formate, and glycolate, and decreased methanol, valine and glutamine, but they did not consider any confounders.

These studies used a classical epidemiological approach to data analysis and estimated the effect of individual metabolites. More evidence with large sample sizes is needed to confirm the above findings because only glycoprotein acetyls and cholesterol in large HDL related to incident ischaemic stroke were replicated.

2.1.3 Summary and gaps of the first objective: Associations of metabolites with ischaemic stroke risk

Associations between metabolites and ischaemic stroke were only evident in at least two studies for glycoprotein acetyls,[46,47] cholesterol in large HDL,[46,47] histidine,[47,55] LDL cholesterol,[46,50] ApoB,[46,49] and the ratio of ApoB to ApoA1.[46,49] The lack of replication for other metabolites may be explained in part by studies with small sample sizes (thus lower statistical power), different confounders adjusted and differences in lifestyle across populations, which restricted generalisability of the findings.

Adjustment for different sets of confounders can affect replication because confounders influence both metabolite levels and ischaemic stroke risk. When studies adjust for different variables, the estimated effect of a metabolite may differ due to over-adjustment, under-adjustment, or residual confounding. Without harmonised adjustment strategies, effect estimates are not directly comparable, making replication less likely.

In addition, heterogeneity in analytical platforms contributes to inconsistent findings. Some studies used nuclear magnetic resonance NMR-based metabolomics, while others relied on MS or biochemical and immunoassay measures. These technologies differ in sensitivity, metabolite coverage, and quantification methods, meaning that the same metabolite may be measured with different precision or not measured at all. As a result, effect sizes, detection thresholds, and significance levels are not directly comparable across platforms, which complicates replication.

Differences in study designs also limit comparability. For example, cross-sectional studies capture metabolite levels and stroke status at a single time point and are prone to effects of medication and reverse causation, whereas prospective cohort studies measure metabolite levels before stroke onset and provide stronger evidence of

temporality. Nested case-control studies often match participants on specific characteristics, which can bias effect estimates relative to population-based cohort analyses. These methodological differences influence observed associations and can explain why some metabolites show inconsistent results across studies.

The UKB has a number of strengths that enable analyses addressing the limitations identified in previous literature research. The prospective design using linkage to electronic health datasets enables analysis methods more robust to the possibility of reverse causation. The wide range of environmental, lifestyle, physical measures, and genomic information collected enables me to adjust for many possible confounders, including history of diabetes, history of hypertension, smoking, alcohol intake, physical activity, dietary intake and dietary supplement use. The large number of participants with long-term follow-up within the single metabolomics cohort provides sufficient statistical power, enhancing the ability to reliably detect and provide evidence for true associations between metabolites and ischaemic stroke.

2.2 Literature review for the second objective of the thesis

The second objective of my thesis is to assess to what extent metabolites can jointly predict incident ischaemic stroke with machine learning algorithms and evaluate the additive predictive value of metabolites over conventional risk factors.

2.2.1 Review methodology

The search terms included ("stroke"[All Fields] AND ("metabolite*"[All Fields] OR "metabolomic*"[All Fields]) AND ("predict"[All Fields] OR "prediction"[All Fields] OR "concordance index"[All Fields] OR "machine learning"[All Fields])) OR ((metabolomic*[Text Word] OR metabolite*[Text Word] OR "metabolic level*"[Text Word] OR "metabolic profile*"[Text Word] OR "metabolic biomarker*"[Text Word]

OR "metabolic marker*"[Text Word] OR "blood metabol*"[Text Word] OR "plasma metabol*"[Text Word] OR "serum metabol*"[Text Word]) AND ("uk biobank*")). Inclusion criteria comprised: adult humans; evaluating the additional predictive value of metabolites on the outcome over conventional risk factors; a wide range of metabolites or any specific type of metabolites in blood as the predictors; ischaemic stroke or stroke as the outcome; and prospective study. Exclusion criteria contained: animal studies; cohorts of selected groups (such as those with a specific disease). The studies included in my literature review table were identified from the PubMed databases from their initiation to December 2024. After exporting the initial search results, titles and abstracts were then screened based on the pre-specified inclusion and exclusion criteria. Studies deemed potentially relevant were retrieved in full text and assessed again for final inclusion. Studies that met all eligibility criteria were extracted and included in **Table 2.2**.

2.2.2 Results

As mentioned in the previous chapter, metabolomics have the potential to increase the predictive ability of risk scores for stroke, but numerous unique challenges emerge from risk prediction based on hundreds of metabolites, which make application of conventional regression methods challenging or inappropriate. First, metabolites are highly correlated, making their simultaneous use in a regression model precarious. Second, metabolites may have nonlinear effects on disease. Third, it is difficult to manually detect interactions among hundreds of variables. Machine learning techniques have been proposed as cutting-edge alternatives to conventional regression methods to overcome these issues by simultaneously handling many variables and taking intricate interactions between variables into account without prior assumptions.[63-65] Machine

learning offers new opportunities in epidemiological research to deal with highly correlated metabolites, which have non-linear effects and may interact.

Out of 474 reports, 433 studies were excluded after title screening (**Supplementary Figure 2.2**). Among the 41 articles assessed at the abstract screening stage, 33 were excluded due to the non-eligible population (n = 4), exposure (n = 2), outcome (n = 17), or study design (n = 10). Of the eight full-text articles reviewed, six were excluded because the outcomes did not meet the inclusion criteria, leaving two studies[66,67] selected for this review (**Table 2.2**). Buergel et al.[67] applied 168 directly measured metabolites of the Nightingale NMR platform and eight established atherosclerotic CVD risk factors (age, sex, smoking status, prevalent diabetes mellitus, SBP, total cholesterol, HDL cholesterol, and antihypertensive drugs) to predict cerebral stroke by residual neural network in UKB (n = 117,981), including 1739 incident patients with all types of stroke. The discriminatory ability of the model with only metabolites was moderate (C-statistic 0.686, 95% CI 0.673-0.698). The 20 most important metabolites included creatinine, LA, albumin, glutamine, glycine, cholesteryl esters in chylomicrons and extremely large very low-density lipoprotein (VLDL), free cholesterol in very large HDL, saturated fatty acids (SFA), MUFA, tyrosine, phospholipids in small LDL, acetoacetate, glycoprotein acetyls, glucose, acetone, leucine, histidine, average diameter for LDL particles, lactate, and free cholesterol in intermediate-density lipoprotein (IDL).[67] The eight atherosclerotic CVD risk factors yielded a C-statistic of 0.747 (0.736-0.758). Adding the 168 metabolites to that model did not significantly elevate the C-statistic [0.754 (0.742-0.764)]. More importantly, in four independent cohorts that were used to validate the findings, the metabolites did not add predictive information over age and sex.[67] However, they did not take 81 indirectly

measured metabolites (i.e. metabolites ratios) into consideration. Since this publication, NMR data have become available in around 280,000 participants.

Zhang et al.[66] also used the initial release of the NMR metabolomics in UKB (N~120,000) and applied the metabolic risk score from 18 selected metabolites based on least absolute shrinkage and selection operator regularized Cox regression models and extreme gradient boosting model (XGBoost) for prediction. After incorporating the metabolic risk score into the traditional stroke risk prediction model (age, sex, healthy lifestyle score, history of atrial fibrillation or flutter, family history of stroke, use of blood pressure medication, use of lipid-lowering medication, HDL cholesterol, and SBP) did not improve the performance of 10-year (AUC 0.771, 0.724-0.817 vs. 0.811, 0.761-0.855) or 15-year (0.771, 0.724-0.818 vs. 0.813, 0.763-0.856) ischaemic stroke risk prediction models. Furthermore, metabolic risk scores did not enhance the performance of the Framingham Stroke Risk Profile[22] in 10-year ischaemic stroke risk prediction in men (0.767, 0.716-0.817 vs. 0.769, 0.721-0.820) or women (0.727, 0.631-0.812 vs. 0.727, 0.629-0.812).

Table 2.2 Literature review for the second objective of the thesis

The second objective of my thesis is to assess to what extent metabolites can jointly predict incident ischaemic stroke with machine learning algorithms and evaluate the additive predictive value of metabolites over conventional risk factors.

Study type	Analysis platform and sample preparation	Author	Title	Geographic location and ethnicity	Age, years (mean)	Age (healthy subjects)	Age, years (mean)	N (cases/total)	Sex (% male)	Outcomes	Follow-up, years (total)	Key outcomes
Prospective cohort study	NMR; Plasma.	Zhang[66] (2024)	Plasma metabolites, systolic blood pressure, lifestyle, and stroke risk: A prospective cohort study	UK White, Asian, Black and other	56	56	61	77315/1048	49%	Incident ischaemic stroke	13.45 (median)	Adding the metabolic risk score to the traditional stroke risk prediction model (age, sex, healthy lifestyle score, family history of stroke, history of atrial fibrillation or flutter, use of blood pressure medication, use of lipid-lowering medication, HDL-C, and SBP) did not improve the performance of 10-year or 15-year risk prediction models. Metabolic risk scores did not enhance the performance of FSRP in IS risk prediction in men or women.
Prospective cohort study	NMR; Plasma.	Buergel (2022)	Metabolomic profiles predict individual multidisease outcomes	UK White, Asian, Black and other	58	Not available	Not available	117981/1048	46%	Incident stroke	12.2 (median)	The combination of age, sex, and metabolomic state did not contain more predictive information for cerebral stroke than age and sex only.

NMR, nuclear magnetic resonance. SBP, systolic blood pressure. HDL-C, high-density lipoprotein cholesterol. IS, ischaemic stroke. FSRP, Framingham Stroke Risk Profile.

2.2.3 Summary and gaps of the second objective

Regarding the prediction of stroke, previous researchers studied the incidence of overall stroke comprising a heterogeneous group of patients with different subtypes that were not always classified correctly in UKB data sources.[45] In contrast, I have focused on incident ischaemic stroke patients because the positive predictive value is over 80% for incident ischaemic stroke in UKB.[45] Further, the data are now available on approximately 280,000 UKB participants and a different analysis approach is applied from Buergel et al.[67] based on gradient boosting model rather than neural networks. In the gradient boosting model, the learning scheme successively fits new models to give a more precise estimate of the predictors.[68] Neural networks were powerful for deep learning focusing on image, language, or even audio datasets.[69] However, for epidemiological data, regression trees like the extreme gradient boosting model may be more powerful.[65,69-72] While random forest constructs numerous trees independently on different random subsets of samples and aggregates the individual predictions into a collective one,[73] gradient boosting models iteratively build fixed small sized regression trees (weak learners) based on the previous tree's error and integrate them altogether to get better performance.[73,74] Compared to gradient boosting models, the neural networks are less robust to uninformative features, which are frequent in metabolic datasets.[69] Gradient boosting models more easily generate better predictions, e.g. accuracy and R^2 scores, than random forest and the neural network.[69,73] Furthermore, a powerful potential added value of gradient boosting model is the capability to deal with and visualize interactions between variables,[65] which are difficult to deconvolute in neural network.

2.3 Literature review for the third objective of the thesis

The third objective of the thesis is to asssess the causality of the relationships examined in the first and second objectives using MR.

2.3.1 Review methodology

The search terms included ((“ischaemic stroke”[All Fields] OR “ischemic stroke”[All Fields]) AND (“metabolite*”[All Fields] OR “metabolomic*”[All Fields]) AND (“Mendelian randomisation”[All Fields] OR “Mendelian randomization”[All Fields])) OR ((metabolomic*[Text Word] OR metabolite*[Text Word] OR “metabolic level*”[Text Word] OR “metabolic profile*”[Text Word] OR “metabolic biomarker*”[Text Word] OR “metabolic marker*”[Text Word] OR “blood metabol*”[Text Word] OR “plasma metabol*”[Text Word] OR “serum metabol*”[Text Word]) AND (“uk biobank*”)). Inclusion criteria comprised: adult humans; examining the causal relationship between exposures and outcome through MR; a wide range of metabolites or any specific type of metabolites in blood as the exposures; and ischaemic stroke as the outcome. Exclusion criteria contained: animal studies; and cohorts of selected groups (such as those with a specific disease). The studies included in my literature review table were identified from the PubMed databases from their initiation to December 2024. After exporting the initial search results, titles and abstracts were then screened based on the pre-specified inclusion and exclusion criteria. Studies deemed potentially relevant were retrieved in full text and assessed again for final inclusion. Studies that met all eligibility criteria were extracted and included in **Table 2.3**.

2.3.2 Results

Out of 305 reports, 286 studies were excluded after title screening (**Supplementary Figure 2.3**). Of the 19 articles that proceeded to abstract screening, eight were excluded because one investigated a non-eligible exposure and seven investigated non-eligible outcomes. Among the 11 full-text articles reviewed, three were

excluded due to the non-eligible exposure, resulting in eight studies[41,75-81] selected in the final review (**Table 2.3**).

Zhang et al.[75] conducted a two-sample MR study to evaluate the causal effects of 452 serum metabolites ($N_{European\ participants}$ = 7824; Analysis platform: MS) on ischaemic stroke (MEGASTROKE consortium: 34,217 ischaemic stroke cases and 406,111 controls of European ancestry). After correcting for multiple testing, genetically predicted levels of ADpSGEGDFXAEGGGVR* ($C_{63}H_{98}N_{19}O_{29}P$) were found to be positively associated with ischaemic stroke only in the inverse-variance weighted method, but not in the weighted median method, MR-Egger method, or MR pleiotropy residual sum and outlier (MR-PRESSO) test.

Qiao et al.[79] conducted a two-sample MR study to evaluate the causal effects of 486 serum metabolites ($N_{European\ participants}$ = 7824; Analysis platform: MS) on ischaemic stroke (60,341 ischaemic stroke cases and 406,111 controls of European ancestry). After correcting for multiple testing, genetically predicted levels of bradykinin, des-arg(9) were found to be positively associated with ischaemic stroke only in the inverse-variance weighted method, but not in the MR-Egger method or weighted median method.

Zheng et al.[76] conducted a two-sample MR study to evaluate the causal effects of 151 plasma metabolites expressed in absolute concentrations (N = 115,078 in UKB; Analysis platform: NMR) on ischaemic stroke (GIGASTROKE consortium: 62,100 ischaemic stroke cases and 1,234,808 controls of European ancestry). After correcting for multiple testing, genetically predicted levels of 7 metabolites (i.e. cholesterol in medium HDL, ratio of apolipoprotein B to apolipoprotein A1, cholesteryl esters in medium HDL, phospholipids in medium HDL, total lipids in medium HDL, concentration of medium HDL particles, and ratio of LA to total fatty acids) were found

to be associated with ischaemic stroke only in the inverse-variance weighted method, but not in the simple mode method, weighted median method, weighted mode method, or MR-Egger method. In other five studies,[41,77,78,80,81] the sample sizes of the exposures (two[41,82] from NMR and three[77,78,80] from MS; one[41] from UK Biobank) ranged from 2076[80] to 115,078[41], and the sample sizes of ischaemic stroke ranged from 29,633[82] to 484,121[78]. They found that the genetically predicted metabolites (e.g. cholesterol, glycerate, indolelactate, leucine, ornithine, taurodeoxycholate, threonate, and trimethylamine-N-oxide) were not associated with ischaemic stroke.

In summary, the available evidence showed that some genetically predicted metabolic levels associated with ischaemic stroke only in the inverse-variance weighted method, while little evidence of associations was found from other biomarkers.

Table 2.3 Literature review for the third objective of the thesis

The third objective of the thesis is to assess the causality of the relationships examined in the first and second objectives using MR.

Analysis platform and sample preparation	Author	Title	N (exposure)	N (outcome)	Key outcomes
MS; Serum.	Zhang[75] (2023)	Assessing the causal effect of genetically predicted metabolites and metabolic pathways on stroke	452 metabolites from 7824 European participants	MEGASTROKE consortium: 34217 ischaemic stroke cases and 406111 controls of European ancestry	Genetically predicted levels of ADpSGEGDFXAEGGGVR* were found to be associated with ischaemic stroke.
MS; Serum.	Qiao[79] (2021)	Evaluating Causal Relationship Between Metabolites and Six Cardiovascular Diseases Based on GWAS Summary Statistics	486 metabolites from 7824 people of European ancestry	60341 ischaemic stroke cases and 406111 controls of European ancestry	Genetically predicted levels of bradykinin, des-arg(9) were found to be associated with ischaemic stroke.
NMR; Plasma.	Zheng[76] (2024)	A combined observational and Mendelian randomization investigation reveals NMR-measured analytes to be risk factors of major cardiovascular diseases	151 metabolites from UK Biobank: 115078 participants	GIGASTROKE: 62100 cases and 1234808 controls of European ancestry	Genetically predicted levels of 7 metabolites were found to be associated with ischaemic stroke.
MS; Plasma or serum.	Lin[77] (2023)	Causal associations between gut microbiota, gut microbiota-derived metabolites, and cerebrovascular diseases: a multivariable Mendelian randomization study	81 gut microbiota-derived metabolites from Twins UK and KORA cohort: 7824 people of European ancestry	MEGASTROKE consortium: 34217 ischaemic stroke cases and 406111 controls of European ancestry	Genetically predicted levels of cholesterol, glycerate, indolelactate, leucine, ornithine, taurodeoxycholate, and threonate were not associated with ischaemic stroke.
MS; Plasma or serum.	He[78] (2024)	Causal relationship between human blood metabolites and risk of ischemic stroke: a Mendelian randomization study	486 metabolites from 7824 people of European ancestry	11929 ischaemic stroke cases and 472192 controls of European ancestry	After accounting for multiple testing, no genetically predicted metabolites were associated with ischaemic stroke.
LC-MS; Plasma.	Hu[80] (2023)	The association between trimethylamine N-oxide levels and ischemic stroke occurrence: a meta-analysis and Mendelian randomization study	Framingham Heart Study: 2076 European participants	MEGASTROKE consortium: 34217 ischaemic stroke cases and 406111 controls of European ancestry	No causal effect of trimethylamine-N-oxide on ischaemic stroke was observed in the Mendelian randomisation analysis.
NMR; Plasma.	Harshfield[41] (2023)	Association of baseline metabolomic profiles with incident stroke and dementia and with imaging markers of cerebral small vessel disease	325 metabolites from UK Biobank: 115078 participants	GIGASTROKE: 62100 cases and 1234808 controls of European ancestry	No genetically predicted metabolites were associated with ischaemic stroke.
NMR; Plasma.	Si[82] (2022)	Exploring the Causal Roles of Circulating Remnant Lipid Profile on Cardiovascular and Cerebrovascular Diseases: Mendelian Randomization Study	15 circulating lipid profiles from 24,925 individuals with European ancestry	International Stroke Genetics Consortium (ISGC): 29,633 Caucasian participants	No genetically predicted circulating remnant lipids were associated with ischaemic stroke.

MS, mass spectrometry. NMR, Nuclear magnetic resonance. LC-MS, liquid chromatography mass spectrometry. MR, Mendelian randomisation.

2.3.3 Summary and gaps of the third objective

Because of the improvement in metabolomic studies over the past decades, ischaemic stroke pathogenic pathways are now better understood. Most studies could demonstrate an association of the metabolite with ischaemic stroke but could not confirm its causal relationship. Regarding limitations in residual confounding and the potential for reverse causality, the causal relationship between ischaemic stroke and metabolites cannot be determined in observational studies. While randomised controlled trials (RCTs) contribute to the strongest causal evidence, evaluating the relationship between metabolites and ischaemic stroke poses challenges with respect to ethical considerations and cost constraints in participant recruitment.[75,83] The similarity between MR and RCTs is frequently described. However, MR studies investigate the lifelong cumulative influences of a risk factor (e.g. the metabolic level), while RCTs investigate the short-term effect of a therapy. Thus, the effect estimates from RCTs and MR studies are not directly comparable. MR can evaluate the causal inference of metabolic risk factors and provide predicted relative effects of metabolites on various disease outcomes through investigating genetic variants that have direct effects on a metabolic level.

Despite the strengths of MR methods for establishing causal associations, existing research on the effects of metabolites on stroke risk has several important limitations. First, most previous MR studies used relatively small metabolite GWAS datasets, often ranging from only 2,000 to 25,000 participants,[75,77-80,82] which weakened statistical power and reduced the ability to detect true causal effects. For example, He et al[78] claimed that out of 486 analyzed metabolites ($N_{\text{European ancestry}}$ = 7,824), the genetically predicted levels of 23 known metabolites and 5 unknown metabolites were identified to be causally associated with ischaemic stroke using the inverse variance weighted

method. However, none of these associations remained significant after multiple testing correction, indicating limited power. Second, many metabolic exposures are highly correlated, which compromises the precision of causal estimates in traditional MR frameworks. The standard inverse variance weighted method is limited in addressing horizontal pleiotropy, particularly when shared genetic regulation exists across lipids and other metabolite subclasses. Therefore, my thesis using greater GWAS datasets and the up-to-date MR method (i.e. multivariable MR) tackling the horizontal pleiotropy can generate enhanced statistical power and more precise assessments of the lifelong genetic metabolic effects on ischaemic stroke.

Although the literature search in this chapter was systematic, a forward search strategy was not conducted. This may have limited the inclusion of more recent or influential studies that cited the originally retrieved articles.

In summary, the emerging area of metabolomics research holds substantial promise to the discovery of unique biomarkers for the risk of ischaemic stroke. The major limitations in the current literature reside in the absence of robust replications because of heterogeneity in study designs, small sample sizes, and features of the metabolomics platforms. Consequently, more convincing evidence from pooled results of studies in unified candidate profiling and sample collection, as well as the large cohorts with diverse ethnic groups, are the pivotal issues for generalisation of findings. Genome-wide association studies and MR benefit from shared databases, more standardised methods, and less vulnerability to confounding and reverse causality encountered in current observational studies. Hence, I would also apply these genetic approaches in my thesis to comprehend the mechanisms and the causal pathways from metabolites to ischaemic stroke.

Chapter 3. Data and Methods

Chapter summary

This chapter details the data sources, design, and analytic framework underpinning the thesis. Analyses draw on the UK Biobank prospective cohort (n = 502,507), focusing on a random subset with Nightingale Health Nuclear magnetic resonance (NMR) profiling (n = 274,355). After predefined exclusions (withdrawn consent, biomarker outliers, missingness, baseline cardiovascular disease), 184,564 participants contributed to main analyses. The NMR panel comprised 249 biomarkers (168 directly measured and 81 derived). Measurements were log-transformed, then standardised across metabolites. Genotyping and large-scale imputation followed established UK Biobank pipelines to create dense genome-wide data.

Incident ischaemic stroke was identified from linked hospital and death records using validated diagnostic codes. Covariates were taken from baseline questionnaires (e.g. ethnicity, smoking, alcohol intake, physical activity, and diet), physical measurements (e.g. body mass index and systolic blood pressure), and medical records. Missing data were imputed with a machine-learning method based on chained random forests.

The chapter shows strong agreement between NMR and routine clinical chemistry for most overlapping biomarkers, extensive correlation within and between lipoprotein subclasses and other metabolite groups, and systematic associations of metabolites with age, sex, lifestyle, adiposity, blood pressure, medication use, and diabetes. These results guide confounder adjustment, model choice for association and prediction, and provide a validated platform for the genetic and causal analyses that follow.

3.1 Study design and participants

UKB is a population-based prospective cohort study comprising 502,507 participants aged from 37 to 69 years recruited between 2006 and 2010 in 22 assessment centres throughout the UK.[84] A random subset of 274,496 individuals were characterised using the high-throughput Nuclear Magnetic Resonance (NMR)-based cardiometabolic biomarker profiling platform (Nightingale Health, Helsinki, Finland) and available for inclusion in metabolic analysis (Figure 3.1). Written informed consent was provided by all participants. Participants were able to withdraw consent at any time, so I further excluded 141 participants who withdrew consent. The ethical approvals were procured from the National Health Service North West Multi-Centre Research Ethics Committee. The current study was authorized from UKB project 61054.

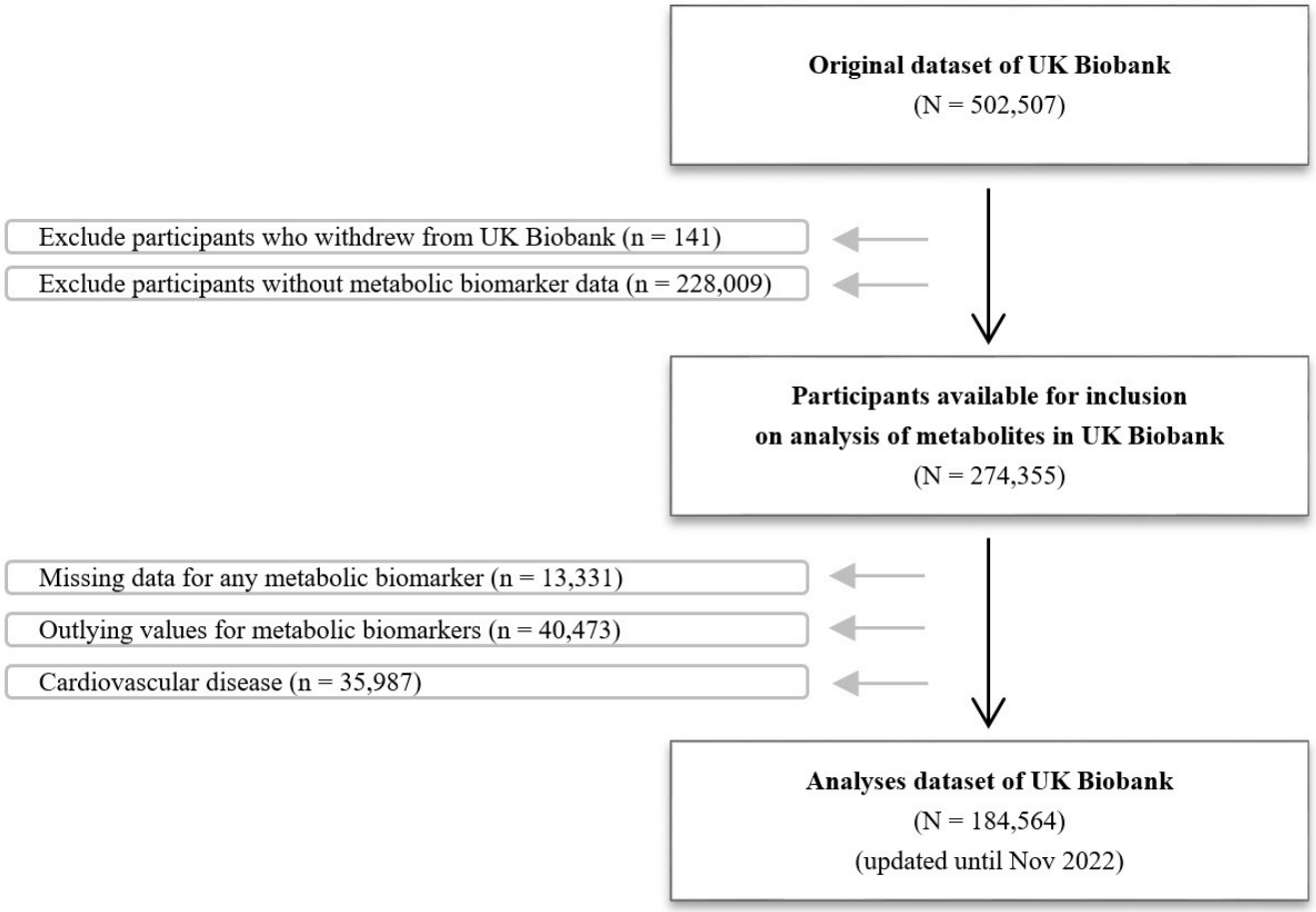


Figure 3.1 A flowchart describing the general framework

Metabolic levels outside four standard deviations from the mean were considered as outliers. Participants with missing data for any biomarker, outlying values for metabolic biomarkers and cardiovascular disease at baseline were sequentially excluded in the analyses dataset. Metabolic biomarker data were extended to the full cohort of approximately 500,000 UK Biobank participants in November 2025.

3.2 Measurement of metabolites using the Nightingale platform

The Nightingale Health's metabolic biomarker profiling platform has been qualified for clinical use (IVD 98/79/EC; Directive on In Vitro Diagnostic Medical Devices) and the laboratory information management system is accredited from ISO 13485.[55] Quality control and technical details of the metabolic biomarker data have been previously published.[35,36,55] A total of 168 directly measured metabolites and 81 derived measurements are simultaneously quantified (**Supplementary Table 3.1**), including 175 lipoprotein subclasses, 53 other lipid-related fractions, and low-molecular-weight molecules, such as 10 amino acids, 4 ketone bodies, 4 glycolysis related metabolites, 3 inflammation and fluid balance related metabolites. The UKB plasma samples characterised were obtained at baseline. A hypothesis-free approach was used to allow unbiased identification of novel metabolic biomarkers beyond established lipid or amino acid pathways, reflecting the exploratory strength of high-throughput NMR metabolomics. Metabolic levels outside four SD from the mean were considered as outliers and were excluded.[24] Metabolic measures were natural logarithmically transformed after the zero values were replaced by half of the lowest value and standardised to have mean zero and a SD of one in **Chapter 4.** due to having a non-normal distribution.[24,47] Log-transformation helps reduce skewness, making the distributions approximately normal and improving the validity of statistical inference when using linear or regression-based models. This transformation also stabilises variance across the range of metabolite concentrations, ensuring that effect estimates are less influenced by extreme outliers. After log-transformation, the interpretation of regression coefficients changes. Effect estimates represent the change in the outcome per one-unit increase in the natural log of the metabolite, which approximates a relative rather than an absolute change. Scaling all transformed metabolite measures to have a

mean of zero and a standard deviation of one further facilitates comparison across metabolites by expressing estimates per one standard deviation increase. This standardisation allows the results to be directly comparable even when metabolites are measured on different absolute concentration scales.

3.3 Genomic data processing

UKB genotyping was conducted on the Affymetrix UK BiLEVE Axiom array (N = 49,950) and the UKB Axiom array (N = 438,427).[85] Marker-based quality control (e.g., tests for plate effects, batch effects, array effects, sex effects, discordance across control replicates, and departures from Hardy–Weinberg equilibrium), sample-based quality control (e.g., 968 samples with > 5% missing rate or unusually high heterozygosity, 835 identified as sample duplicates or probably sample mishandling in the laboratory), and the genetic data imputation have been done as described previously.[85] The imputation reference panels contained (1) Haplotype Reference Consortium (HRC) with 39,235,157 single nucleotide polymorphisms (SNPs),[86] and (2) the merged UK10K,[87] and 1000 Genomes phase 3 reference panels with 87,696,888 biallelic markers.[88] The two sets of imputed data were combined, and the HRC imputation reference panel was used when a SNP was present in both panels because of its better imputation performance. The IMPUTE4 program,[89] which is a re-coded version of the haploid imputation functionality implemented in IMPUTE2,[90] was applied to perform imputation.[85] A dataset with 93,095,623 autosomal SNPs, large structural variants and short indels were procured in 487,442 individuals. The additional 3,963,705 markers were imputed on the X chromosome. The UCSC genome annotation database[91] for the Genome Reference Consortium Human Build 37 (GRCh37) assembly of the human genome was used to determine as many markers as possible in the SNP database (dbSNP) reference SNP (rs) IDs.[85]

3.4 Definitions of outcome

Hospital inpatient data from Hospital Episode Statistics (HES; Censoring date: 31 October 2022), Patient Episode Database for Wales (PEDW; Censoring date: 31 May 2022), and Scottish Morbidity Record (SMR; Censoring date: 31 August 2022) and the death data from NHS England, NHS Central Register, and National Records of Scotland (Censoring date: 30 November 2022) were used to define the first-ever incident disease based on the 9th and 10th edition of the International Classification of Disease. Prevalent diseases were identified through hospital inpatient data or baseline self-report. The ischaemic stroke outcome[45] was defined by codes 434 and 436 in International Classification of Diseases (ICD)-9 and codes I63 and I64 in ICD-10. Regarding the accuracy of ischaemic stroke phenotyping, Woodfield et al.[92] systematically extracted data on positive predictive values (PPV) with these codes based on 11 studies. They have demonstrated that PPVs of close to or greater than 90% can be achieved for ischaemic stroke, and such PPVs are sufficient for numerous large scale epidemiological studies of the determinants of ischaemic stroke.[92] Furthermore, ischaemic stroke cases in UKB can be ascertained through their full free-text electronic patient records reviewed by two stroke specialists, and adequate accuracy with PPVs greater than 80% was demonstrated.[45] Censoring for the ischaemic stroke outcome occurred at the earliest of three events: loss to follow-up, death from any cause, or the administrative end of study follow-up based on the dataset-specific censoring dates.

3.5 Definitions of covariates and baseline data

Covariates were obtained from baseline questionnaires, physical measurements, and linked health records. Questionnaire-based variables included ethnicity, smoking status, alcohol intake, physical activity, and diet. Ethnicity[84,93] was categorised as ‘White’, ‘Asian or Asian British’, ‘Black or black British’, ‘Mixed’, ‘Chinese’, and

‘Other ethnic group’. Smoking status was classified as ‘Never’, ‘Previous’ and ‘Current’. The alcohol intake frequency was grouped as "Non-drinker", "Special occasions only", "Daily or almost daily", "Three or four times a week", "Once or twice a week", and "One to three times a month." The responses "Prefer not to answer" were treated as missing data. Physical activity was classified as "low", "moderate", and "high" based on the International Physical Activity Questionnaire.[94] Thirteen self-reported baseline dietary intakes were categorised into groups: fresh fruit (< 1 servings/ day, ≥ 1 and < 2 servings/day, ≥ 2 and < 3 servings/day, ≥ 3 servings/day), whole grain foods (0 serve/day, > 0 and < 1 serve/day, ≥ 1 and < 3 serves/day, ≥ 3 serves/day), refined grain foods (0 time/day, > 0 and < 1 serve/day, ≥ 1 and < 3 serves/day, ≥ 3 serves/day), caffeinated coffee (0 cup/day, > 0 and < 2 cups/day, ≥ 2 and < 4 cups/day, ≥ 4 cups/day), decaffeinated coffee (0 cup/day, > 0 and < 3 cups/day, ≥ 3 cups/day), tea (0 cup/day, > 0 and < 3 cups/day, ≥ 3 and < 5 cups/day, ≥ 5 cups/day), vegetables (< 1 serving/day, ≥ 1 and < 2 servings/day, ≥ 2 and < 3 servings/day, ≥ 3 servings/day), cheese (< 1 time/week, ≥ 1 and < 2 times/week, ≥ 2 and < 5 times/week, ≥ 5 times/week), unprocessed red meat (0 time/week, < 1 time/week, ≥ 1 and < 2 times/week, ≥ 2 times/week), processed meat (0 time/week, < 1 time/week, ≥ 1 and < 2 times/week, ≥ 2 times/week), oily fish (0 time/week, < 1 time/week, ≥ 1 and < 2 times/ week, ≥ 2 times/week), non-oily fish (0 time/week, < 1 time/week, ≥ 1 and < 2 times/week, ≥ 2 times/week), and butter (never/rarely use spread, butter/spreadable butter, other type of spread/margarine).[24] Measurement-based variables included BMI and systolic blood pressure (SBP). BMI was constructed from dividing UKB Field 21002 "Weight" in kilograms by the square of UKB Field 12144 "Height" in metres, both measured during the initial assessment centre visit. SBP was defined by the average of two automated readings of blood pressure. Prevalent CVD was defined using

hospital and self-reported records: (1) codes 390-398, and 410-459 in ICD-9, (2) codes I00-I09, and I20-I99 in ICD-10, and (3) self-reported illness (code 1066:1068, 1074:1094, 1425, 1426, 1471, 1479, 1483:1487, 1491:1493, 1583, 1584, 1586, and 1588:1590) in UKB Field "Non-cancer illness code, self-reported." Other covariates were defined as described in more detail in **Table 3.1**.

Table 3.1 The definitions of outcome and the covariates in the study

Variable	Model	Definition
Outcome		
Incident ischaemic stroke	NA	(1) Codes 434 and 436 in ICD-9. (2) Codes I63, and I64 in ICD-10.
Covariates		
Demographic and study specific risk factors		
Age at recruitment	1	UK Biobank Field 21022 "Age at recruitment."
Sex	1	UK Biobank Field 31 "Sex."
Ethnicity	1	Ethnicity was categorized as ‘White’, ‘Asian or Asian British’, ‘Black or black British’, ‘Mixed’, ‘Chinese’, and ‘Other ethnic group’.
Lipid modifying medication	1	Medication status of lipid modifying agents from the verbal interview was classified according to Anatomical Therapeutic Chemical (ATC) codes C10.
Fasting time	1	Fasting time was defined as the interval between the consumption of food or drink and blood sampling in UK Biobank Field 74.
Spectrometers	1	UK Biobank Field 23650 "Spectrometer."
Assessment centres	1	UK Biobank Field 54 "UK Biobank assessment centre."
Behavioural risk factors		
Smoking status	2	Smoking status was classified as ‘Never’, ‘Previous’ and ‘Current’. The responses “Prefer not to answer” were treated as missing data.
Alcohol intake frequency	3	The alcohol intake frequency was grouped as "Non-drinker", "Special occasions only", "Daily or almost daily", "Three or four times a week", "Once or twice a week", and "One to three times a month." The responses "Prefer not to answer" were treated as missing data.
Dietary supplement use	4	UK Biobank Field 6155 "Vitamin and mineral supplements" and 6179 "Mineral and other dietary supplements." The responses “Prefer not to answer” were treated as missing data.
Fresh fruit	4	UK Biobank Field 1309 "Fresh fruit intake."

Variable	Model	Definition
Whole grain foods	4	Whole grain foods were defined by "Wholemeal or Wholegrain" in UK Biobank Field 1438 "Bread intake", 1448 "Bread type" and ,"Bran cereal (e.g. All Bran, Branflakes)", "Biscuit cereal (e.g. Weetabix)", "Oat cereal (e.g. Ready Brek, porridge)", and "Muesli" in 1458 "Cereal intake", and 1468 "Cereal type." One serve of grain food indicates one slice of bread, or one bowl of cereal.
Refined grain foods	4	Refined grain foods were defined by "White", "Brown", and "Other type of bread" in UK Biobank Field 1438 "Bread intake", 1448 "Bread type" and "Other (e.g. Cornflakes, Frosties)" in 1458 "Cereal intake" and 1468 "Cereal type." One serve of grain food indicates one slice of bread, or one bowl of cereal.
Caffeinated coffee	4	UK Biobank Field 1498 "Coffee intake" and 1508 "Coffee type."
Decaffeinated coffee	4	UK Biobank Field 1498 "Coffee intake" and 1508 "Coffee type."
Tea	4	UK Biobank Field 1488 "Tea intake."
Vegetables	4	UK Biobank Field 1289 "Cooked vegetable intake." One serve of vegetables indicates two heaped tablespoons of cooked vegetables.
Cheese	4	UK Biobank Field 1408 "Cheese intake."
Unprocessed red meat	4	UK Biobank Field 1369 "Beef intake", 1379 "Lamb/mutton intake", 1389 "Pork intake."
Processed meat	4	UK Biobank Field 1349 "Processed meat intake."
Oily fish	4	UK Biobank Field 1329 "Oily fish intake."
Non-oily fish	4	UK Biobank Field 1339 "Non-oily fish intake."
Butter	4	UK Biobank Field 1428 "Spread type" and 10767 "Spread type (pilot)."
Physical activity	5	Physical activity was classified as "low", "moderate", and "high" based on International Physical Activity Questionnaire of UK Biobank Field 22032 "IPAQ activity group."
Physiological risk factors		
Body mass index	6	Body mass index is constructed from UK Biobank Field 12144 "Height" and UK Biobank Field 21002 "Weight" measured during the initial assessment centre visit.
Pre-existing medical conditions		

Variable	Model	Definition
Prevalent diabetes	7	(1) Taking anti-diabetes medications (ATC code A10 Drugs used in diabetes). (2) Answer to the question “Has a doctor ever told you that you have diabetes?” as yes. (3) Self-reported diabetes, gestational diabetes, type 1 diabetes, and type 2 diabetes (1220, 1221, 1222, and 1223) in UK Biobank Field 20002 "Non-cancer illness code, self-reported." (4) HbA1c ≥ 6.5% (48 mmol/mol). (5) Code 250 (diabetes mellitus) in ICD-9. (6) Codes E10-E14 (diabetes mellitus) or O24 (diabetes mellitus in pregnancy) in ICD-10.
Systolic blood pressure	7	The average of two measures of blood pressure were used from UK Biobank Field 4080 "Systolic blood pressure, automated reading."
Others		
Prevalent cardiovascular disease	NA	(1) Codes 390-398, and 410-459 in ICD-9. (2) Codes I00-I09, and I20-I99 in ICD-10. (3) Self-reported illness (code 1066:1068, 1074:1094, 1425, 1426, 1471, 1479, 1483:1487, 1491:1493, 1583, 1584, 1586, and 1588:1590) in UK Biobank Field "Non-cancer illness code, self-reported."
Blood pressure medication	NA	Answers to the touchscreen questionnaire "Do you regularly take any of the following medications? (you can select more than one answer)" including "Blood pressure medication". (in UK Biobank Field 6153 "Medication for cholesterol, blood pressure, diabetes, or take exogenous hormones" and 6177 "Medication for cholesterol, blood pressure or diabetes")
Education	NA	UK Biobank Field 6138 "Quantifications." Education was categorized into ‘College or University degree’, ’A levels/AS levels or equivalent’, ‘Certificates of Secondary Education (CSEs) or equivalent’, ‘National Vocational Qualification (NVQ) or Higher National Diploma (HND) or Higher National Certificate (HNC) or equivalent’, ‘O levels/General Certificate of Secondary Education (GCSEs) or equivalent’, ‘Other professional qualifications’, and ‘None of the above’.
Hypertension	NA	(1) Codes 401-405 in ICD-9. (2) Codes I10, I11, I12, I13, and I15 in ICD-10. (3) Self-reported illness (code 1065, 1072, and 1073) in UK Biobank Field "Non-cancer illness code, self-reported."

The model column indicates the Cox proportional hazards regression model used in **Chapter 4.**. The pathologies were defined based on (1) the self-reported non-cancer illness using field 20002 in UK Biobank, (2) the earliest codes from hospital inpatient records and death register using the ICD coding system across all their hospital inpatient records. Participants who had a disease code in hospital admissions predating recruitment to UK Biobank or who self-reported a disease at baseline recruitment were classified as having history of disease. ICD, International Classification of Diseases.

3.6 Imputation of missing values

Missing values imputation by chained random forests of the R package *missRanger*[95,96] was performed to impute the missing values for the variables in **Supplementary Table 3.2,** and the imputed data were used in epidemiological analyses and machine learning. Initially a random forest model was produced for each variable to predict its missing values from the remaining variables in the data set. Then, this was iteratively performed for all variables and the entire process was repeated until a stopping criterion is attained (i.e. the maximum number of chained iterations up to 10 and the number of trees to 200). This non-parametric method can deal with heterogeneous data containing categorical and continuous variables simultaneously.[95] It is especially suitable when a normal distribution for all variables cannot be assumed. Furthermore, it gives the highest imputation accuracy and lowest imputation error when compared with other leading methods, including multiple imputation by chained equations (MICE),[97] nearest neighbour estimation (NN),[98] and mean imputation.[99] The information used in the imputation included baseline age, sex, BMI, ethnicity, education, diastolic blood pressure, SBP, alcohol intake frequency, smoking status, pack years of smoking, physical activity, fasting time, waist-hip ratio, antihypertensive medication, vitamin and mineral supplements, and 13 dietary variables (**Supplementary Table 3.3**).

3.7 Description of NMR metabolites by key socio-demographic and lifestyle characteristics

3.7.1 Background

Compared to the conventional clinical chemistry biomarker measurement, which has been routinely used in disease diagnosis and prediction during medical practice,

blood metabolites in the present study (i.e. cholesterol, TG, phospholipids, cholesteryl esters, free cholesterol, total lipids, lipoprotein particle concentrations, apolipoproteins, and lipoprotein subclasses) were characterised from NMR in more detail. Recently, numerous researchers have applied NMR metabolites to evaluate their associations with ischaemic stroke (see **Chapter 2.**). However, when metabolites were applied as exposures, the high correlations among metabolites may cause collinearity problems and make the interpretation of the model difficult. Furthermore, previous epidemiological research on the blood metabolome has shown age-related alterations in metabolic levels based on MS[100] and NMR[101,102] with the age-metabolite relationship usually influenced by sex.[103-105] In addition, Li et al.[106] found that circulating metabolic profiles were associated with alcohol consumption in the Framingham Heart Study Offspring cohort (N = 2,428), and via these metabolites, alcohol consumption had different influences on CVD risk. Nevertheless, few studies have explicitly investigated the relationship between the NMR metabolites and socio-demographic risk factors (e.g. age and sex) or major behavioural risk factors (e.g. smoking and alcohol intake) in a single large cohort. Understanding the characteristics of the metabolic exposures and their relationships with these potential confounding factors, will inform the statistical models used to study the associations of metabolites with ischaemic stroke in the following chapters and will facilitate the understanding and interpretation of the results of such analyses.

3.7.2 Methods

The clinical chemistry biomarkers were measured from serum samples (N ~ 480,000) of UKB participants, mainly from aliquot 1 almost without dilution.[55,107] On the other hand, the NMR was measured from ethylenediamine tetraacetic acid (EDTA) plasma samples (N = 274,355; **Figure 3.1**) from aliquot 3, expected to be 5-10%

diluted.[108] According to UKB's own quality control reports, this degree of dilution has limited impact on epidemiological or genetic analyses.[55] To compare metabolic measurements between NMR and clinical chemistry, Spearman correlation coefficients (ϱ) were used among 274,355 participants with data on NMR metabolites and clinical chemistry and eligible for inclusion in the analyses of metabolites with ischaemic stroke.

After sequential exclusion of participants with missing data for any metabolic biomarker (n = 13,331), with outlying values for metabolic biomarkers (n = 40,473), and with CVD (e.g. acute rheumatic fever, chronic rheumatic heart diseases, ischaemic heart diseases, atrial fibrillation and flutter, pulmonary heart disease and diseases of pulmonary circulation) at baseline (n = 35,987), 184,564 participants were left in the main analyses dataset. Participants with CVD were excluded to minimise the possibility of reverse causation. Spearman correlation coefficients were used to evaluate cross-correlations of the metabolites in the main analyses dataset (**Figure 3.1**). The trajectories of mean biomarker levels by age and sex were plotted in line charts using generalised additive models to smooth the lines.

Finally, linear regression was used to evaluate the relationship between selected baseline characteristics (exposure) and metabolic biomarkers (outcome) adjusting for age and sex. The baseline characteristics were sex, ethnicity, lipid modifying medication, fasting time, BMI, smoking status, alcohol intake frequency, dietary supplement use, 13 dietary variables, physical activity, prevalent diabetes, and SBP (**Table 3.1**).

3.7.3 Comparison of metabolic measurements between NMR and clinical chemistry

Nine metabolic biomarkers were measured by both NMR and clinical chemistry in UKB (**Figure 3.2**).

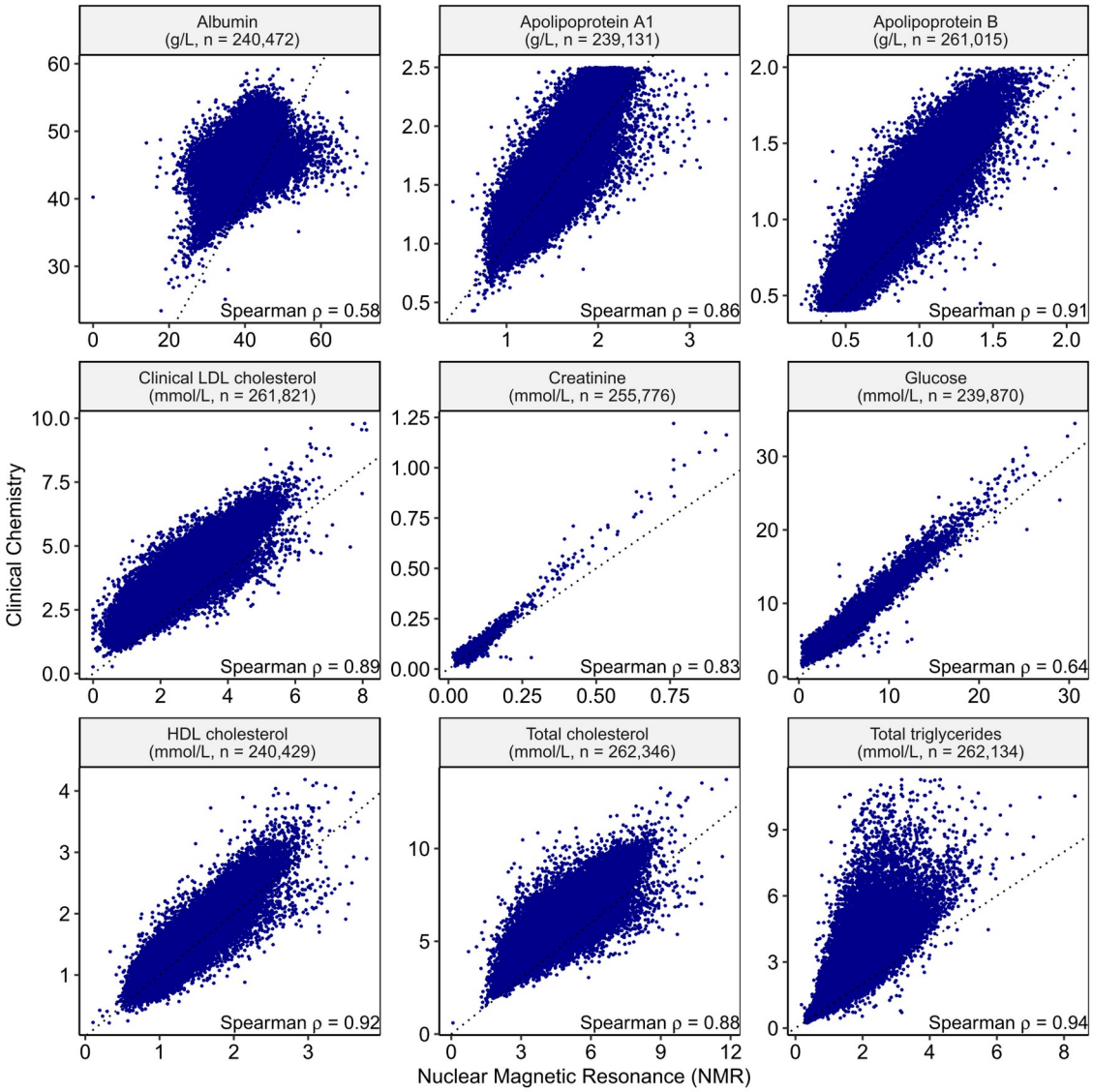


Figure 3.2 Comparisons of the NMR measurements with clinical chemistry

Scatterplots of nine metabolic biomarkers for which both NMR and clinical chemistry measurements are available in UK Biobank. The dotted line represents perfect correlation. Direct LDL from clinical chemistry was measured by enzymatic protective selection analysis and compared with the corresponding "clinical LDL cholesterol" from NMR.

There were very high correlations between NMR and most clinical chemistry measurements for ApoA1 (ϱ = 0.86), ApoB (ϱ = 0.91), clinical LDL cholesterol (ϱ = 0.89), creatinine (ϱ = 0.83), HDL cholesterol (ϱ = 0.92), total cholesterol (ϱ = 0.88), and total TG (ϱ = 0.94). The correlations between NMR and clinical chemistry were lower for albumin (ϱ = 0.58) and glucose (ϱ = 0.64).

3.7.4 Metabolic biomarkers at baseline

3.7.4.1 Cross-correlations

The Spearman correlation coefficients (ϱ) among VLDLs are illustrated in **Figure 3.3 (A),** and the full names of the metabolite abbreviations are shown in **Supplementary Table 3.1.** Strong positive correlations ($\varrho \geq 0.7$) were commonly observed between lipoprotein subclasses (e.g. concentration of particles, total lipids, phospholipids, cholesterol, cholesteryl esters, and free cholesterol) in very small, small, medium, and total VLDL. Strong positive correlations were also detected overall among lipoprotein subclasses in VLDL, chylomicrons and extremely large, very large, and large VLDL and average diameter for VLDL particles. Strong negative correlations were observed between TG to total lipids ratio and cholesterol to total lipids ratio, cholesteryl esters to total lipids ratio, free cholesterol to total lipids ratio in chylomicrons and extremely large, very large, large, medium, small, and very small VLDL.

Correlation among IDLs and LDLs are illustrated in **Figure 3.3 (B)**. Strong positive correlations were noted between lipoprotein subclasses in IDL, total, large, medium, and small LDL. Strong positive correlations were also noted between TG in IDL, total, large, medium, and small LDL.

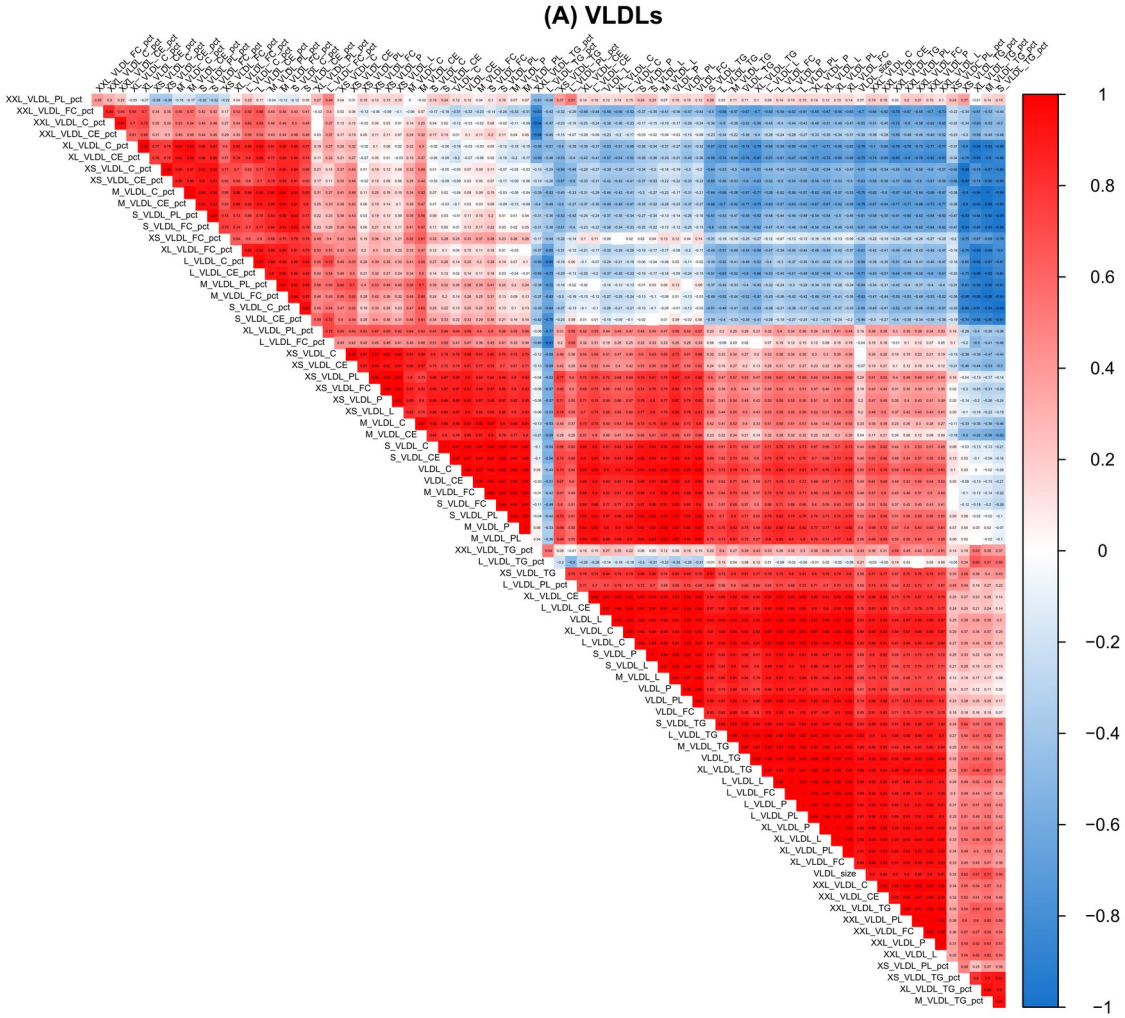


Figure 3.3 The correlation matrix of the metabolites in the analyses dataset (n = 184,567)

The Spearman's rank-order correlation was applied, and the correlation matrix was shown in the hierarchical clustering order. Blue, negative association; Red, positive association; the number in the box and the palette of colours denote the values of Spearman correlation coefficients (ϱ): range from –1 (blue), 0 (white) to 1 (red). (A) VLDLs, (B) IDLs and LDLs, (C) HDLs, (D) Amino acids, fatty acids, and other metabolites, (E) All metabolites. Abbreviations are in line with **Supplementary Table 3.1**.

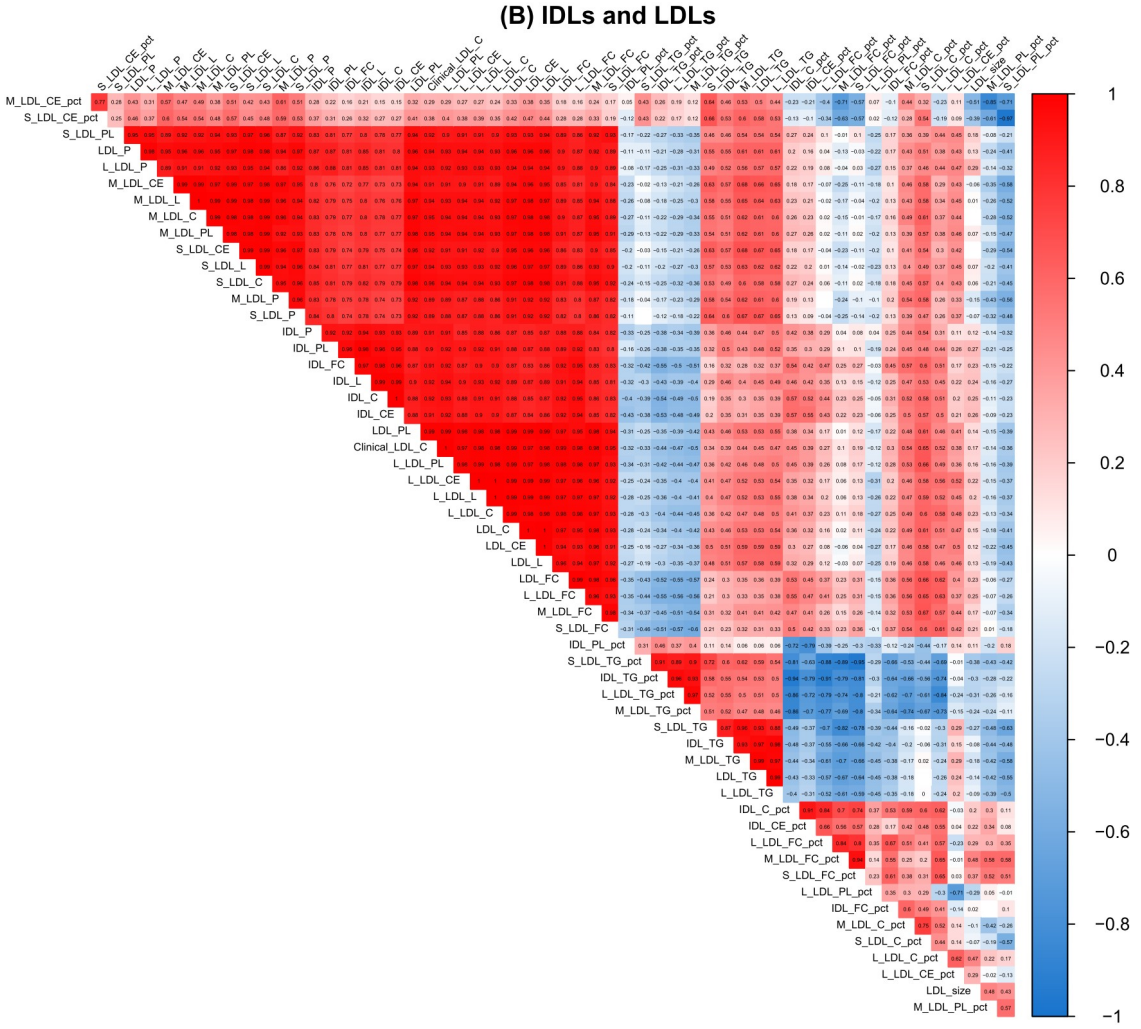


Figure 3.3 (Continued) The correlation matrix of the metabolites in the analyses dataset (n = 184,567)

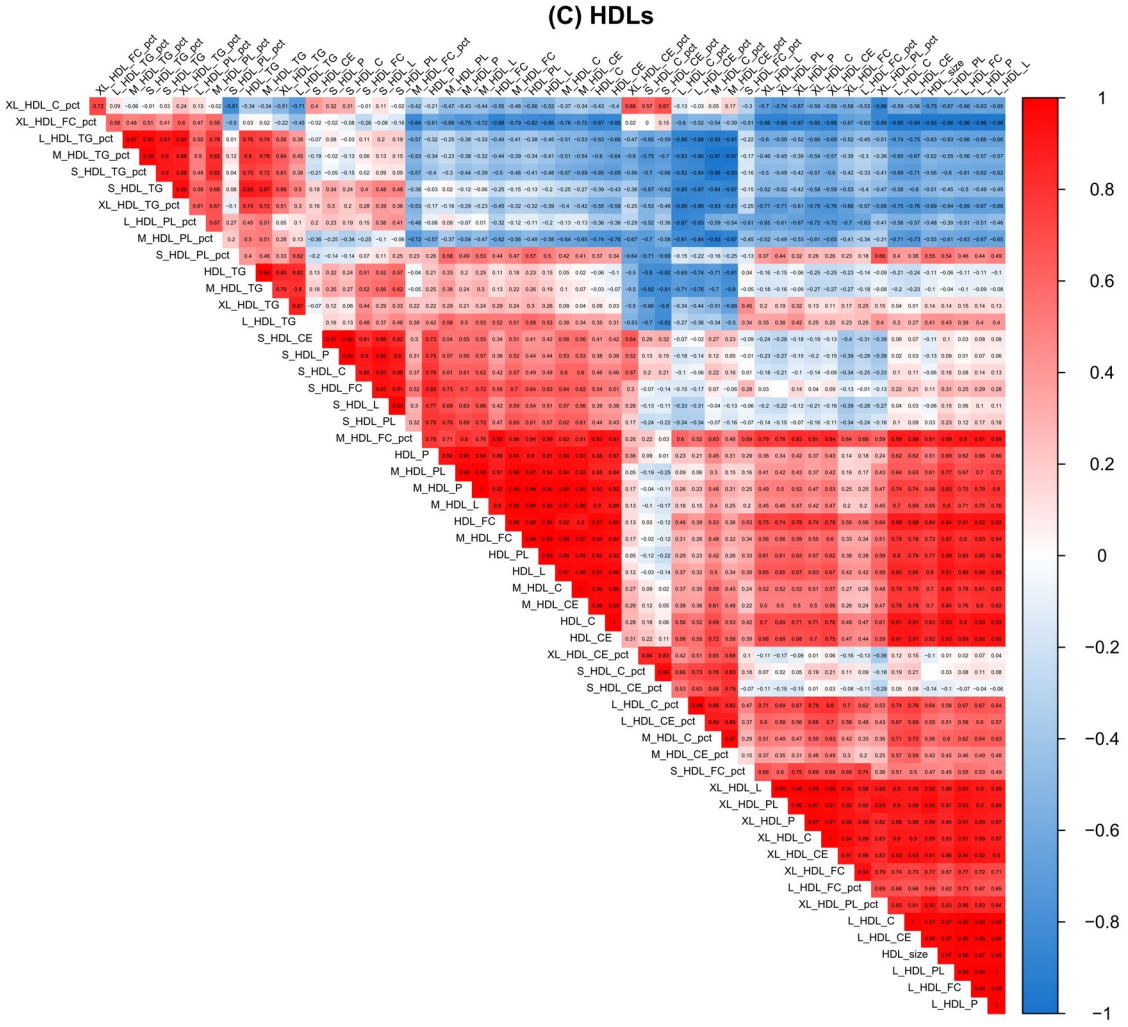


Figure 3.3 (Continued) The correlation matrix of the metabolites in the analyses dataset (n = 184,567)

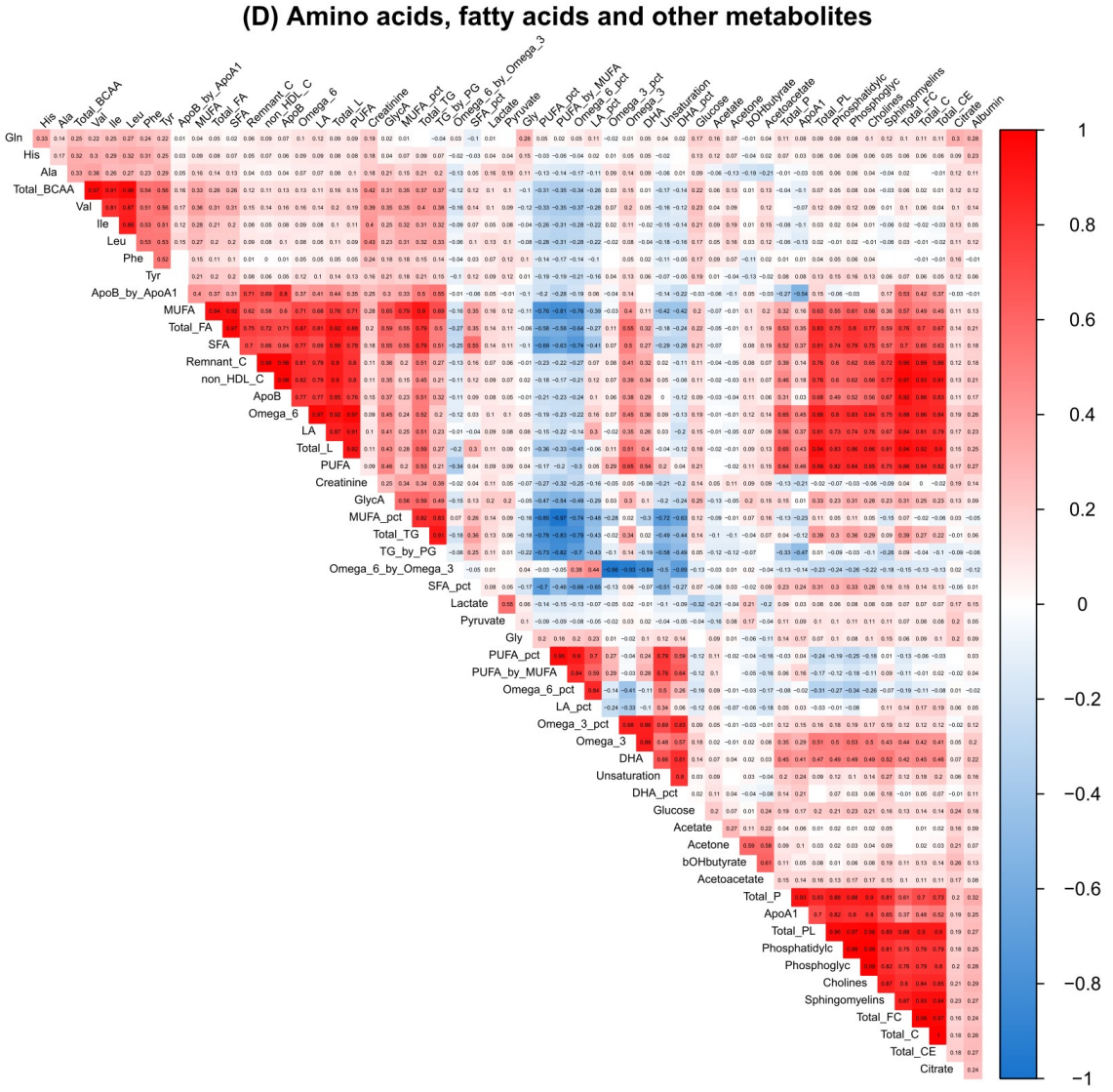


Figure 3.3 (Continued) The correlation matrix of the metabolites in the analyses dataset (n = 184,567)

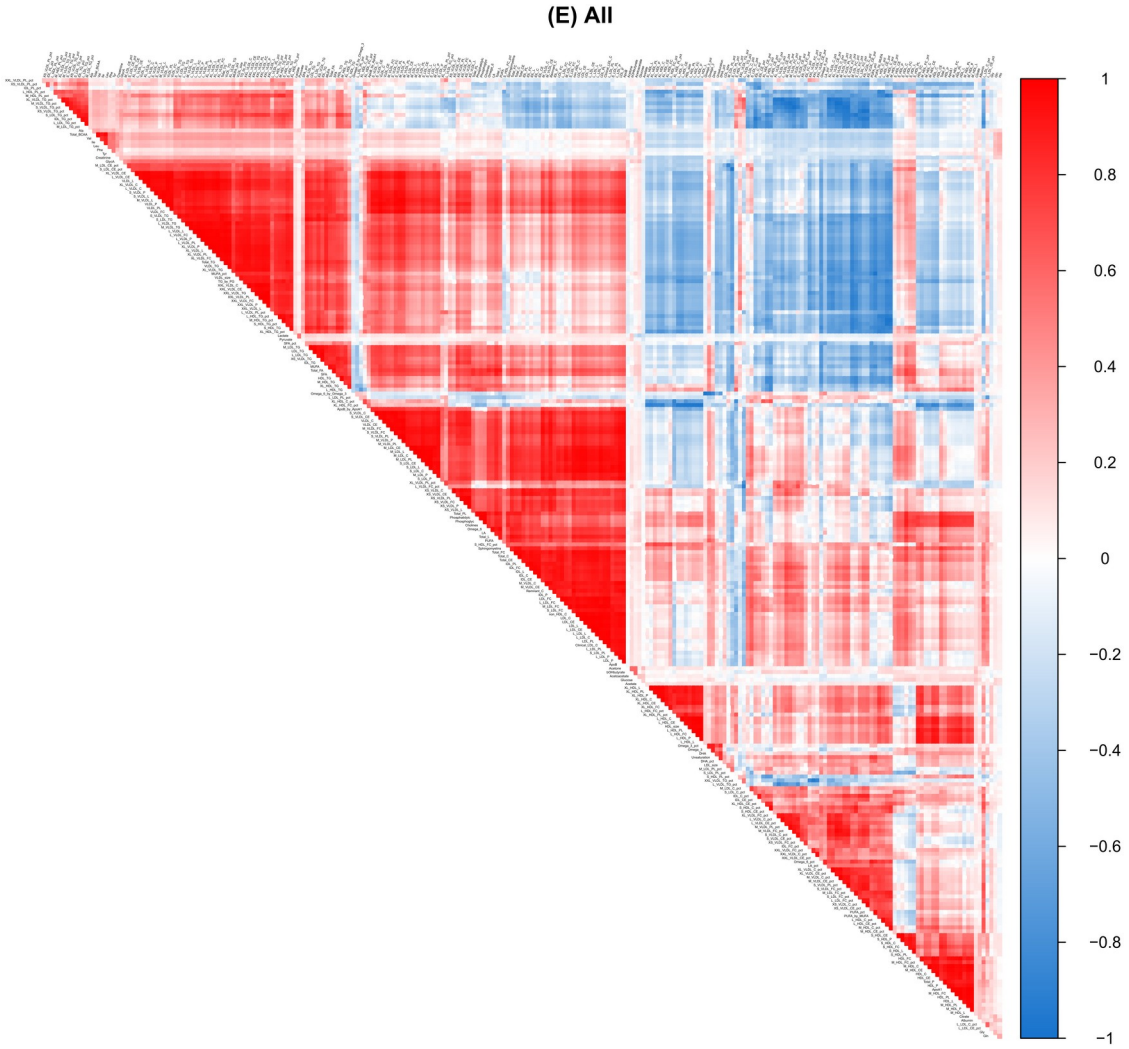


Figure 3.3 (Continued) The correlation matrix of the metabolites in the analyses dataset (n = 184,567)

There were also a great number of strong correlations among HDLs, amino acids, fatty acids (FA) and other metabolites (**Figure 3.3 (C), (D), and (E)**). For instance, levels of ApoB were positively related to lipoprotein subclasses in total, medium, small VLDL, IDL, total, large, medium, and small LDL. And the concentration of ApoA1 was positively correlated to lipoprotein subclasses in total, very large, large, medium, and small HDL.

3.7.4.2 Mean NMR biomarkers by age and sex

The trends of mean NMR biomarkers by age and sex are illustrated in **Supplementary Figure 3.1**. Mean levels of most metabolites changed by age and sex. In terms of VLDLs, the mean levels of (1) concentration of chylomicrons and extremely large, very large, and large VLDL particles; (2) phospholipids, total lipids, cholesterol, cholesteryl esters, free cholesterol, TG in chylomicrons and extremely large, very large, and large VLDL; (3) TG to total lipids ratio in chylomicrons and extremely large, very large, medium, small, and very small VLDL; (4) TG in VLDL and average diameter for VLDL particles were consistently higher in men than in women across all age groups. The largest differences in mean metabolite levels between men and women were generally greater than 0.6 standard deviations at age 40, and these differences diminished with increasing age. While the mean concentrations of some metabolites were higher in men compared to women at age 40 years, they reversed by age 70 years, including (1) concentration of total, medium, small, and very small VLDL particles; (2) total phospholipids, lipids, cholesterol, cholesteryl esters, free cholesterol in total, medium, and small VLDL; (3) phospholipids, total lipids, free cholesterol, and TG in very small VLDL.

For IDLs and LDLs, most of them showed higher mean levels in men than in women at age 40 years, then reversed by age 70 years, including (1) concentration of total, large, medium, and small LDL particles; (2) phospholipids, total lipids, cholesterol, cholesteryl esters, free cholesterol, and TG in total, large, medium, and small LDL; (3) clinical LDL cholesterol, concentration of IDL particles, and TG in IDL; (4) cholesteryl esters to total lipids ratio in large, medium, and small LDL; (5) cholesterol to total lipids ratio in medium and small LDL. Some of IDLs and LDLs were consistently higher in women than in men, including (1) phospholipids, total lipids, cholesterol, cholesteryl esters, free cholesterol in IDL (2) free cholesterol to total lipids ratio in IDL, large, medium, and small LDL; (3) cholesterol to total lipids ratio in IDL and large LDL; (4) average diameter for LDL particles.

Regarding HDLs, most of their mean concentrations were consistently higher in women than in men, (1) total lipids, phospholipids, cholesterol, cholesteryl esters, free cholesterol in total, very large, large, and medium HDL; (2) cholesterol to total lipids ratio, cholesteryl esters to total lipids ratio, and free cholesterol to total lipids ratio in large and medium HDL; (3) TG in large HDL, free cholesterol to total lipids ratio in small HDL, and phospholipids to total lipids ratio in very large and small HDL.

Almost all metabolites (244/249 = 98%) were significantly associated with sex after adjusting for age (**Supplementary Figure 3.2 (A)**). One hundred and nine metabolites (109/249 = 44%) had higher mean levels in men compared to women, and 135 metabolites (135/249 = 54%) had lower mean levels in men compared to women. The strongest positive association was for creatinine (1.09 SD higher in men than in women; 95% CI 1.09, 1.10). Conversely, the strongest inverse associations were for total lipids in large HDL (0.93 SD lower in men than in women, 95% CI -0.94, -0.92),

phospholipids in large HDL (-0.93 [-0.94, -0.92]), free cholesterol in large HDL (-0.93 [-0.94, -0.92]), and concentration of large HDL particles (-0.93 [-0.94, -0.92]).

3.7.4.3 Possible confounders in the association of ischaemic stroke

In order to understand the potential confounders of the associations between metabolites and ischaemic stroke, the first step was based on evidence of possible confounders from literature, which included a standardised international case-control study from 32 countries (INTERSTROKE Study),[10] a literature review,[11] and the World Health Organization (**Chapter 1..2**).[12] The possible confounders included age, sex, ethnicity, BMI, prevalent diabetes mellitus, prevalent hypertension, current smoking, alcohol intake, physical activity, diet intake, and dietary supplement use. Some variables (e.g. BMI, prevalent diabetes mellitus, and prevalent hypertension) were not treated as mediators because they are established vascular risk factors that may share upstream biological determinants with metabolites and operate as confounders of the metabolite-stroke association rather than lying on the causal pathway between metabolites and ischaemic stroke. Treating such variables as mediators would risk adjusting away part of the true metabolic effect or inducing collider bias, particularly when these conditions and metabolites both reflect broader cardiometabolic processes. Therefore, they were appropriately handled as confounders to isolate the independent association between metabolites and ischaemic stroke risk.

Age and sex are well-established and strong risk factors for stroke risk. For example, a prospective cohort study[109] (N_{total} = 14,357, N_{cases} = 929, mean follow-up time = 19.6 years) from 4 US communities revealed that crude incidence rates of ischaemic stroke were greater with increasing age. During 2005-2007, the incidence rate was 1.1 (95% CI 0.5-2.7) per 1000 person-years in participants aged 60 to 64, and 8.3 (95% CI 5.6-12.2) per 1000 person-years among participants aged 80 to 85. Another prospective

cohort study[110] (N_{total} = 25,789, N_{cases} = 939, mean follow-up time = 8.7 years) with black (39.9%) and white (60.1%) participants showed that at age 45 through 64 years, women were at lower ischaemic stroke risk than men (incidence rate ratio (IRR) = 0.68, 95% CI 0.49-0.94 in white individuals; 0.72, 0.52-0.99 in black individuals). At age 65 through 74 years, lower risk of ischaemic stroke in women than men remained in white participants (0.71, 0.55-0.94) but not in black participants (0.94, 0.68-1.30). However, there was no sex difference at 75 years or older for both ethnicities (0.92, 0.70-1.22 in white individuals; 0.95, 0.61-1.47 in black individuals). This study showed that the sex difference pattern of ischaemic stroke varied by ethnicity within different periods of age.

Regarding BMI, Shiozawa et al.[111] analysed JMDC Claims Database (N_{total} = 2,740,778, N_{cases} = 22,942, mean follow-up time = 3.5 years in men and 3.0 years in women) using multivariable Cox regression analysis with adjustment for age, cigarette smoking, alcohol consumption, physical inactivity, hypertension, diabetes mellitus, and dyslipidemia. They found that, compared with the normal weight group, obesity was associated with a higher incidence of ischaemic stroke (HR 1.14, 95% CI 1.06-1.23 in men; 1.13, 1.01-1.27 in women).[111]

In terms of prevalent diabetes, Kannel et al.[112] reported that, compared to participants without diabetes, 2.47-fold age-adjusted incidence of ischaemic stroke occurred in men with diabetes and 3.65-fold age-adjusted incidence in women with diabetes in the prospective study of the Framingham cohort (N_{total} = 5209, N_{cases} = 138, mean follow-up time = 20 years).

In the case-control study in 22 countries (INTERSTROKE study, N_{total} = 5337, N_{cases} = 2337),[113] diabetes mellitus (OR = 1.60, 99% CI 1.29-1.99), history of hypertension (OR = 2.37, 99% CI 2.00-2.79), current smoking (OR = 2.32, 99% CI

1.91-2.81), alcohol intake compared with never or former drinkers (OR = 1.41, 99% CI 1.09-1.82 for more than 30 drinks per month or binge drinking), and regular physical activity, defined as engaging in at least 4 hours per week of moderate or strenuous exercise, (OR = 0.68, 99% CI 0.51-0.91) were found to be significant risk factors for ischaemic stroke.

Diet quality has been strongly associated with incident stroke. Dehghan et al.[114] applied the modified Alternative Healthy Eating Index (mAHEI) derived from a qualitative food frequency questionnaire, containing meat, eggs, green leafy vegetables, raw vegetables, whole grain, refined grain, sugar, fried food, dessert, nuts, and fruit in randomised trials (N_{total} = 31,546, $N_{stroke\ events}$ = 1395, median follow-up time = 4.67 years). They found that the highest quintile of mAHEI scores was related to a 19% reduction in risk of stroke compared with the lowest quintile. Regarding dietary supplement use, folic acid was reported to be correlated to a reduced risk of stroke compared with the control condition in a systematic review of 7 randomised controlled trials (N_{total} = 24,525, $N_{stroke\ events}$ = 694, relative risk = 0.80, 95% CI 0.69-0.93).[115]

In summary, ethnicity, current smoking, alcohol intake, age, sex, BMI, prevalent diabetes mellitus, prevalent hypertension, physical activity, diet intake, and dietary supplement use were viewed as the possible confounders for associations with ischaemic stroke. In the next section, their associations with 249 metabolites were examined.

3.7.4.4 Metabolic biomarkers and selected baseline characteristics as possible confounders

3.7.4.4.1 Ethnicity

Supplementary Figure 3.2 (B) and **Table 3.2** show the association of metabolic biomarkers with ethnicity, which includes White (reference), Asian or Asian British, Black or Black British, mixed Chinese, and other ethnic groups. All metabolites were significantly associated with ethnicity after adjusting for age and sex. The strongest positive association by SD was for the degree of unsaturation (i.e. the number of double bonds in the fatty acid chain, indicated by 0 for saturated fatty acids, 1 for monounsaturated fatty acids, and 2 or more for polyunsaturated fatty acids).[116] Being of Black ethnicity was significantly associated with an average of 1.10 SD (95% CI 1.06, 1.14) higher the degree of unsaturation compared to White ethnicity. Conversely, the strongest inverse association by SD was for ratio of SFA to total FA. Being of Chinese ethnicity was significantly related to an average of 1.07 SD (-1.15, -0.99) lower in ratio of SFA to total FA compared to White ethnicity.

Table 3.2 Association of metabolic biomarkers with selected baseline characteristics

Baseline characteristics		Significant associations between metabolites and baseline characteristics		Metabolites not related to the selected baseline characteristic*
Demographic and study specific risk factors				
Sex	Male Female (reference)	244	97.99%	L_LDL_PL, XL_HDL_CE_pct, HDL_TG, S_HDL_C_pct, and LDL_P
Ethnicity	White (reference) Asian or Asian British Black or Black British Mixed Chinese Other ethnic group	249	100.00%	
Lipid modifying medication		243	97.59%	bOHbutyrate, XXL_VLDL_P, M_HDL_P, M_HDL_C, XXL_VLDL_FC, and S_HDL_P
Fasting time		234	93.98%	Unsaturation, cholines, omega_6, XS_VLDL_L, PUFA_by_MUFA, ApoA1, PUFA_pct, VLDL_FC, L_VLDL_PL_pct, SFA, Total_L, Total_PL, L_LDL_TG, MUFA_pct, and M_LDL_CE_pct
Physiological risk factors				
Body mass index		247	99.20%	Omega_3 and L_LDL_L
Behavioural risk factors				
Smoking status	Never (reference) Previous Current	241	96.79%	L_LDL_CE, L_LDL_PL, Omega_6, Total_C, Total_FC, and XXL_VLDL_TG_pct

Baseline characteristics		Significant associations between metabolites and baseline characteristics		Metabolites not related to the selected baseline characteristic*
Alcohol intake frequency	Non-drinker (reference) Daily or almost daily Three or four times a week Once or twice a week One to three times a month Special occasions only	249	100.00%	
Dietary supplement use		182	73.09%	M_VLDL_TG_pct, IDL_TG_pct, IDL_TG, S_VLDL_P, M_VLDL_C_pct, S_VLDL_L, IDL_C, S_LDL_FC, IDL_L, L_HDL_FC_pct, Omega_6_pct, IDL_CE, HDL_TG, M_LDL_PL_pct, L_LDL_PL, XL_HDL_FC, LDL_L, IDL_PL, IDL_C_pct, Acetoacetate, LDL_FC, L_LDL_FC, M_VLDL_CE_pct, M_VLDL_FC_pct, L_LDL_P, M_HDL_TG, S_VLDL_PL, M_LDL_C_pct, S_HDL_PL_pct, M_LDL_FC, LDL_PL, Clinical_LDL_C, non_HDL_C, LDL_C, LDL_CE, XS_VLDL_L, M_LDL_TG_pct, IDL_PL_pct, Acetone, S_HDL_CE_pct, XL_HDL_TG, XS_VLDL_CE_pct, IDL_FC, L_HDL_C_pct, M_LDL_C, L_LDL_TG, M_LDL_L, M_LDL_PL, S_VLDL_FC, XS_VLDL_CE, LA_pct, XS_VLDL_C, L_HDL_CE_pct, LDL_P, XS_VLDL_C_pct, XS_VLDL_FC, XS_VLDL_TG, L_LDL_L, Total_FC, S_LDL_C_pct, L_HDL_PL_pct, LDL_TG, LA, His, XS_VLDL_P, M_LDL_CE, and S_LDL_PL

Baseline characteristics		Significant associations between metabolites and baseline characteristics		Metabolites not related to the selected baseline characteristic*
Fresh fruit	< 1 servings/day (reference) ≥ 1 and < 2 servings/day ≥ 2 and < 3 servings/day ≥ 3 servings/day	240	96.39%	Gly, L_LDL_PL_pct, M_VLDL_PL_pct, S_HDL_PL_pct, S_VLDL_C_pct, S_VLDL_TG_pct, Tyr, Val, and XXL_VLDL_CE_pct
Whole grain foods	0 serve/day (reference) > 0 and < 1 serve/day ≥ 1 and < 3 serves/day ≥ 3 serves/day	247	99.20%	Gly and Phe
Refined grain foods	0 time/day (reference) > 0 and < 1 serve/day ≥ 1 and < 3 serves/day ≥ 3 serves/day	241	96.79%	ApoB_by_ApoA1, Gly, HDL_FC, IDL_CE_pct, Ile, M_VLDL_CE, Pyruvate, and Total_BCAA
Caffeinated coffee	0 cup/day (reference) > 0 and < 2 cups/day ≥ 2 and < 4 cups/day ≥ 4 cups/day	246	98.80%	Gly, Glucose, and XXL_VLDL_CE_pct
Decaffeinated coffee	0 cup/day (reference) > 0 and < 3 cups/day ≥ 3 cups/day	228	91.57%	LDL_size, PUFA, His, Glucose, Acetate, Albumin, M_VLDL_CE, S_LDL_FC, XXL_VLDL_C_pct, XXL_VLDL_CE_pct, XXL_VLDL_FC_pct, XXL_VLDL_TG_pct, M_VLDL_PL_pct, M_VLDL_FC_pct, M_VLDL_TG_pct, S_VLDL_C_pct, S_VLDL_TG_pct, XS_VLDL_FC_pct, L_LDL_PL_pct, XL_HDL_C_pct, and L_HDL_FC_pct

Baseline characteristics		Significant associations between metabolites and baseline characteristics		Metabolites not related to the selected baseline characteristic*
Tea	0 cup/day (reference) > 0 and < 3 cups/day ≥ 3 and < 5 cups/day ≥ 5 cups/day	242	97.19%	Gly, M_VLDL_CE, XXL_VLDL_C_pct, XXL_VLDL_CE_pct, S_VLDL_TG_pct, M_LDL_PL_pct, and S_LDL_C_pct
Vegetables	< 1 serving/day (reference) ≥ 1 and < 2 servings/day ≥ 2 and < 3 servings/day ≥ 3 servings/day	229	91.97%	Total_FC, Ala, Glucose, Pyruvate, XS_VLDL_CE, IDL_L, IDL_PL, L_LDL_C, L_LDL_FC, S_HDL_P, S_HDL_L, S_HDL_PL, S_HDL_C, S_HDL_CE, XXL_VLDL_CE_pct, M_VLDL_PL_pct, S_VLDL_C_pct, S_VLDL_TG_pct, XS_VLDL_FC_pct, and IDL_PL_pct
Cheese	< 1 time/week (reference) ≥ 1 and < 2 times/week ≥ 2 and < 5 times/week ≥ 5 times/week	243	97.59%	Val, bOHbutyrate, Acetoacetate, S_VLDL_CE_pct, XS_VLDL_FC_pct, and IDL_FC_pct
Unprocessed red meat	0 time/week (reference) < 1 time/week ≥ 1 and < 2 times/week ≥ 2 times/week	238	95.58%	His, XXL_VLDL_TG_pct, M_VLDL_PL_pct, M_VLDL_FC_pct, S_VLDL_C_pct, S_VLDL_TG_pct, XS_VLDL_C_pct, XS_VLDL_CE_pct, XS_VLDL_TG_pct, IDL_FC_pct, and M_LDL_TG_pct
Processed meat	0 time/week (reference) < 1 time/week ≥ 1 and < 2 times/week ≥ 2 times/week	247	99.20%	HDL_FC and His

Baseline characteristics		Significant associations between metabolites and baseline characteristics		Metabolites not related to the selected baseline characteristic*
Oily fish	0 time/week (reference) < 1 time/week ≥ 1 and < 2 times/week ≥ 2 times/week	248	99.60%	Omega_6_pct
Non-oily fish	0 time/week (reference) < 1 time/week ≥ 1 and < 2 times/week ≥ 2 times/week	239	95.98%	LDL_TG, VLDL_FC, M_VLDL_L, S_VLDL_P, S_VLDL_L, L_LDL_TG, S_HDL_L, S_HDL_PL, L_VLDL_FC_pct, and L_VLDL_TG_pct
Spread type	Never/rarely use spread (reference) Butter/spreadable butter Other type of spread/margarine	243	97.59%	Gln, Gly, His, Total_BCAA, XL_VLDL_PL_pct, and L_LDL_C_pct
Physical activity	Low Moderate High	241	96.79%	non_HDL_C, LDL_PL, LDL_CE, Total_L, LDL_L, bOHbutyrate, M_VLDL_C, and IDL_P
Pre-existing medical conditions				
Prevalent diabetes		239	95.98%	LDL_TG, Omega_3_pct, Acetone, XL_VLDL_FC, L_VLDL_L, L_VLDL_FC, M_VLDL_L, M_VLDL_TG, XXL_VLDL_TG_pct, and L_LDL_PL_pct
Systolic blood pressure		247	99.20%	Leu and M_HDL_FC_pct

*Abbreviations are in line with **Supplementary Table 3.1**.

3.7.4.4.2 Smoking Status

The associations of metabolic biomarkers with smoking status, including never smokers (reference), previous smokers, and current smokers, are illustrated in **Supplementary Figure 3.2 (C)**. The mean levels of 125 metabolites were higher among current smokers compared to never smokers, including LDL cholesterol, VLDL cholesterol, TG in LDL, total TG, ApoB, ratio of ApoB to ApoA1, SFA, ratio of MUFA to total FA, lactate, alanine, 3-hydroxybutyrate, acetoacetate, acetone, glycoprotein acetyls, and other metabolites. Of these, the strongest positive association was for ratio of MUFA to total FA. Being a current smoker was significantly associated with an average of 0.44 SD (95% CI 0.43, 0.46) higher ratio of MUFA to total FA compared to never smokers. Conversely, the levels of 104 metabolites were lower among current smokers compared to never smokers, including HDL cholesterol, ApoA1, PUFA, ratio of PUFA to total FA, ratio of PUFA to MUFA, degree of unsaturation, albumin, omega-3 FA, ratio of omega-3 FA to total FA, ratio of omega-6 FA to total FA, tyrosine, and other metabolites. Of these biomarkers, the strongest inverse association was for degree of unsaturation. On average, current smokers had 0.484 SD (-0.499, -0.469) lower degree of unsaturation compared to never smokers.

3.7.4.4.3 Alcohol intake

The associations of metabolic biomarkers with alcohol intake frequency, including non-drinker (reference), special occasions only, daily or almost daily, three or four times a week, once or twice a week, and one to three times a month, are illustrated in **Supplementary Figure 3.2 (D)**. All metabolites were associated with alcohol intake frequency. Of these, the strongest positive association was for total lipids in medium HDL (0.831 SD higher in participants with alcohol intake daily or almost daily than in

non-drinkers; 95% CI 0.813, 0.848). Conversely, the strongest inverse association was for phospholipids to total lipids ratio in very small VLDL (0.560 SD lower in participants with alcohol intake daily or almost daily than in non-drinkers; 95% CI -0.579, -0.540).

3.7.4.4.4 Other baseline characteristics

For other baseline characteristics, most of them were associated with over 90% of metabolites (**Table 3.2** and **Supplementary Figure 3.2**), including lipid modifying medication (243/249 = 96%, $N_{positive\ association}$ = 62, $N_{negative\ association}$ = 181), fasting time (234/249 = 94%, $N_{positive\ association}$ = 123, $N_{negative\ association}$ = 111), BMI (247/249 = 99%, $N_{positive\ association}$ = 136, $N_{negative\ association}$ = 111), fresh fruit consumption (240/249 = 96%), whole grain foods consumption (247/249 = 99%), refined grain foods consumption (241/249 = 97%), caffeinated coffee consumption (246/249 = 99%), decaffeinated coffee consumption (228/249 = 92%), tea consumption (242/249 = 97%), vegetables consumption (229/249 = 92%), cheese consumption (243/249 = 98%), unprocessed red meat consumption (238/249 = 96%), processed meat consumption (247/249 = 99%), oily fish consumption (248/249 = 100%), non-oily fish consumption (239/249 = 96%), type of spread (243/249 = 98%), physical activity (241/249 = 97%), prevalent diabetes (239/249 = 96%, $N_{positive\ association}$ = 65, $N_{negative\ association}$ = 174), and SBP (247/249 = 99%, $N_{positive\ association}$ = 180, $N_{negative\ association}$ = 67). Regarding lipid modifying medication, the strongest positive association was for TG to total lipids ratio in very small VLDL (0.83 SD higher in participants using lipid modifying medication, 95% CI 0.82, 0.84). Conversely, the strongest inverse association was for cholesteryl esters in medium VLDL (0.97 SD lower in participants using lipid modifying medication, 95% CI -0.98, -0.96).

In terms of oily fish intake and non-oily fish intake, the strong positive associations were for omega-3 FA to total FA (1.31 SD higher in participants using oily fish intake ≥ 2 times/week compared to those without oily fish intake, 95% CI 1.30, 1.33; 1.01 SD higher in participants using non-oily fish intake ≥ 2 times/week compared to those without non-oily fish intake, 95% CI 0.99, 1.03), ratio of DHA to total FA (1.28 [1.26, 1.30]; 0.93 [0.90, 0.95]), DHA (1.20 [1.19, 1.22]; 0.92 [0.90, 0.94]), omega-3 FA (1.05 [1.03, 1.07]; 0.85 [0.83, 0.87]). Conversely, the strongest inverse associations were for ratio of omega-6 fatty acids to omega-3 fatty acids (1.23 SD lower in participants using oily fish intake ≥ 2 times/week compared to those without oily fish intake, 95% CI -1.25, -1.22; 0.99 SD lower in participants using non-oily fish intake greater than or equal to 2 times/week compared to those without non-oily fish intake, 95% CI -1.01, -0.97).

3.7.5 Discussion

3.7.5.1 Comparison of correlations with chemistry biomarkers

High consistency was found between apolipoproteins, cholesterols, creatinine, and TG measured using both routine clinical chemistry and NMR. However, there was moderate consistency for albumin and glucose between routine clinical chemistry and NMR. Some possible reasons have been described in previous literature.[55,117] First, clinical biochemistry biomarkers were measured using serum samples, and NMR biomarkers were measured from EDTA plasma samples. There were subtle differences in absolute concentrations in UKB between EDTA plasma and serum and longer differences in sample storage time.[55] In addition, the concentration of glucose can be affected by the method used for blood processing.[118-120] Second, clinical biochemistry biomarkers were measured using blood sample aliquot 1, and NMR biomarkers were measured from aliquot 3. Blood sample aliquot 3 in the UKB was reported to be more

affected by dilution compared with aliquot 1, especially for a metabolite like albumin with a narrow biological range.[121] Third, Ritchie et al.[122] found that the origin of the technical variation was caused by the UKB sample plating process, rather than the NMR quantification pipeline. Two recent papers based on plasma samples of the China Kadoorie Biobank (CKB)[46] study ($\varrho_{(albumin)}$ = 0.8; N = 4,662) and serum samples of the FinHealth 2017 cohort[55] ($\varrho_{(albumin)}$ = 0.81; $\varrho_{(glucose)}$ = 0.98; N = 6,000), the Health 2000 study[123] ($\varrho_{(glucose)}$ = 0.77; N = 5,895), and the FINRISK 2012 cohort[123] ($\varrho_{(albumin)}$ = 0.85; N = 5,229) revealed high consistency between NMR and clinical chemistry measurement, which also implied that the low correlation for albumin and glucose in UKB resulted from the difference in the technical variations or sample types rather than the measurement technique. Because most of the NMR and clinical chemistry measurements were highly consistent, this standard NMR approach markedly facilitated the interpretation of association magnitudes when comparing with results from previous studies using clinical chemistry measurements.

3.7.5.2 Cross-correlations of metabolic biomarkers at baseline

Cross-correlations between metabolic groups and each NMR metabolite reflects the physiological and biochemical characteristics of the biomarkers. The levels of biochemically analogous metabolites, such as LDLs, were positively related with each other. These biochemical similarities of metabolites were specifically quantified by NMR.[124,125] On the other hand, metabolites that are physiologically related but biochemically different were also correlated with each other. For example, around 70% of the total HDL protein content was constituted with ApoA1, which was the most functionally major protein in HDL and important for reverse cholesterol transport, anti-thrombotic, anti-inflammatory, anti-apoptotic, and anti-atherogenic properties.[126] Although HDLs were chemically distinct from ApoA1, the fractions of the lipids within

HDLs were associated with ApoA1. Likewise, ApoB accounted for nearly 30%, 60%, and 95% of VLDL, IDL, and LDL and plays a crucial role in their catabolism.[127] In UKB, ApoB was also related to the fractions of the lipids within VLDL, IDL, and LDL although they were biochemically disparate.

In general, the cross-correlation structure within NMR metabolites follows either physiological or biochemical relationships that are applicable to the interpretation of epidemiological findings. Investigating the association of phenotypes with individual metabolic biomarkers is beneficial for the initial biomarker assessment and replication because it allows adjustment for relevant confounders and facilitates biological interpretability.[36] These high cross-correlations also inform the importance to select appropriate analytical methods to study these metabolites in isolation (**Chapter 4.**) or jointly (**Chapter 5.**).

3.7.5.3 Mean NMR biomarkers by age and sex

The mean concentrations of lipoproteins and lipids tend to differ by age and sex.[128,129] Men tend to have higher average TG and LDL cholesterol levels compared with women, which may in part be attributed to variations in sexual hormones,[129] yet men also tend to have lower average HDL cholesterol. However, the sex-attributed alterations in biomarker levels are reduced before puberty and in the old age. In line with previous reports in the Mexico City Prospective Study[124] and Avon Longitudinal Study of Parents and Children[130] about the sex and age differences in NMR metabolite levels, higher VLDLs and lower HDLs in men than in women were evident in UKB, although some of these differences diminished or even reversed by age 70 years. Mean levels of numerous VLDLs and LDLs were higher in men at younger age and then higher in women at older age, suggesting that these fluctuations may be attributed to the effects of the sexual hormones and menopause. In addition, the variations of metabolites

by age and sex could also result from pre-existing medical conditions (e.g. diabetes mellitus and hypertension) and behavioural factors (e.g. physical activity, diet, smoking status, and alcohol intake frequency).

3.7.5.4 Metabolic biomarkers and selected baseline characteristics as possible confounders

A vast majority of NMR biomarkers in UKB showed strong correlation with baseline characteristics. In addition to age and sex, behavioural factors have been demonstrated to associate with specific metabolites. Regarding smoking status, the results of UKB align with previous reports[124,131,132] which have linked smoking with higher concentrations of LDL cholesterol, TG in LDL, ApoB, SFA, and glycoprotein acetyls and lower concentrations of HDL cholesterol and ApoA1. Insulin resistance has been suggested to play a key role in the aberration of TG and HDLs and CVD in smokers.[133] Furthermore, oxidative modification of LDL can be aggravated by cigarette smoking, and autoantibody titers to oxidized LDL as well as circulating compounds of lipid peroxidation significantly elevate in smokers.[134,135] Of note, the higher levels of glycoprotein acetyls, the inflammation markers, among current smokers compare to never smokers in UKB also demonstrated that smoking may contribute to vascular inflammation.[131] With regard to alcohol intake, the results of these analyses in UKB were in line with previous reports[124,136,137] in which alcohol intake was associated with higher concentrations of HDL cholesterol, ApoA1, MUFA, total cholines and lower concentration of citrate and glutamine. The results also showed strong positive association between consumption of oily fish and non-oily fish intake with omega-3 FA and DHA, which was in line with former research.[138,139] Berr et al.[139] demonstrated that high fish consumption was associated with omega-3 FA and DHA in EVA study (N = 200). Although these associations were only adjusted for age and sex and may suffer

from residual confounding (e.g. hypertension, liver disease, adiposity, and kidney function), the results provide useful information on potential adjustment consideration for the Cox regression analyses in **Chapter 4.**.

3.7.6 Conclusion

Among 274,355 participants available for inclusion of metabolic analysis in UKB, this chapter describes the high correlation between the NMR and clinical chemistry measurements, indicating that the NMR-based quantification is consistent with established clinical assays and supports the validity and reliability of the NMR metabolite data for subsequent analyses. Furthermore, results demonstrated that NMR metabolites were associated with each other not only by their biochemical characteristics but also by pathophysiological inter-relationships. In addition, the cross-sectional associations of NMR metabolites with demographic factors, physiological factors, behavioural factors, and pre-existing medical conditions provide evidence on the importance of accounting for potential confounders when evaluating the relationships between NMR metabolites and ischaemic stroke.

Chapter 4. Epidemiological associations of individual metabolites with incident ischaemic stroke

Chapter summary

This chapter examined epidemiological associations between 249 NMR-measured metabolites and incident ischaemic stroke in 184,564 UK Biobank participants (3,002 events; median follow-up 13.8 years). Using sequentially adjusted Cox models (from demographics and technical factors to lifestyle, adiposity, diabetes, and systolic blood pressure), 40 metabolites showed significant associations in the fully adjusted model. Higher risk was observed for phenylalanine and multiple cholesterol-enriched VLDL ratios, selected LDL lipid ratios, and very large HDL measures; lower risk was seen for triglyceride-enriched VLDL ratios, specific LDL and HDL subclasses, and albumin. Findings were robust in competing-risk models, after excluding the first three years of follow-up, and across lipid-lowering strata; sex interactions were not evident.

For external context, associations aligned more closely with a European-dominant consortium (directional concordance ~81% of available comparisons) than with China Kadoorie Biobank (directional concordance ~30%), underscoring potential population and ethnic differences and the value of harmonised covariate adjustment. Overall, the chapter identifies a focused set of lipid subclass and amino acid biomarkers, together with albumin, that are associated with ischaemic stroke independent of established risk factors. These findings guide instrument selection and hypotheses for the Mendelian randomisation analyses in **Chapter 6.**.

4.1 Background

Stroke is a leading cause of death globally, and ischaemic stroke accounts for about two-thirds of all strokes in most populations.[1,2] Dyslipidaemia plays an important role in the pathogenesis of ischaemic stroke, and there are recommended lipid-lowering therapies available for preventing ischaemic stroke.[140,141] However, the relevance of other metabolites remains unclear. Human blood consists of a great number of chemically diverse low molecular weight molecules, the metabolome, which can provide several opportunities in the discovery of causal biomarkers for ischaemic stroke. Multiple metabolomic studies have explored the mechanisms of ischaemic stroke, but most of them were limited by their cross-sectional study design. Furthermore, previous studies that investigated associations of ischaemic stroke with NMR metabolites have shown inconsistent findings.[46,142] The objective of this chapter is to conduct a comprehensive analysis of circulating NMR metabolites and incident ischaemic stroke in a single prospective population-based setting involving 3,002 incident ischaemic stroke events among 184,564 participants of UKB. Subsequently, I compare the findings with the two largest relevant published studies to date: a nested case-control study within the China Kadoorie Biobank (CKB) and a consortium of 7 prospective cohort studies on metabolomics and stroke. The study design for this chapter, as well as **Chapter 5.**, which assesses the predictive ability of the metabolomics for the risk of ischaemic stroke over and above established risk factors is shown in **Figure 4.1**.

4.2 Methods

The analyses of UKB excluded participants with missing data for fasting time, missing data for any metabolic biomarker, outlying values of metabolic biomarkers, or

history of CVD (**Figure 4.1**). All data were analysed using R version 4.1.2 (R Foundation for Statistical Computing, Vienna, Austria).

4.2.1 Summary of baseline characteristics

The descriptive characteristics in the total sample and by ischaemic stroke status at the end of follow-up were assessed using the "CBCgrps" package[143] and shown as median and the interquartile range (IQR) for continuous variables and proportions for categorical variables.

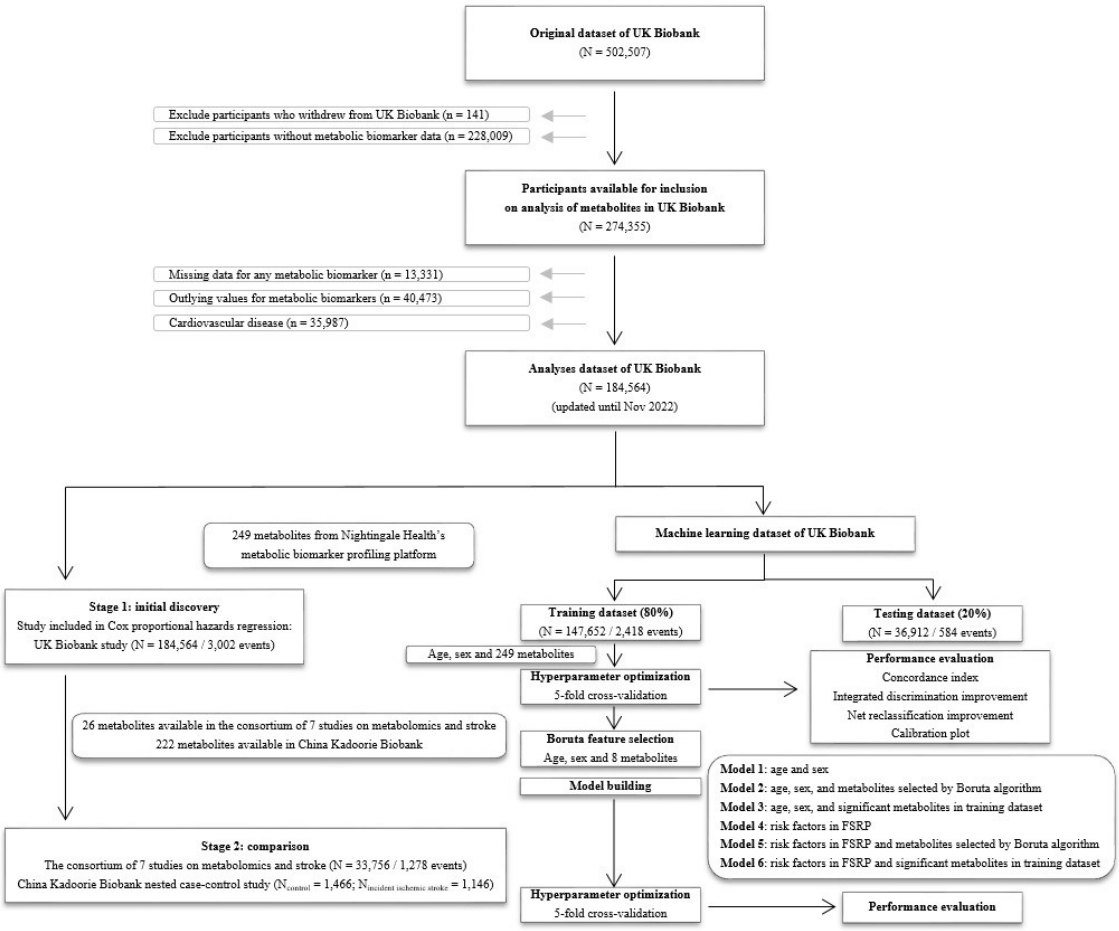


Figure 4.1 A flowchart describing the general framework of the epidemiological analysis (current Chapter) and the prediction modelling (Chapter 5.)
Metabolic levels outside four standard deviations from the mean were considered as outliers. Participants with missing data for any biomarker, outlying values for metabolic biomarkers and cardiovascular disease at baseline were sequentially excluded in the analyses dataset. Risk factors in FSRP include age, sex, current smoking, prevalent diabetes mellitus, blood pressure medication, and systolic blood pressure. XGBoost, extreme gradient boosting model; FSRP, Framingham Stroke Risk Profile.

4.2.2 Associations of NMR metabolites with incident ischaemic stroke

Cox proportional hazards regression models were used to assess the association of each metabolite with incident ischaemic stroke in UKB (Stage 1 in **Figure 4.1**). Using a sequence of increasingly adjusted models allows the evaluation of how different groups of potential confounders influence the observed associations, helping to distinguish the independent relationship between each metabolite and ischaemic stroke from those that may be explained by sociodemographic, lifestyle, physiological, or clinical factors. I used seven models to adjust for putative confounders, each additionally including potential confounders that could be categorised in the same groups (e.g. sociodemographic, study-specific, lifestyle and physiological risk factors as well as pre-existing medical conditions; **Table 3.1**).[2,144] In model 1, I adjusted for sociodemographic factors and study-specific factors that can substantially influence metabolite levels i.e. age, sex, ethnicity, lipid modifying medication, fasting time, spectrometers, and assessment centres. Age, sex, and ethnicity are strong non-modifiable risk factors for stroke and metabolite levels.[109,110] Fasting time, spectrometer differences, and assessment centres are technical variations in sample collection and measurement.[55] Lipid modifying medication was initially adjusted for because most of the metabolites are lipids and such medication might substantially affect their levels. I further explored this in the sensitivity analyses by stratifying participants based on the lipid modifying medication status. Then I included the confounders of model 1 but sequentially further adjusted for groups of lifestyle factors, anthropometry, and physiological factors or diseases and specifically smoking status (model 2), alcohol intake (model 3), dietary supplement use and 13 dietary variables (model 4), physical activity (model 5), BMI (model 6), prevalent diabetes and SBP (model 7). Smoking is a major modifiable lifestyle risk factor for ischaemic stroke[113] and also associated with

metabolite profiles (**Chapter 3..7.4.4.2**). It's conceptually more upstream (i.e. a behavioural factor) and should be adjusted before biological or clinical measurements (e.g. BMI or SBP). Like smoking, alcohol consumption is a lifestyle behaviour and associated with both ischaemic stroke risk[113] and metabolic profiles (**Chapter 3..7.4.4.3**). Smoking and alcohol are often correlated but still distinct behaviours, so alcohol is added separately after smoking. Diet is another modifiable exposure and associated with stroke risk[114] and metabolic levels (**Chapter 3..7.4.4.4**). Dietary intake is typically assessed with more complexity and less accuracy than smoking or alcohol consumption, so it is adjusted for after these simpler behaviours. Adjusting at this stage refines the model for nutrition-driven metabolic effects before adjusting for more clinical or physiological factors. Physical activity is a key independent behaviour but logically follows diet because dietary choices and physical activity patterns often cluster. BMI is a physiological measure that is downstream of lifestyle behaviours (e.g. smoking, alcohol intake, diet, and physical activity). BMI is influenced by these lifestyle behaviours and itself is associated with ischaemic stroke risk[111] and metabolic profiles (**Chapter 3..7.4.4.4**). Prevalent diabetes mellitus[112,114] and hypertension[114] are pre-existing clinical conditions and more proximal risks of ischaemic stroke. SBP, rather than hypertension, was chosen as a covariate because it was the most important modifiable risk factor and there was strong evidence of a linear association between SBP and ischaemic stroke, including within both normotensive and hypertensive ranges.[12] By adjusting for diabetes and SBP last, I can understand how much of the association is independent of these powerful clinical confounders, which might also be mediators (or both). These confounders were considered because they had been found to be correlated with metabolic levels and identified as strong risk factors for ischaemic stroke.[10-12] In addition, I used quintiles of levels of individual metabolites to evaluate the

nonlinear relationship between the outcome and the metabolites in the fully adjusted model (model 7). As I conducted a hypothesis free analysis of the individual metabolites, the false discovery rate (FDR) correction with the Benjamini-Hochberg method was used to adjust for multiple testing.[145] The proportional hazards assumption was tested using an estimate of the time-dependent coefficient with its p value from the "survival" package and was satisfied in the Cox models for each metabolite (FDR-corrected $p > 0.05$).[146] Variance inflation factors were used to diagnose any multi-collinearity of the variables in the fully adjusted model, and none of them showed collinearity (all variance inflation factors < 2).

From model 1 to model 7, the percentage change in the beta coefficients for each metabolite was calculated to assess the extent to which adjustment for additional groups of confounders attenuated or amplified the association with incident ischaemic stroke. To ensure that these changes were interpreted only for robust associations, percentage changes were reported only when the FDR for the corresponding association was less than 0.001. This stringent threshold was chosen to minimise the likelihood of false-positive findings arising from multiple testing across a large number of metabolites and to focus interpretation on associations with strong statistical evidence. Reporting changes in beta coefficients only for metabolites meeting this threshold avoids overemphasising random variation in weaker associations, thereby improving the reliability and biological plausibility of the conclusions.

To examine potential effect modification by sex, I also tested multiplicative interaction terms between sex and each metabolite in the fully adjusted model (model 7), which included all covariates described above. This analysis was conducted for all 249 metabolites, and FDR-corrected p values were calculated for each interaction term.

4.2.3 Secondary analyses

In sensitivity analyses, I fitted the proportional subdistribution hazards model, also known as the Fine-Gray model,[147] to estimate the risk of incident ischaemic stroke over time while treating all-cause mortality as a competing risk in competing risk regression using the R package "cmprsk."[148] The effect of covariates on the subdistribution of all-cause mortality in a competing risks setting was assessed in the model. In addition, the effect of excluding the first 3 years of follow-up (n = 1,375 events) was evaluated to adjust for existing pre-clinical vascular pathology that may have changed the metabolomics spectrum in the blood causing possible reverse causation bias. Furthermore, I performed analyses stratified by lipid modifying agents (Yes/No) and ethnicity (White/Non-white) to explore any potential effect modification. Sensitivity analysis was also conducted by further adjusting for another potential confounder, blood pressure medication.

To compare the effect of a single metabolite on ischaemic stroke between two groups (e.g. the effect from Cox regression vs. the effect from competing risks regression; the group with full follow-up time vs. the group excluding the first 3 years of follow-up; the white ethnic group vs. the others; lipid modifying agent nonusers vs. users) in sensitivity analyses, I used a linear mixed effects model to evaluate the association between the beta coefficients for the metabolites. The effective number of independent metabolic biomarkers was 67 using the method described by Li & Ji (**Supplementary Table 3.1**) which estimated the effective number of independent variables based on the eigenvalue variance of the correlation matrix.[149] This method controlled for type I error in multiple testing scenarios. To further reflect this structure in linear mixed effects model, the 67 clusters from the Euclidean distance matrix and the hierarchical cluster analysis[150] were incorporated into the linear mixed effects model

as random intercepts to constrain the inflation for p value of the slope by reason of high correlation between the metabolites.[151] Although there are other methods for variable reduction and feature selection, such as sparse regression methods (e.g. elastic net) and principal component analysis, the goals of these approaches are mainly for dimensionality reduction for prediction, visualising samples, and detecting dominant patterns of metabolic levels.[152,153] The Li & Ji method was selected specifically to balance statistical power and control for false positive errors in the context of multiple testing with highly correlated metabolites.[149] I will use the 67 independent metabolic biomarkers as random intercepts in linear mixed effects models to evaluate differences in metabolic effects across different groups.

4.2.4 Comparison of the UK Biobank findings with other studies

To assess the consistency of associations of UKB with studies in other populations, additional analyses were undertaken to adjust models for the same set of covariates in other key studies for comparability of results.

The CKB is a prospective cohort of 512,891 Chinese adults 30 to 79 years of age at recruitment, enrolled during 2004 to 2008 in 10 geographically defined regions (5 rural, 5 urban), as previously described (**Chapter 2..1.2.3**).[154] A subgroup of 4662 participants was selected for the metabolomics study from a larger nested case-control study of coronary heart disease and ischaemic stroke subtypes consisting of 23,000 CKB participants, which included standard clinical biochemistry analysis and genome-wide array data.[46] Limited adjustment (i.e. age, sex, fasting time, assessment centres, smoking status, and education) was performed for the association of ischaemic stroke with each metabolite in the main analyses of CKB,[46] and the additional adjustment for BMI and SBP, which was performed in subsequent analyses in CKB, was used in the comparison with the results of this thesis.

Another study is a meta-analysis of seven prospective cohort studies[47] based on individual-participant data with 1278 incident ischaemic stroke events among 38,797 individuals (**Chapter 2..1.2.3**), including the Rotterdam Study (N_{total} = 2565, N_{cases} = 183, mean follow-up time = 6 years), Whitehall II study (N_{total} = 5989, N_{cases} = 126, mean follow-up time = 13 years), FINRISK study 1997 (FINRISK97, N_{total} = 6858, N_{cases} = 405, mean follow-up time = 15 years), Dietary, Lifestyle and Genetic determinants of Obesity and Metabolic syndrome (DILGOM, N_{total} = 4531, N_{cases} = 84, mean follow-up time = 7 years), PROspective Study of Pravastatin in the Elderly at Risk (PROSPER, N_{total} = 4824, mean follow-up time = 2 years), Estonian biobank (N_{total} = 10,576, N_{cases} = 261, mean follow-up time = 7 years), and Framingham Heart Study (N_{total} = 3454, N_{cases} = 219, mean follow-up time = 15 years). Age, sex, BMI, lipid modifying medication, ethnicity, smoking status, prevalent diabetes, and prevalent hypertension were considered confounding factors and were adjusted for in the analyses.

A metabolite is considered replicated if the association with ischaemic stroke in the UKB is significant after multiple testing correction (FDR-corrected $p < 0.05$), and the association in other studies is nominally significant.

4.3 Results

4.3.1 Baseline characteristics

This cohort study included 184,564 participants with a median age of 57 years (IQR 49-63; **Table 4.1**), of whom 45.1% were male and 95.3% self-identified as having White ethnicity. At baseline, 23.5% had hypertension and 3.8% had diabetes mellitus. The majority reported moderate (42%) to high (39%) physical activity levels, while 56.7% were never smokers and 7.3% were non-drinkers. Dietary intakes showed that over half consumed dietary supplements (50.2%) and approximately one-third ate at

least three servings of fresh fruit (35.8%) and vegetables (30.8%) daily. During 2,499,861 person-years of follow-up (median follow-up = 13.8 years, IQR 13.1-14.5), 12,593 participants (6.8%) died and 3002 participants developed ischaemic stroke. Those who experienced ischaemic stroke were significantly older compared to those who did not (63, IQR 58-66 vs. 57, 49-62 years respectively; $p < 0.001$) with a higher proportion of male participants (58.2% vs. 44.9%; $p < 0.001$). These participants were also more likely to be smokers at baseline, have higher BMI (median 27.4 kg/m^2, IQR 24.9-30.4 vs. 26.6, 24.1-29.7; $p < 0.001$), higher SBP (145 mmHg, IQR 132.5-158.5 vs. 136, 124.5-149), higher TG (1.6 mmol/l, IQR 1.2-2.2 vs. 1.5, 1.1-2.1; $p < 0.001$; from clinical chemistry measurement), and lower HDL cholesterol (1.3 mmol/l, IQR 1.1-1.6 vs. 1.4, 1.2-1.7; $p < 0.001$; from clinical chemistry measurement) than healthy subjects. Diabetes mellitus (8.3% vs. 3.8%) and diagnosed hypertension (39% vs. 23.3%) at baseline were more prevalent among participants with incident ischaemic stroke than those without.

Table 4.1 Baseline characteristics of UKB participants included in the epidemiological analyses

	All subjects	Incident ischaemic stroke events		p value*
		No	Yes	
Sample size	**(n = 184,564)**	**(n = 181,562)**	**(n = 3,002)**	
Survival time (ischaemic stroke), median (Q1-Q3)	13.7 (13-14.4)	13.7 (13-14.4)	9.3 (5.9-11.8)	< 0.001
All-cause mortality, n (%)	12593 (6.8)	11794 (6.5)	799 (26.6)	< 0.001
Survival time (all-cause mortality), median (Q1-Q3)	13.8 (13.1-14.5)	13.8 (13.1-14.5)	13.6 (12.6-14.4)	< 0.001
Age, median (Q1-Q3)	57 (49-63)	57 (49-62)	63 (58-66)	< 0.001
Male, n (%)	83236 (45.1)	81490 (44.9)	1746 (58.2)	< 0.001
Ethnicity, n (%)				0.005
White	175904 (95.3)	173004 (95.3)	2900 (96.6)	
Asian or Asian British	3214 (1.7)	3167 (1.7)	47 (1.6)	
Black or Black British	2319 (1.3)	2288 (1.3)	31 (1)	
Mixed	1022 (0.6)	1014 (0.6)	8 (0.3)	
Chinese	564 (0.3)	559 (0.3)	5 (0.2)	
Other ethnic group	1541 (0.8)	1530 (0.8)	11 (0.4)	
Lipid modifying medication, n (%)	26135 (14.2)	25476 (14)	659 (22)	< 0.001
Fasting time, hours, median (Q1-Q3)	3 (2-4)	3 (2-4)	3 (3-4)	0.007
Smoking status, n (%)				< 0.001
Never	104634 (56.7)	103169 (56.8)	1465 (48.8)	
Previous	61262 (33.2)	60159 (33.1)	1103 (36.7)	
Current	18668 (10.1)	18234 (10)	434 (14.5)	
Alcohol intake, n (%)				< 0.001
Daily or almost daily	35711 (19.3)	35076 (19.3)	635 (21.2)	
Three or four times a week	43999 (23.8)	43319 (23.9)	680 (22.7)	
Once or twice a week	49710 (26.9)	48952 (27)	758 (25.2)	
One to three times a month	21146 (11.5)	20867 (11.5)	279 (9.3)	
Special occasions only	20494 (11.1)	20119 (11.1)	375 (12.5)	
Non-drinker	13504 (7.3)	13229 (7.3)	275 (9.2)	
Dietary supplements, n (%)	92574 (50.2)	90978 (50.1)	1596 (53.2)	< 0.001
Fresh fruit, n (%)				< 0.001
< 1 servings/day	16779 (9.1)	16446 (9.1)	333 (11.1)	
≥ 1 and < 2 servings/day	48926 (26.5)	48102 (26.5)	824 (27.4)	
≥ 2 and < 3 servings/day	52697 (28.6)	51839 (28.6)	858 (28.6)	
≥ 3 servings/day	66162 (35.8)	65175 (35.9)	987 (32.9)	

	All subjects	Incident ischaemic stroke events		p value*
		No	Yes	
Whole grain foods, n (%)				< 0.001
0 serve/day	51732 (28)	50761 (28)	971 (32.3)	
> 0 and < 1 serve/day	72328 (39.2)	71270 (39.3)	1058 (35.2)	
≥ 1 and < 3 serves/day	60335 (32.7)	59364 (32.7)	971 (32.3)	
≥ 3 serves/day	169 (0.1)	167 (0.1)	2 (0.1)	
Refined grain foods, n (%)				< 0.001
0 time/day	3021 (1.6)	2983 (1.6)	38 (1.3)	
> 0 and < 1 serve/day	41022 (22.2)	40470 (22.3)	552 (18.4)	
≥ 1 and < 3 serves/day	109868 (59.5)	108068 (59.5)	1800 (60)	
≥ 3 serves/day	30653 (16.6)	30041 (16.5)	612 (20.4)	
Caffeinated coffee, n (%)				0.391
0 cup/day	67830 (36.8)	66758 (36.8)	1072 (35.7)	
> 0 and < 2 cups/day	41691 (22.6)	41012 (22.6)	679 (22.6)	
≥ 2 and < 4 cups/day	46872 (25.4)	46110 (25.4)	762 (25.4)	
≥ 4 cups/day	28171 (15.3)	27682 (15.2)	489 (16.3)	
Decaffeinated coffee, n (%)				0.342
0 cup/day	156381 (84.7)	153844 (84.7)	2537 (84.5)	
> 0 and < 3 cups/day	17079 (9.3)	16812 (9.3)	267 (8.9)	
≥ 3 cups/day	11104 (6)	10906 (6)	198 (6.6)	
Tea, n (%)				0.34
0 cup/day	26556 (14.4)	26123 (14.4)	433 (14.4)	
> 0 and < 3 cups/day	46926 (25.4)	46205 (25.4)	721 (24)	
≥ 3 and < 5 cups/day	54274 (29.4)	53368 (29.4)	906 (30.2)	
≥ 5 cups/day	56808 (30.8)	55866 (30.8)	942 (31.4)	
Vegetables servings, n (%)				0.166
< 1 serving/day	11176 (6.1)	10969 (6)	207 (6.9)	
≥ 1 and < 2 servings/day	53858 (29.2)	53009 (29.2)	849 (28.3)	
≥ 2 and < 3 servings/day	62710 (34)	61707 (34)	1003 (33.4)	
≥ 3 servings/day	56820 (30.8)	55877 (30.8)	943 (31.4)	
Cheese, n (%)				< 0.001
< 1 time/week	36124 (19.6)	35498 (19.6)	626 (20.9)	
≥ 1 and < 2 times/week	39941 (21.6)	39214 (21.6)	727 (24.2)	
≥ 2 and < 5 times/week	84305 (45.7)	83008 (45.7)	1297 (43.2)	

	All subjects	Incident ischaemic stroke events		p value*
		No	Yes	
≥ 5 times/week	24194 (13.1)	23842 (13.1)	352 (11.7)	
Unprocessed red meat, n (%)				< 0.001
0 time/week	12267 (6.6)	12127 (6.7)	140 (4.7)	
< 1 time/week	6206 (3.4)	6106 (3.4)	100 (3.3)	
≥ 1 and < 2 times/week	15075 (8.2)	14871 (8.2)	204 (6.8)	
≥ 2 times/week	151016 (81.8)	148458 (81.8)	2558 (85.2)	
Processed meat, n (%)				< 0.001
0 time/week	16508 (8.9)	16293 (9)	215 (7.2)	
< 1 time/week	55760 (30.2)	54917 (30.2)	843 (28.1)	
≥ 1 and < 2 times/week	54535 (29.5)	53613 (29.5)	922 (30.7)	
≥ 2 times/week	57761 (31.3)	56739 (31.3)	1022 (34)	
Oily fish, n (%)				0.004
0 time/week	20250 (11)	19905 (11)	345 (11.5)	
< 1 time/week	62929 (34.1)	62000 (34.1)	929 (30.9)	
≥ 1 and < 2 times/week	70213 (38)	69019 (38)	1194 (39.8)	
≥ 2 times/week	31172 (16.9)	30638 (16.9)	534 (17.8)	
Non-oily fish, n (%)				0.002
0 time/week	8372 (4.5)	8235 (4.5)	137 (4.6)	
< 1 time/week	54055 (29.3)	53248 (29.3)	807 (26.9)	
≥ 1 and < 2 times/week	92025 (49.9)	90425 (49.8)	1600 (53.3)	
≥ 2 times/week	30112 (16.3)	29654 (16.3)	458 (15.3)	
Spread type, n (%)				< 0.001
Never/rarely use spread	19694 (10.7)	19446 (10.7)	248 (8.3)	
Butter/spreadable butter	67811 (36.7)	66662 (36.7)	1149 (38.3)	
Other type of spread/margarine	97059 (52.6)	95454 (52.6)	1605 (53.5)	
Physical activity base on International Physical Activity Questionnaire, n (%)				0.316
Low	34974 (18.9)	34397 (18.9)	577 (19.2)	
Moderate	77597 (42)	76375 (42.1)	1222 (40.7)	
High	71993 (39)	70790 (39)	1203 (40.1)	
BMI, kg/m2, median (Q1-Q3)	26.6 (24.1-29.7)	26.6 (24.1-29.6)	27.3 (24.9-30.4)	< 0.001
Baseline diabetes mellitus, n (%)	7058 (3.8)	6809 (3.8)	249 (8.3)	< 0.001
Systolic blood pressure, mmHg, median (Q1-Q3)	136 (124.5-149)	136 (124.5-149)	145 (132.5-158.5)	< 0.001

	All subjects	Incident ischaemic stroke events		p value*
		No	Yes	
Diastolic blood pressure, mmHg, median (Q1-Q3)	82 (75.5-89)	82 (75.5-89)	85 (78-91.5)	< 0.001
Baseline hypertension, n (%)	43432 (23.5)	42260 (23.3)	1172 (39)	< 0.001
Blood pressure medication, n (%)	30034 (16.3)	29131 (16)	903 (30.1)	< 0.001
Triglycerides, mmol/l, median (Q1-Q3)	1.5 (1.1-2.1)	1.5 (1.1-2.1)	1.6 (1.2-2.2)	< 0.001
LDL cholesterol, mmol/l, median (Q1-Q3)	3.6 (3.1-4.2)	3.6 (3.1-4.2)	3.6 (3.1-4.2)	0.104
HDL cholesterol, mmol/l, median (Q1-Q3)	1.4 (1.2-1.7)	1.4 (1.2-1.7)	1.3 (1.1-1.6)	< 0.001

*Chi-squared test was used for categorical variables. Student's t-test or Wilcoxon rank-sum (Mann-Whitney) test were used for normally or non-normally distributed continuous variables. BMI, body mass index; LDL, low-density lipoprotein; HDL, high-density lipoprotein.

4.3.2 Associations of the NMR metabolites with incident ischaemic stroke

The associations of incident ischaemic stroke with quintiles of levels of individual metabolites in the fully adjusted model (model 7) were used to evaluate the shape of the relationship between the outcome and metabolic biomarkers (**Supplementary Figure 4.1**). This analysis showed little evidence for nonlinear relationships between each metabolite and incident ischaemic stroke, so the continuous variable of each metabolite was applied from model 1 to model 7.

After adjusting for age, sex, ethnicity, lipid modifying medication, fasting time, spectrometers, and assessment centres in model 1, 84 (33.7%, **Table 4.2**) of the 249 metabolites were significantly associated with the incidence of ischaemic stroke (FDR-corrected $p < 0.05$; **Supplementary Figure 4.2** and **Supplementary Table 4.1**). After further adjusting for smoking status in model 2, associations with incident ischaemic stroke remained significant for 68 (27.3%) metabolites, including 27 positive and 41 inverse associations. The largest percentage changes in effect estimates from model 1 to model 2 were observed for ratio of docosahexaenoic acid (DHA) to total fatty acids (23%), DHA (22%), degree of unsaturation (22%), and triglycerides to total lipids ratio in large LDL (-22%). Additionally adjusting for alcohol intake in model 3 yielded 48 (19.3%) metabolites significantly related to incident ischaemic stroke (22 positive and 26 inverse associations; **Supplementary Table 4.2**). The largest percentage changes in effect estimates from model 2 to model 3 were observed for cholesterol in medium HDL (26%), cholesteryl esters in medium HDL (25%), and HDL cholesterol (25%). Further adjustment for dietary supplement use and 13 dietary variables (model 4) and physical activity (model 5, **Supplementary Table 4.3)** resulted in 6 metabolites being related to incident ischaemic stroke, including acute-phase reaction markers glycoprotein acetyls (HR 1.11, 95% CI 1.06-1.15 in model 4 and 1.11, 1.07-1.15 in model 5), and with the

same effect sizes for both models aromatic amino acid phenylalanine (1.07, 1.03-1.11), ratio of PUFA to total FA (0.94, 0.91-0.97), ratio of PUFA to MUFA (0.94, 0.90-0.97), ratio of LA to total FA (0.93, 0.90-0.97), and fluid balance-related metabolite albumin (0.87, 0.84-0.91). The largest percentage changes in effect estimates from model 3 to model 4 were observed for ratio of monounsaturated fatty acids to total fatty acids (-28%), ratio of polyunsaturated fatty acids to total fatty acids (27%), degree of unsaturation (27%), and ratio of polyunsaturated fatty acids to monounsaturated fatty acids (27%).

Table 4.2 Overall association results between metabolites and incident ischaemic stroke

Model (Covariate adjustment)	Significant metabolites (n)	Significant metabolites (%)	Direction (+/-)	Largest positive association	Largest inverse association
Model 1 (Age, sex, ethnicity, lipid modifying medication, fasting time, spectrometers, and assessment centres)	84	33.7	35 / 49	Glycoprotein acetyls 1.16 (1.11-1.2)	Albumin 0.85 (0.82-0.88)
Model 2 (Model 1 + smoking status)	68	27.3	27 / 41	Glycoprotein acetyls 1.13 (1.09-1.18)	Albumin 0.86 (0.83-0.89)
Model 3 (Model 2 + alcohol intake)	48	19.3	22 / 26	Glycoprotein acetyls 1.12 (1.08-1.17)	Albumin 0.87 (0.84-0.9)
Model 4 (Model 3 + dietary supplement use and 13 dietary variables)	6	2.4	2 / 4	Glycoprotein acetyls 1.11 (1.06-1.15)	Albumin 0.87 (0.84-0.91)
Model 5 (Model 4 + physical activity)	6	2.4	2 / 4	Glycoprotein acetyls 1.11 (1.07-1.15)	Albumin 0.87 (0.84-0.91)
Model 6 (Model 5 + body mass index)	1	0.4	0 / 1	-	Albumin 0.89 (0.86-0.92)
Model 7 (Model 6 + prevalent diabetes mellitus and hypertension)	40	16.1	22 / 18	Free cholesterol in very large HDL 1.09 (1.04-1.14)	Albumin 0.87 (0.84-0.91)

Hazard ratios with 95% confidence intervals are shown for the strongest positive and inverse associations.

After additional adjustment for BMI (model 6), only the inverse association of the fluid balance-related metabolite albumin with incident ischaemic stroke remained significant (0.89, 0.86-0.92). The greatest percentage changes in effect estimates from model 5 to model 6 were found for glycoprotein acetyls (-35%) and albumin (12%).

In the fully adjusted model further controlling for prevalent diabetes and SBP (model 7; **Table 4.3**, **Supplementary Table 4.4**, and **Supplementary Figure 4.3)**, 40 metabolites were significantly associated with incident ischaemic stroke, which indicated that prevalent diabetes and SBP were important negative confounding factors.[155] These comprised 22 VLDLs, 3 LDLs, 13 HDLs, aromatic amino acid phenylalanine, and fluid balance-related albumin. Twenty-two metabolites (i.e. 15 VLDLs, phospholipids to total lipids ratio in small LDL, 5 HDLs, and phenylalanine) were associated with higher risk of ischaemic stroke, with adjusted HRs per 1 SD higher increment ranging from 1.05 (95% CI 1.02-1.10) for phenylalanine to 1.09 (1.04-1.14) for free cholesterol in very large HDL; and 18 metabolites (i.e. 7 VLDLs, 2 LDLs, 8 HDLs, and albumin) were associated with lower risk of ischaemic stroke, with adjusted HRs per 1 SD increment ranging from 0.87 (0.84-0.91) for albumin to 0.95 (0.91-0.99) for TG to total lipids ratio in small VLDL. For TG concentrations within VLDLs, there were inverse associations with incident ischaemic stroke for TG to total lipids ratio in very large VLDL (0.94, 0.91-0.98), medium VLDL (0.94, 0.90-0.98), small VLDL (0.95, 0.91-0.99), and very small VLDL (0.94, 0.91-0.98). The average diameter for VLDL particles was also inversely related to incident ischaemic stroke. In contrast to TG, cholesterol to total lipids ratio and cholesteryl esters to total lipids ratio in all VLDL particles, other than large VLDL, were positively associated with incident ischaemic stroke. For lipoprotein subclasses of HDL, very large HDLs (i.e. total lipids, free cholesterol, and cholesterol in very large HDL) were positively associated with

incident ischaemic stroke, whereas medium HDLs (i.e. total lipids and phospholipids in medium HDL) and small HDLs (i.e. total lipids, phospholipids, cholesterol, and cholesteryl esters in small HDL; concentration of small HDL particles) were inversely associated with incident ischaemic stroke. The greatest percentage changes in effect estimates from model 6 to model 7 were found for albumin (-14%). There was no evidence that the association of each metabolite with incident ischaemic stroke was modified by sex (all FDR-corrected p-interaction > 0.05). There were no marked differences in the metabolic associations among lipid modifying agent users compared with lipid modifying agent non-users (n = 26,135 vs. 158,429; **Supplementary Table 4.5** and **Supplementary Figure 4.4**).

Table 4.3 Association of significant individual metabolites with incident ischaemic stroke in model 7 (n = 184,564)

Metabolite		Beta	SE	HR (95% CI)	p value	FDR
Amino acids						
Phe	Phenylalanine	0.053	0.019	1.05 (1.02-1.1)	5.84E-03	4.59E-02
VLDLs						
XXL_VLDL_C_pct	Cholesterol to total lipids ratio in chylomicrons and extremely large VLDL	0.056	0.020	1.06 (1.02-1.1)	5.14E-03	4.41E-02
XXL_VLDL_CE_pct	Cholesteryl esters to total lipids ratio in chylomicrons and extremely large VLDL	0.068	0.020	1.07 (1.03-1.11)	9.00E-04	1.48E-02
XL_VLDL_C_pct	Cholesterol to total lipids ratio in very large VLDL	0.062	0.021	1.06 (1.02-1.11)	2.93E-03	3.32E-02
XL_VLDL_CE_pct	Cholesteryl esters to total lipids ratio in very large VLDL	0.063	0.021	1.07 (1.02-1.11)	2.38E-03	2.83E-02
M_VLDL_PL_pct	Phospholipids to total lipids ratio in medium VLDL	0.069	0.020	1.07 (1.03-1.11)	6.41E-04	1.23E-02
M_VLDL_C_pct	Cholesterol to total lipids ratio in medium VLDL	0.069	0.020	1.07 (1.03-1.11)	6.20E-04	1.23E-02
M_VLDL_CE_pct	Cholesteryl esters to total lipids ratio in medium VLDL	0.071	0.020	1.07 (1.03-1.12)	4.03E-04	1.19E-02
M_VLDL_FC_pct	Free cholesterol to total lipids ratio in medium VLDL	0.068	0.020	1.07 (1.03-1.11)	7.75E-04	1.38E-02
S_VLDL_C_pct	Cholesterol to total lipids ratio in small VLDL	0.063	0.020	1.07 (1.03-1.11)	1.32E-03	1.79E-02
S_VLDL_CE_pct	Cholesteryl esters to total lipids ratio in small VLDL	0.064	0.019	1.07 (1.03-1.11)	9.48E-04	1.48E-02
S_VLDL_FC_pct	Free cholesterol to total lipids ratio in small VLDL	0.054	0.020	1.06 (1.02-1.1)	7.38E-03	4.59E-02
XS_VLDL_C	Cholesterol in very small VLDL	0.054	0.020	1.06 (1.02-1.1)	7.36E-03	4.59E-02
XS_VLDL_C_pct	Cholesterol to total lipids ratio in very small VLDL	0.055	0.020	1.06 (1.02-1.1)	6.15E-03	4.59E-02
XS_VLDL_CE	Cholesteryl esters in very small VLDL	0.058	0.020	1.06 (1.02-1.1)	4.47E-03	4.04E-02
XS_VLDL_CE_pct	Cholesteryl esters to total lipids ratio in very small VLDL	0.055	0.020	1.06 (1.02-1.1)	7.23E-03	4.59E-02
VLDL_size	Average diameter for VLDL particles	-0.078	0.021	0.93 (0.89-0.96)	1.50E-04	9.36E-03
XXL_VLDL_L	Total lipids in chylomicrons and extremely large VLDL	-0.055	0.020	0.95 (0.91-0.98)	6.33E-03	4.59E-02
XXL_VLDL_FC	Free cholesterol in chylomicrons and extremely large VLDL	-0.060	0.021	0.94 (0.9-0.98)	3.31E-03	3.58E-02
XL_VLDL_TG_pct	Triglycerides to total lipids ratio in very large VLDL	-0.059	0.021	0.94 (0.91-0.98)	4.49E-03	4.04E-02
M_VLDL_TG_pct	Triglycerides to total lipids ratio in medium VLDL	-0.064	0.021	0.94 (0.9-0.98)	1.84E-03	2.29E-02
S_VLDL_TG_pct	Triglycerides to total lipids ratio in small VLDL	-0.054	0.020	0.95 (0.91-0.99)	6.78E-03	4.59E-02
XS_VLDL_TG_pct	Triglycerides to total lipids ratio in very small VLDL	-0.059	0.021	0.94 (0.91-0.98)	4.00E-03	3.99E-02

Metabolite		Beta	SE	HR (95% CI)	p value	FDR
LDLs						
S_LDL_PL_pct	Phospholipids to total lipids ratio in small LDL	0.070	0.020	1.07 (1.03-1.12)	4.79E-04	1.19E-02
M_LDL_CE_pct	Cholesteryl esters to total lipids ratio in medium LDL	-0.056	0.020	0.95 (0.91-0.98)	4.54E-03	4.04E-02
S_LDL_CE_pct	Cholesteryl esters to total lipids ratio in small LDL	-0.063	0.020	0.94 (0.9-0.98)	1.36E-03	1.79E-02
HDLs						
XL_HDL_L	Total lipids in very large HDL	0.059	0.022	1.06 (1.02-1.11)	7.34E-03	4.59E-02
XL_HDL_FC	Free cholesterol in very large HDL	0.085	0.021	1.09 (1.04-1.14)	6.63E-05	8.25E-03
XL_HDL_C	Cholesterol in very large HDL	0.064	0.022	1.07 (1.02-1.11)	3.71E-03	3.85E-02
L_HDL_FC_pct	Free cholesterol to total lipids ratio in large HDL	0.072	0.020	1.08 (1.03-1.12)	3.34E-04	1.19E-02
S_HDL_FC_pct	Free cholesterol to total lipids ratio in small HDL	0.077	0.020	1.08 (1.04-1.12)	1.31E-04	9.36E-03
HDL_P	Concentration of HDL particles	-0.058	0.021	0.94 (0.91-0.98)	6.21E-03	4.59E-02
M_HDL_L	Total lipids in medium HDL	-0.057	0.021	0.95 (0.91-0.99)	7.22E-03	4.59E-02
M_HDL_PL	Phospholipids in medium HDL	-0.058	0.021	0.94 (0.91-0.98)	5.51E-03	4.57E-02
S_HDL_P	Concentration of small HDL particles	-0.069	0.020	0.93 (0.9-0.97)	5.95E-04	1.23E-02
S_HDL_L	Total lipids in small HDL	-0.074	0.020	0.93 (0.89-0.97)	2.62E-04	1.09E-02
S_HDL_PL	Phospholipids in small HDL	-0.075	0.020	0.93 (0.89-0.97)	2.22E-04	1.09E-02
S_HDL_C	Cholesterol in small HDL	-0.066	0.020	0.94 (0.9-0.97)	1.03E-03	1.50E-02
S_HDL_CE	Cholesteryl esters in small HDL	-0.070	0.020	0.93 (0.9-0.97)	4.69E-04	1.19E-02
Other metabolites						
Albumin	Albumin	-0.136	0.019	0.87 (0.84-0.91)	4.21E-13	1.05E-10

SE, standard error. HR, hazard ratio. CI, confidence intervals. FDR, false discovery rate.

4.3.3 Secondary analyses

When treating all-cause mortality as a competing risk, the association of individual metabolites with incident ischaemic stroke were not materially altered (**Supplementary Table 4.6** and **Supplementary Figure 4.5**). Moreover, metabolic associations remained mostly unaltered after excluding the first three years of follow-up (n_{events} = 1,375), indicating that they were not driven by clinically nascent disease at baseline (**Supplementary Table 4.7** and **Supplementary Figure 4.6**). When the analyses were stratified by ethnicity, the metabolic associations with incident ischaemic stroke were statistically significantly different between the White ethnic group and the non-White (n = 175,904 vs. 8,660; **Supplementary Figure 4.7**), but since 95% of the participants were White, there were no marked differences between analyses of all participants and those restricted just to White participants (n = 184,564 vs. 175,904; **Supplementary Table 4.8** and **Supplementary Figure 4.8**). Further adjustment for another potential confounder, blood pressure medication, did not materially change the main associations (**Supplementary Table 4.9** and **Supplementary Figure 4.9**).

4.3.4 Comparison of the UK Biobank findings with other studies

In the external and previously published analysis of CKB,[46] 1 of 225 metabolites was associated with incident ischaemic stroke (acetoacetate [mmol/L], OR 1.31, 1.16-1.47, FDR-corrected $p < 0.05$). When adjusting for the same covariates as done in CKB, 25 metabolites were significantly associated with incident ischaemic stroke in UKB (**Figure 4.2** and **Supplementary Table 4.10**). Of the 25 metabolites, 21 were available in CKB, and of these, albumin (4.8%; OR 0.906, 0.822-0.999) was nominally significant and none was significant when adjusting for multiple testing. Furthermore, for 67 of 222 (30.2%) available metabolites, the directions of associations were consistent between UKB and CKB.

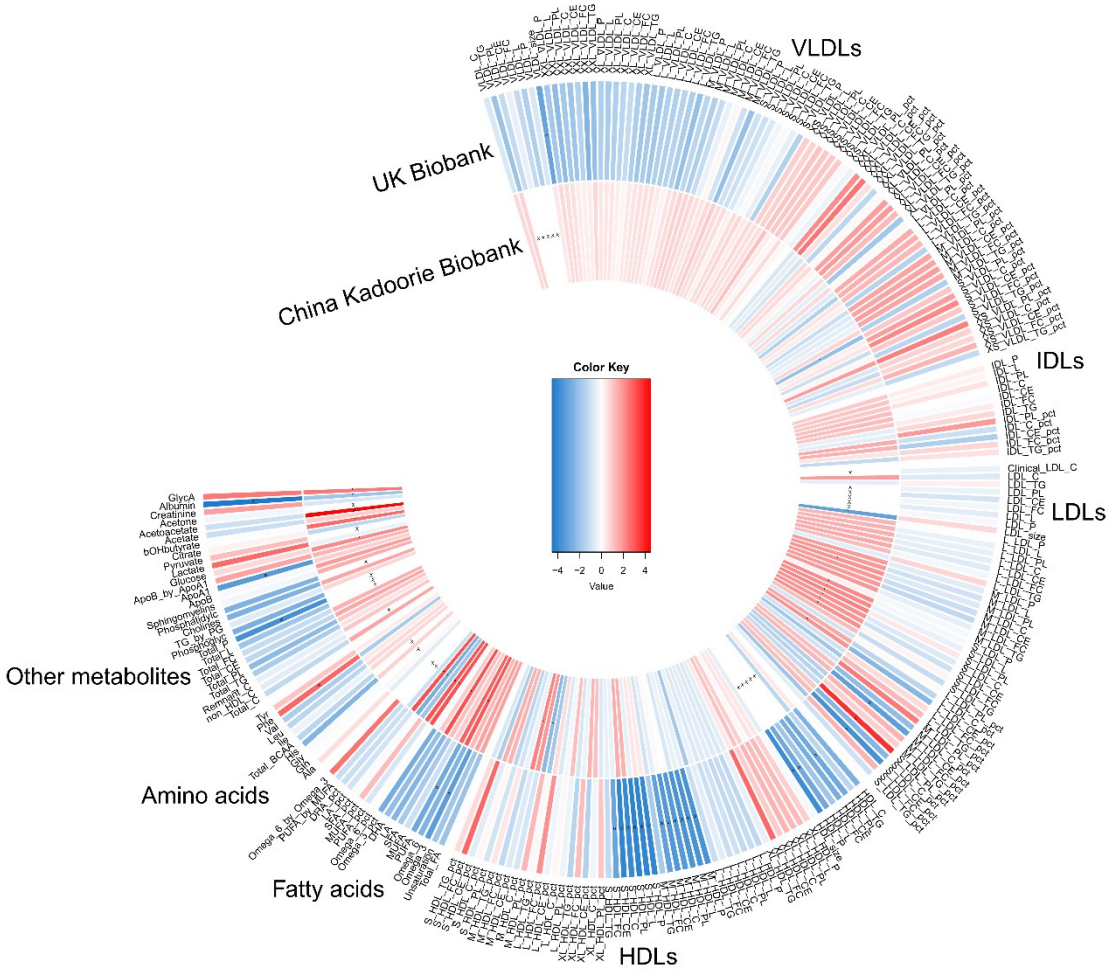


Figure 4.2 Comparison of significant associations of incident ischaemic stroke and metabolites between UK Biobank and China Kadoorie Biobank

The intensity of colours denote the Wald statistic values per one standard deviation higher metabolic levels on the natural log scale between incident ischaemic stroke and individual metabolites from Cox regression in UK Biobank and logistic regression in China Kadoorie Biobank, adjusting for age, sex, fasting time, assessment centres, smoking status, education, body mass index, and systolic blood pressure. Red: positive association; blue: negative association. Based on the findings of UK Biobank, stars (*) and circles (o) are applied to indicate nominal significance and significance after multiple testing correction. Crosses (x) denote unavailability. Abbreviations are in line with **Supplementary Table 3.1**.

In the analysis of the consortium of 7 studies on metabolomics and stroke,[47] 7 of 147 metabolites were associated with incident ischaemic stroke (i.e. phenylalanine, pyruvate, glycoprotein acetyls, histidine, cholesterol in HDL2, cholesterol in large HDL, and free cholesterol in large HDL; Bonferroni-corrected $p < 0.05/30$ independent metabolites $= 1.7 \times 10^{-3}$). When adjusting for the same covariates as the consortium, 57 metabolites were significantly associated with incident ischaemic stroke in UKB (**Figure 4.3** and **Supplementary Table 4.10**). Two out of 57 were available in the published paper, and glycoprotein acetyls (HR 1.13, 1.06-1.20) was significant when adjusting for multiple testing. Of note is that 21 of 26 (80.8%) available associations were directionally consistent between UKB and the consortium.

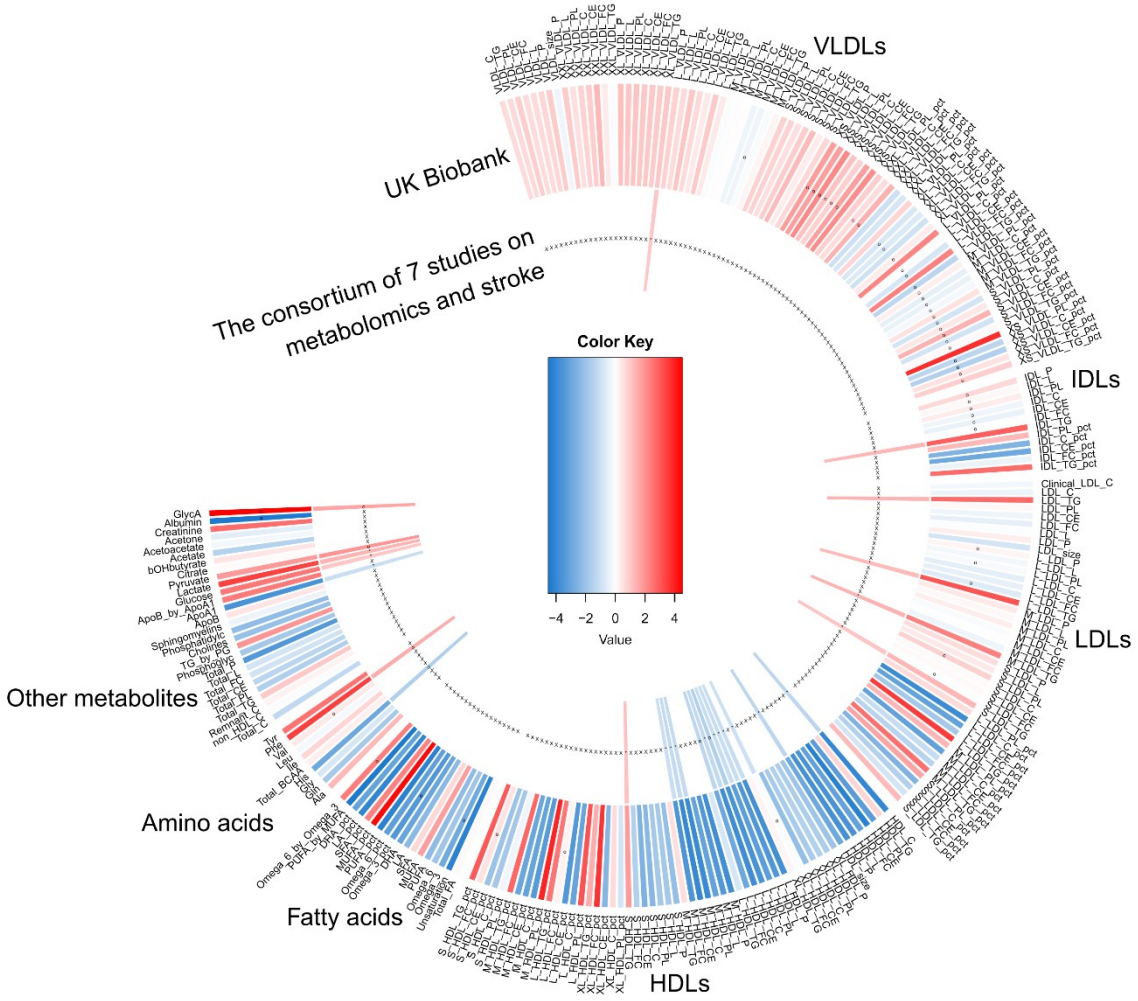


Figure 4.3 Comparison of significant associations of incident ischaemic stroke and metabolites between UK Biobank and the consortium of 7 studies on metabolomics and stroke

The intensity of colours denote the Wald statistic values per one standard deviation higher metabolic levels on the natural log scale between incident ischaemic stroke and individual metabolites from the Cox proportional hazards regression, adjusting for age, sex, body mass index, lipid modifying medication, ethnicity, smoking status, prevalent diabetes, and prevalent hypertension. Red: positive association; blue: negative association. Stars (*) and circles (o) indicate significant associations and significance after multiple testing correction. Crosses (x) denote unavailability. Abbreviations are in line with **Supplementary Table 3.1**.

4.4 Discussion

4.4.1 Association of the metabolites with incident ischaemic stroke

After accounting for multiple testing, incident ischaemic stroke was positively associated with 22 metabolites and inversely associated with 18 metabolites in UKB. These were phenylalanine, 15 VLDLs, phospholipids to total lipids ratio in small LDL, and 5 HDLs, which were associated with increased risk of ischaemic stroke, and 7 VLDLs, 2 LDLs, 8 HDLs, and albumin, which were associated with lower risk of ischaemic stroke. These results were independent of conventional risk factors, such as smoking, alcohol intake, dietary consumption, BMI, high blood pressure, and diabetes.

The strongest association was found between albumin and risk of ischaemic stroke. A 1-SD higher concentration of albumin was associated with 12.3% lower risk of ischaemic stroke, demonstrating that the reduced serum albumin level was an indicator of incident ischaemic stroke, even after controlling for BMI and multiple cardiovascular risk factors.[156] This finding is in line with the Northern Manhattan Study (N_{total} = 2986, N_{cases} = 271, median follow-up time = 12 years), which found that serum albumin levels of 27 to 42 g/L were associated with increased risk of ischaemic stroke (HR 1.67, 95% CI 1.21-2.29) compared with levels of 46 to 55 g/L. Results from the Copenhagen General Population Study[157] (N_{total} = 100,520, N_{cases} = 8247, median follow-up time = 8.5 years) also showed that each 10 g/L lower plasma concentration of albumin was associated with a 46% higher risk of ischaemic stroke (HR 1.46, 95% CI 1.28-1.68). Albumin, a recognised marker of frailty and chronic comorbidity burden,[158] as well as a regulator of colloid osmotic pressure and a transporter of free fatty acids, amino acids, and hormones, can antagonize leukocyte adhesion and stagnation within the postcapillary microcirculation,[159] promote the effect of thrombolysis,[160] and facilitate blood flow perfusion of the brain.[161] It has been hypothesized that

urinary loss of antithrombin,[162] reduced free protein S levels,[162] and venous thrombotic risk of hypoalbuminemia[163,164] may explain the relationship between stroke and low albumin level.

Aromatic amino acid phenylalanine levels were observed to be positively associated with risk of ischaemic stroke. A 1-SD increase in the concentration of phenylalanine was related to 5.4% higher risk of ischaemic stroke. Sidorov et al.[31] also found that concentration of phenylalanine was elevated in patients post ischaemic stroke relative to healthy controls, suggesting that the phenylalanine metabolism pathway is aberrant in patients with ischaemic stroke. It may result from inflammation[165,166] as well as stroke-related oxidative stress[165,167] with diminished phenylalanine hydroxylase activity. After stroke, the excitatory amino acid was released in the ischaemic tissue and led to neurotoxicity within the brain,[168] and the phenylalanine suppressed the excitatory synaptic transmission by way of presynaptic and postsynaptic mechanisms.[169] It has been hypothesised that the elevated levels of phenylalanine consequently serve as a compensatory adaptation to neutralize excitatory amino acid induced neurotoxicity.[170]

Most metabolic biomarkers in my results were constituted by 14 lipoprotein subclasses and 22 relative lipoprotein lipid concentrations. Four derived measurements of VLDLs were associated with reduced risk of ischaemic stroke, including TG to total lipids ratio in very large, medium, small, and very small VLDL. Si et al. also found the inverse association of TG in largest and very large VLDL with ischaemic stroke in a MR study.[82] It has been hypothesized that individuals are protected from atherogenesis because the considerable TG carried in large lipoproteins (i.e. VLDL) cannot penetrate the arterial wall and thereby avoid the risk of atherosclerosis.[171-173]

For HDLs, inverse associations of risk of ischaemic stroke were observed with 8 directly measured medium and small HDL subclasses (total lipids and phospholipids in medium HDL; total lipids, phospholipids, cholesterol, cholesteryl esters in small HDL and concentration of small

HDL particles). On the other hand, positive associations of incident ischaemic stroke were noted with higher levels of 3 directly measured very large HDL subclasses (total lipids, cholesterol, and free cholesterol in very large HDL). In a consortium of seven studies on metabolomics and stroke, Vojinovic et al. also observed an inverse relationship between cholesterol in medium HDL and risk of ischaemic stroke.[47] The possible mechanisms involve distinct physiochemical antiapoptotic and antioxidant activities from the small dense HDLs, which play a crucial role in the protection of LDL against oxidation.[174-176] Furthermore, concentration of small and medium HDL particles was significantly and inversely correlated with carotid intima-media thickening in a cross-sectional study.[177] In UKB my results demonstrate many associations with relative lipoprotein lipid concentrations, which were not the focus in the previous publications.[46,47] Compared with clinical chemistry measurement, the NMR measurement contributed to novel insights on mechanisms by looking at different particle sizes of lipoproteins. I will evaluate their causal relationships with ischaemic stroke in **Chapter 6.**.

The sensitivity analyses showed that associations of individual metabolites with incident ischaemic stroke differed between the White ethnic group and the other ethnic group in UKB, but there was the unbalanced composition of ethnicities compared with the general population.[178] Because all other ethnic groups combined accounted for less than 5%, the CIs of the regression coefficients were much broader than those of the White ethnic group. The interpretation was restricted by the relatively small sample size of all other ethnic groups combined, and a more heterogeneous ethnic background of the cohort was needed to study the risk of ischaemic stroke with metabolites.

Of note, the effect estimates and the number of significant associations changed across the models with different levels of adjustment. Intriguingly, the number of significant associations was initially decreasing, but in the final model it increased ($N_{Model\ 1}$: 84, $N_{Model\ 2}$: 68, $N_{Model\ 3}$: 48,

$N_{Model\ 4}$: 6, $N_{Model\ 5}$: 6, $N_{Model\ 6}$: 1, and $N_{Model\ 7}$: 40; **Table 4.2**). A potential reason to explain this increase is the negative confounding whereby the product of the associations between the exposure and the outcome, the exposure and the confounder, and the confounder and the outcome is negative. This can result in adjusted estimate of the primary association of incident ischaemic stroke with the metabolites to be further from the null hypothesis than the unadjusted measure.[155] It is also possible that partial adjustment without SBP and diabetes in Model 6 introduced collider bias, particularly if BMI lies downstream of both metabolic exposure and cardiometabolic disease pathways. By additionally adjusting for SBP and diabetes in Model 7, some of this collider structure is reduced, allowing re-emergence of genuine associations. The intention was to account for pre-existing clinical risk factors comprehensively in model 7, rather than isolating the contribution of individual components. Consequently, the specific extent to which the re-emergence of associations in Model 7 was driven by SBP, as opposed to diabetes mellitus, was not formally quantified. However, given that SBP is a continuous variable with greater statistical variability and stronger associations with metabolic profiles than diabetes (Number of significant associations between metabolites and systolic blood pressure was 247, compared with 239 for diabetes mellitus; **Table 3.2**), it is plausible that SBP contributed more substantially to the change in the number of significant associations. In any case, for this analysis I wanted to investigate the associations of the metabolites with ischaemic stroke independent of all these factors that might be important confounders in theory. The sequential adjustment from model 1 to model 7 provided important insights into the extent to which lifestyle factors, adiposity, and cardiometabolic risk factors explain the observed associations of ischaemic stroke with metabolites. The substantial reduction in the number of significant metabolites from 84 in the minimally adjusted model to only 6 after adjusting for diet- and lifestyle-related covariates (models 4 and 5, **Table 4.2**) suggests that part of the observed associations may be attributable to confounding by these factors.

The 40 significant metabolites in the fully adjusted model (model 7), after additionally accounting for BMI, prevalent diabetes, and SBP, highlights the complex role of these clinical risk factors as negative confounders, where adjustment unmasks associations otherwise attenuated in intermediate models. The re-emergence of 40 significant metabolites in the fully adjusted Model 7 highlights the complex interplay of confounding, and potential collider bias, particularly in relation to SBP and diabetes, which may act as negative confounders in these associations. This pattern implies that certain lipid and amino acid metabolites, as well as albumin, are associated with ischaemic stroke risk independently of established vascular risk factors, while others may act through pathways linked to obesity, glycaemic control, or blood pressure regulation. By comparing hazard ratios across models, it is also evident that some metabolites, particularly albumin, phenylalanine, and specific VLDLs and HDLs, exhibit robust associations that persist despite extensive covariate adjustment, strengthening the case for their potential causal or mediating role in stroke pathophysiology.

4.4.2 Comparison with other studies

Although there were more than 50 metabolites related to ischaemic stroke with limited adjustment (i.e. age, sex, fasting time, assessment centres, smoking status, and education) in the main analyses of CKB,[46] additionally adjusting for BMI and SBP in subsequent analyses in CKB (used in the comparison with the results of this thesis) tremendously attenuated almost all significant associations of ischaemic stroke with metabolites, and only acetoacetate remained significant. The evidence on considerable alterations of these partial adjusted findings indicates the value of fully assessing for residual confounding.[179] Of the 21 significant metabolites in UKB, albumin (4.8%) was also nominally significant in the subsequent analyses in CKB if the same confounding factors (i.e. age, sex, fasting time, assessment centres, smoking status, education, BMI, and SBP) were adjusted for. Although further controlling for confounders (e.g. prevalent

diabetes) has been suggested to influence observational estimates,[47] my results demonstrate substantial confounding may remain from alcohol intake and dietary consumption. Despite the issues of residual confounding, the directions of available associations were more concordant between UKB and the consortium of seven studies on metabolomics and stroke (80.8%) than between UKB and CKB (30.2%). Few studies have evaluated the ethnic differences in metabolomics. Vardarajan et al. found distinct metabolic profiles of Alzheimer's disease between Caribbean Hispanics, Non-Hispanic Whites, and African Americans.[180] In addition, differential metabolic pathways of lipid and amino acid were reported between European American and African American patients with bladder cancer.[181] Therefore, ethnic disparities could help to explain more concordant metabolic profiles between UKB and the consortium than those between UKB and CKB.

4.4.3 Strengths and limitations

This study adds several novel insights to the existing literature on metabolomics and ischaemic stroke, including the largest sample size with NMR metabolic biomarkers and prospective population-based design which enabled the analysis of cardiometabolic biomarkers before the onset of ischaemic stroke. Forty metabolic biomarkers were robustly associated with incident ischaemic stroke after adjustment for multiple testing and several confounding factors (**Chapter 4..3.2**). Prior studies either adjusted for limited confounding factors in the associations of ischaemic stroke and metabolites[46,55] or reported unreliable effect estimates (e.g. extremely large, small, or infinite HRs; **Chapter 2..1.2.3**)[41]. While albumin and phenylalanine have been previously found to be associated with ischaemic stroke risk by Guo et al.[56] and Vojinovic et al.[47] respectively, this study provides the largest validation of the 38 novel metabolic biomarkers linking ischaemic stroke to downstream subclasses of VLDL, IDL, and HDL profiles even after full covariate adjustment (**Table 4.3**). Furthermore, the study directly compared its findings with

both the CKB[46] and the consortium of 7 studies on metabolomics and stroke[47] using the same covariates, revealing that metabolic profiles of ischaemic stroke appear to vary by ethnicity and population structure, with greater consistency between UKB and European-dominant cohorts (**Chapter 4..3.4**). This highlights potential ethnic differences in ischaemic stroke-related metabolic pathways, a relatively underexplored area.

Limitations of the study involve the possible systematic underestimation of estimated associations due to within-person variability from measurement error or short-term biological variability (i.e. regression dilution bias).[182,183] In addition, the participants in UKB were mostly from White ethnic groups, recruited on a volunteer basis, and were generally healthier than the wider population,[184,185] which may limit the generalisability of the results. However, the analysis of risk associations may be still generalisable.[186] Finally, the above findings, although robust, came from an observational study which might suffer from residual confounding. Further randomised evidence is needed to ascertain causality.

4.5 Conclusion

The current chapter quantifies the associations of individual metabolites with incident ischaemic stroke from 184,564 participants of UKB. The findings suggest positive associations of phenylalanine, VLDLs, one LDL, and HDLs, and inverse association of VLDLs, LDLs, HDLs, and albumin with risk of ischaemic stroke. Each 1-SD higher observed levels of free cholesterol in very large HDL and of albumin were associated with 9% higher and 13% lower risk of ischaemic stroke, respectively. Compared to the clinical chemistry measurement, the current NMR metabolic biomarkers with detailed lipoprotein subclasses and the largest sample size to date contributed to a more comprehensive understanding of mechanisms for ischaemic stroke, in addition to the previous studies of CKB[46] and the consortium of 7 studies on metabolomics and stroke.[47] These results inform the specific metabolites selected for further MR analyses in **Chapter 6.**.

Chapter 5. Predictive value of NMR metabolites for ischaemic stroke

Chapter summary

This chapter evaluated whether adding 249 nuclear magnetic resonance-derived metabolites improves prediction of incident ischaemic stroke beyond established clinical factors in 184,564 participants from the UK Biobank. Extreme gradient boosting model and random survival forests were used with an 80/20 training-testing dataset split. Performance was assessed using the concordance index, calibration plots, the absolute integrated discrimination improvement, and the continuous net reclassification improvement.

Across the full cohort and in subgroups defined by age at least 50 years, men, and women, metabolites produced little or no gain in discrimination compared with age and sex alone or with the Framingham Stroke Risk Profile. Typical changes in the concordance index were small and not statistically meaningful. Reclassification frequently worsened after adding metabolites, with negative values for both integrated discrimination improvement and continuous net reclassification improvement. Calibration was generally acceptable but showed overestimation at low predicted risks and some underestimation at high predicted risks.

Feature selection highlighted a small set of influential variables, notably age, sex, albumin, glycoprotein acetyls, omega-3 fatty acids, and several lipid ratios. Dependence plots indicated non-linear relationships and potential interactions, but these patterns did not translate into better predictive accuracy. Overall, nuclear magnetic resonance metabolomics added limited incremental value for ischaemic stroke risk prediction relative to standard clinical factors, underscoring the need to test broader biomarker panels and to evaluate performance in more diverse populations.

5.1 Background

Two ways that metabolomics could be used for ischaemic stroke research are looking at aetiological associations of metabolites with ischaemic stroke (i.e. traditional regression methods) and investigating the predictive value of metabolomics for ischaemic stroke risk (i.e. pure prediction algorithms or machine learning).[187-189] The first one can be used to better understand underlying mechanisms related to stroke risk and the second to discover potentially new biomarkers for stroke risk that could have applications in further research or clinical practice. According to Dr. Breiman,[187] the underpinning of aetiological research as shown in **Chapter 4.** involves the formal apparatus of inference (e.g. hypothesis tests, optimal estimators, and CIs). It commences with observed data and perceives the data as being created by a model which employs a vector of metabolic variables as input and gives an outcome variable (i.e. incident ischaemic stroke). The connections between estimation and attribution relied on CIs and p-values, and the discovery of the disease aetiology could be achieved by this traditional one-size-fits-all model (e.g. comparing averages across people who did or did not have an abnormal metabolic level and aiming to leverage the heterogeneity in a defined framework by applying data to discover which person should receive the treatment for disease).[190,191] On the other hand, in predictive modelling making valuable predictions is more important than other considerations such as the assumption of data distributions or probability theory.[189] Machine learning contrasts sharply with the data modeling culture with the fact that it applies the empirical distribution ingrained in the dataset rather than an assumed distribution.[187,189] Furthermore, **Chapter 3.** demonstrates that metabolites are immensely correlated, which causes inflated estimates when simultaneously introducing them in a regression model. Machine learning can address this problem and in addition detect nonlinear effects and interactions of a set of variables on disease

outcome.[64] Machine learning can also be utilised to drive disease prevention and population health by risk stratification.[192] Precise and early recognition of people at high ischaemic stroke risk is critical for reducing the disease burden by providing prompt and specific countermeasures. Two current validated prediction tools for stroke[19] are the Framingham Stroke Risk Profile (FSRP)[21] and QStroke tools.[20] The FSRP was selected as the baseline comparison model over QStroke due to its extensive validation across international cohorts,[22,193] stroke-specific focus, widespread use in both clinical and research settings, and the recommendation by the American Heart Association for physicians to predict individual risk of patients and to outline treatment plans.[22] In particular, comparative validation has shown that FSRP significantly outperformed QStroke for 5-year stroke risk prediction.[19] Furthermore, a tendency to overestimate stroke risk among women at higher levels of predicted risk was reported for QStroke.[20] These findings underscore limitations in QStroke's calibration across sex and risk strata, suggesting that its predictions may be biased in high-risk female populations. In contrast, the revised FSRP demonstrated improved discrimination and calibration across diverse cohorts.[22]

Metabolomic risk models based on NMR biomarkers have been used to predict type 2 diabetes,[24] CVD,[194] and many common diseases.[67] Based on the first release of the NMR data (N~120,000) in UKB, the type 2 diabetes[24] risk prediction model was modestly enhanced by adding NMR metabolites on top of conventional risk factors, but the NMR metabolites did not improve the risk predictions for CVD[194] and overall stroke.[67] In addition, Buergel et al.[67] used only neural network, while it has been shown that extreme gradient boosting model (XGBoost) outperforms neural network in terms of epidemiological data (**Chapter 2..2.3**),[65,69-72] and did not apply 81 derived NMR metabolic measurements (i.e. metabolite ratios) in UKB to predict overall stroke.

Given these prior findings (**Chapter 3..7.4.3**), this chapter also evaluates subgroup-specific prediction models. Subgroup analyses focus on older participants (aged over 50 years), men, and women, since age and sex are major non-modifiable risk factors for ischaemic stroke and may modify the predictive value of metabolites. Evaluating these subgroups allows assessment of whether metabolomic predictors improve risk stratification differently across demographic strata, which has implications for targeted prevention strategies.

Therefore, this chapter aims to evaluate whether incorporating NMR metabolites into the conventional risk factors of FSRP improves the prediction of incident ischaemic stroke, using two types of machine learning approaches and the most recent UKB dataset available at the time of analysis.

5.2 Methods

To investigate the utility of metabolites for improving ischaemic stroke risk prediction, I used machine learning approaches with XGBoost[74] and random survival forests (RSF)[195] to train the model in the training dataset (80% randomly selected from the analyses dataset) and generate predictions in the testing dataset (the remaining 20%; **Figure 4.1**) for validation. For subgroup analyses, the training and testing datasets were randomly selected from the subgroups (i.e. older participants aged over 50, men, and women), reflecting the importance of these demographic factors in stroke risk and prediction. The Shapley additive explanation (SHAP)[196] was used to quantify the contribution of each predictor in XGBoost. More information on these methods is provided in the following sections. Baseline characteristics were compared for the two subsets to ensure that the random selection created indeed random subsets of the data (**Supplementary Table 5.1)**. The analyses complied with the transparent reporting of a

multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statement (**Supplementary Table 5.2**).[197]

All analyses were conducted in R version 4.1.2 (R Foundation for Statistical Computing, Vienna, Austria) on the Wellcome Centre for Human Genetics Biomedical Research Computing (BMRC) Compute Cluster, using parallel computation to support large-scale data processing. Model development, hyperparameter tuning, and survival analysis were implemented using the xgboost,[74] survival,[198] Boruta,[199] and SHAPforxgboost[65] packages. The computation was executed across multiple cores using batch job submission on the BMRC cluster, with runtime logs and outputs automatically stored and version-stamped for reproducibility.

5.2.1 Variable selection and model building

I used age, sex, and 249 metabolites as predictors in XGBoost, and used the Boruta algorithm to select relevant predictors. In the Boruta procedure, the full set of candidate predictors (age, sex, and all 249 metabolites) was entered, and the algorithm identified the subset of variables to be relevant for predicting ischaemic stroke. Age, sex, and circulating metabolites were used as the primary predictors to evaluate the incremental predictive value of metabolomics in isolation, without influence from conventional clinical predictors or other potential confounders. Age and sex were retained because they are fundamental demographic determinants of ischaemic stroke risk and are typically included in baseline models in prediction studies. By focusing on age, sex, and metabolites alone, the analysis enabled a clearer evaluation of whether metabolic profiling provides independent predictive information rather than simply replicating the effects of established risk factors. Then six models were built:

1. Age and sex only.

2. Age and sex plus metabolites selected by the Boruta algorithm.
3. Age and sex plus metabolites individually associated with ischaemic stroke.
4. FSRP risk factors (age, sex, smoking status, prevalent diabetes mellitus, blood pressure medication, and SBP) only.
5. FSRP risk factors plus metabolites selected by the Boruta algorithm.
6. FSRP risk factors plus metabolites individually associated with ischaemic stroke.

Metabolites individually associated with ischaemic stroke were those statistically significant (FDR-corrected $p < 0.05$) in Cox regression models developed in the training dataset. Instead of implementing the original FSRP using Cox proportional hazards regression model, I treated the FSRP as a source of clinically validated predictor variables (i.e. age, sex, smoking status, prevalent diabetes mellitus, blood pressure medication, and SBP) and included these variables as inputs in the machine learning survival models. This methodological strategy was based on a fundamental distinction between the modelling framework of the FSRP and the machine learning survival models. The original FSRP models were based on Cox proportional hazards regression, and were designed to estimate over fixed time intervals ranging from 1 to 10 years.[22] By contrast, I modelled ischaemic stroke occurrence and survival time jointly, using a time-to-event framework and evaluated model performance using the concordance index, a standard metric for survival models that captures overall discriminative ability across the follow-up period. Rather than re-implementing the FSRP, the individual FSRP predictors were incorporated as part of a broader, nonlinear, nonparametric machine learning framework to flexibly capture interactions between risk factors and metabolites, account for right-censoring in survival data, and avoid assumptions of proportional hazards and additivity required by Cox proportional

hazards regression model.[64] This approach is also aligned with previous literature suggesting that stroke risk is influenced not only by relationships between biological, demographic, and lifestyle variables which are often inadequately captured by traditional regression models, but also by complex, nonlinear dependencies among these factors.[200]

5.2.2 Extreme gradient boosting model

XGBoost builds a series of small simple prediction models (regression trees) by gradient boosting models. Each consecutive tree is iteratively generated based on the previous tree's error to get a better prediction. Parameters of XGBoost consist of the general parameters (i.e. the type of model used to do boosting, such as low level routines for training, prediction and evaluation), the learning task parameters (i.e. the learning task and the corresponding learning objective), and the booster parameters (i.e. guiding the respective booster at each step). The booster parameters involve (1) maximum tree depth (i.e. the maximum number of layers for regression trees); (2) boosting learning rate, which manages the step size at which the optimizer updates the weights (a higher value leads to faster but less precise updates, while a lower value leads to slower but more precise updates); (3) subsample ratio (i.e. the percentage of observations to be random samples of training instances used to build each tree); (4) percentage of features used for building each tree; (5) minimum sum of instance weight needed in a leaf node (i.e. stop attempting to split once the sample size in a node goes below a given threshold); and (6) number of boosted trees to fit. The range for each assessed hyperparameter is listed in the booster parameter rows of **Supplementary Table 5.3**, **Supplementary Table 5.4**, **Supplementary Table 5.5**, and **Supplementary Table 5.6**. Hyperparameter optimization was achieved in the training dataset by a random search algorithm[201] with 1,000 configurations of hyperparameters in the 5-fold

nested cross-validation (i.e. one-fifth of the data was retained as the validation data for testing the model, and the remaining four-fifths of the data were used as training data). The final configuration was determined as the one reaching the best model performance (i.e. lowest negative log-likelihood). Based on the selected hyperparameters, I subsequently trained the XGBoost model for the maximum number of 5,000 iterations (i.e. boosted regression trees) to determine the best number of iterations, while training would be terminated if the performance did not rise after 20 rounds.[202]

5.2.3 Boruta algorithm

I aimed to identify the smallest group of essential predictors that would enable greater and robust explanatory predictive power not only because parsimonious models with fewer predictors tend to have more predictive ability when applied to new data but also because models with fewer parameters are easier to understand, explain, and allow risk prediction from a smaller metabolites panel. I used the Boruta algorithm,[199] recommended in simulation studies as the most powerful variable selection approach,[203] to detect essential features for prediction. In the wrapper approach,[204] the feature subset selection algorithm is executed as a wrapper around the model training instances (e.g. random forest) and performs a search for a proper subset through the model training instances to evaluate feature subsets. Boruta is a random forest based wrapper algorithm. Importances of real features (e.g. each metabolic variable in UKB) are iteratively compared with importances of shadow features, generated by permuting real ones across participants and therefore eradicating the association with the outcome. In each importance source run, the set of features was doubled by adding a copy of each feature. For each original feature, a statistical test was carried out comparing its importance with the maximum value of all the shadow features. If the most important shadow feature was greater than or equal to the importance of the original feature, the

decision was to remove that feature. Otherwise, features that were significantly better than shadows were deemed important and kept. Thereafter all relevant features were finally chosen by the Boruta algorithm.[199] The maximum number of importance source runs was set to 1000.

5.2.4 Random survival forests

Because random forests usually outperform other tree-based methods,[205] I applied an ensemble tree method, RSF,[195] to analyse the right-censored data and compared the results to those from XGBoost. In the training dataset (80% randomly selected from the analyses dataset in **Figure 4.1**), RSF drew bootstrap samples from two-thirds of the data to form the root of the tree by removing one-third of the data called 'out-of-bag' data. It resembled k-fold cross-validation to restrain over-fitting. Based on randomly selected predictors, RSF grew a binary survival tree beginning at the root node, which was the top of the tree containing all the data. Then the root was recursively split into two daughter nodes, a left and right daughter node, by using the log-rank splitting rule which splits the nodes by maximization of the survival difference between two daughter nodes. This allowed each candidate predictor to affirm its capacity to predict ischaemic stroke. When various participants were sorted, and the number of tree nodes raised with each split, each node in the tree was grouped by participants with homogeneous survival. If each terminal node has at least one ischaemic stroke with unique survival times and the data cannot be further split, the splitting paused at the terminal nodes. I estimated an ensemble hazard function and survival probability by averaging all 1000 trees constructed in RSF, and the model performance was assessed from the testing dataset (20% randomly selected from the analyses dataset in **Figure 4.1**).

5.2.5 Validation methods

In the testing dataset (median follow-up = 13.8 years), the predicted HR scale was assigned to each participant from the cumulative score of each tree. I applied the concordance index (C-index) and its 95% CI, calibration plots, absolute integrated discrimination improvement (IDI), and continuous net reclassification improvement (NRI) to evaluate the model performance.[206] The C-index quantifies the relationship between event times and risk predictions by the values of discrimination between early occurrences, associated with higher-risk participants, and later events.[207] The IDI and NRI quantify the degree of correct reclassification of participants across different risk categories due to new predictors.

C-index was calculated as the proportion of concordant pairs (i.e. where the participant with the longer survival time has the greater probability of survival as predicted by XGBoost).[208,209] C-index allows to distinguish, at a glance, between quasi-random and well-behaved models by summarising three different elements of predictions (time, risk, and event occurrence) in a single value.[207]

IDI is defined as the difference in discrimination slopes, which quantified the separation of negative and positive outcomes, between two models (e.g. one with, and another without, the additional predictors).[206] It is calculated as the difference in the mean predicted risks of ischaemic stroke from the two models among those who did develop ischaemic stroke minus the difference in mean predicted risks in those who did not develop it (i.e. it is the difference between two differences).[24] Positive IDI values imply that the addition of metabolites gives a superior model.

The continuous NRI is derived from the summary measure of reclassification tables by removing risk categories and labelling any decrease in model-based

probability caused by the addition of a new predictor as a downward reclassification and any increase as an upward reclassification.[210,211] It quantifies the appropriateness of the change in predicted probabilities of ischaemic stroke between old and new models.[24] Positive NRI values imply better ischaemic stroke risk reclassification following the addition of metabolites to the risk prediction model.

5.2.6 The Shapley additive explanation

The SHAP, defined by desirable properties from game theory,[212] is an explanation method that attributes to each predictor the change in the expected model prediction when conditioning on that predictor.[65] The SHAP value is the average expected marginal contribution of one predictor after all possible combinations have been considered. SHAP values help to quantify precise and consistent contributions for all of the predictors when each predictor might have contributed more or less than the others.[65,196] In SHAP summary plots, the mean absolute SHAP value across all participants denotes the importance of a predictor in the model. In SHAP dependence plots, positive SHAP values would indicate a positive association with ischaemic stroke, while negative SHAP values would indicate a negative association. Locally estimated scatterplot smoothing (LOESS)[213] with the second degree local polynomial was used to fit values in SHAP dependence plots, and the criterion for automatic smoothing parameter selection is generalised cross-validation.[214]

5.3 Results

5.3.1 Predictive value of metabolites for ischaemic stroke risk in the overall population

Including only age and sex, the C-index was 0.707 (95% CI 0.686-0.728; model 1 in **Table 5.1** and **Supplementary Figure 5.1 A**) in the testing dataset. When age, sex

and 249 metabolites were used to predict ischaemic stroke using XGBoost, the C-index did not significantly increase (0.719, 0.699-0.739; **Table 5.1** and **Supplementary Figure 5.1 A**) in the testing dataset. Ten features were selected as pivotal features for prediction of incident ischaemic stroke by the Boruta algorithm: age, sex, albumin, creatinine, glucose, ratio of PUFA to total FA, ratio of LA to total FA, omega-3 FA, glycoprotein acetyls, and free cholesterol to total lipids ratio in very large HDL. When age, sex and these 8 metabolites were used to predict ischaemic stroke using XGBoost, the C-index did not materially change (0.716, 0.696-0.736; model 2 in **Table 5.1** and **Supplementary Figure 5.1 A**) in the testing dataset. When age, sex and 15 significant metabolites in Cox regression in the training dataset were used to predict ischaemic stroke in XGBoost, the C-index was still similar to the previous models (0.706, 0.686-0.726, model 3; **Table 5.1** and **Supplementary Figure 5.2**) in the testing dataset.

Table 5.1 Prediction of ischaemic stroke and components of each model in the overall population (n = 184,564)

Predictors	Age, sex, and 249 metabolites	Age and sex	Age, sex, and metabolites selected by Boruta algorithm	Age, sex, and significant metabolites in the training dataset	FSRP#	FSRP# and metabolites selected by Boruta algorithm	FSRP# and significant metabolites in the training dataset
Model		**1**	**2**	**3**	**4**	**5**	**6**
Extreme gradient boosting model							
C-index in the training dataset	0.77 (0.76, 0.779)	0.714 (0.704, 0.724)	0.741 (0.732, 0.751)	0.741 (0.731, 0.75)	0.738 (0.728, 0.748)	0.756 (0.747, 0.765)	0.76 (0.751, 0.769)
C-index in the testing dataset	0.719 (0.699, 0.739)	0.707 (0.686, 0.728)	0.716 (0.696, 0.736)	0.706 (0.686, 0.726)	0.726 (0.706, 0.746)	0.732 (0.712, 0.751)	0.728 (0.708, 0.747)
Slope of the calibration line	1.456 (1.033, 1.878)	0.931 (0.650, 1.212)	1.004 (0.631, 1.376)	0.917 (0.567, 1.268)	1.114 (0.680, 1.548)	1.139 (0.788, 1.490)	1.245 (0.959, 1.532)
Absolute integrated discrimination improvement			-0.00257 (-0.00487, -0.00121)	-0.00211 (-0.00405, -0.00102)		0.000396 (-0.00231, 0.00514)	-0.000745 (-0.00391, 0.000215)
Continuous net reclassification improvement			-0.243 (-0.287, -0.188)	-0.231 (-0.286, -0.187)		-0.170 (-0.243, 0.202)	-0.175 (-0.237, 0.186)
Random survival forests							
C-index in the training dataset		0.698 (0.688, 0.708)	0.696 (0.685, 0.706)	0.693 (0.683, 0.703)	0.663 (0.652, 0.673)	0.717 (0.706, 0.727)	0.713 (0.703, 0.723)
C-index in the testing dataset		0.699 (0.678, 0.721)	0.706 (0.686, 0.727)	0.686 (0.664, 0.708)	0.656 (0.634, 0.678)	0.717 (0.697, 0.737)	0.712 (0.691, 0.733)
Slope of the calibration line		0.949 (0.699, 1.199)	0.997 (0.585, 1.410)	1.038 (0.804, 1.273)	1.255 (0.620, 1.890)	0.978 (0.602, 1.353)	0.994 (0.707, 1.282)
Absolute integrated discrimination improvement			-0.00277 (-0.00542, 0.00106)	-0.00239 (-0.00479, -0.000684)		0.00150 (000583, 0.00366)	0.00135 (0.000459, 0.00363)
Continuous net reclassification improvement			-0.230 (-0.283, -0.176)	-0.238 (-0.291, -0.181)		0.280 (0.227, 0.332)	0.287 (0.222, 0.344)
Age	√	√	√	√	√	√	√
Sex (male)	√	√	√	√	√	√	√
Albumin	√	-	√	√	-	√	√
Creatinine	√	-	√	-	-	√	-
Glucose	√	-	√	-	-	√	-
Ratio of polyunsaturated fatty acids to total fatty acids	√	-	√	-	-	√	-
Ratio of linoleic acid to total fatty acids	√	-	√	-	-	√	-
Omega-3 fatty acids	√	-	√	-	-	√	-
Glycoprotein acetyls	√	-	√	-	-	√	-
Free cholesterol to total lipids ratio in very large HDL	√	-	√	-	-	√	-
Systolic blood pressure	-	-	-	-	√	√	√
Blood pressure medication	-	-	-	-	√	√	√
Smoking status	-	-	-	-	√	√	√
Prevalent diabetes mellitus	-	-	-	-	√	√	√

Absolute integrated discrimination improvement (IDI) and continuous net reclassification improvement (NRI) were calculated for selected model comparisons. Specifically, Model 2 and Model 3 were compared against Model 1 (age and sex only), and Model 5 and Model 6 were compared against Model 4 (FSRP risk factors only). These comparisons assess whether adding metabolites improves predictive performance beyond a minimal baseline model and beyond a clinically validated risk score (FSRP). Components in model 2 and 5 were selected from age, sex, and all metabolites using Boruta algorithm. FSRP, Framingham Stroke Risk Profile. C-index, concordance index.
Framingham Stroke Risk Profile: age, sex, smoking status, prevalent diabetes mellitus, blood pressure medication, and systolic blood pressure.

Within the beeswarm summary plot of model 2 (**Figure 5.1**), the SHAP values for age, sex, albumin, glycoprotein acetyls, and omega-3 fatty acids signify the highest contributions to the model prediction. Their means of absolute SHAP values are 0.588, 0.177, 0.131, 0.067, and 0.063, respectively. The beeswarm summary plot of model 3 (Figure 5.2) shows that the age, sex, albumin, free cholesterol in very large HDL, and total lipids in small HDL had the highest contributions to the model prediction. The calibration plots demonstrated good agreement between predicted and observed risk across all models (**Supplementary Figure 5.3 A**), and the slope of the calibration lines ranged from 0.917 (95% CI 0.567-1.268) in model 3 to 1.004 (0.631-1.376) in model 2, indicating satisfactory calibration without systematic overestimation or underestimation (**Table 5.1)**. Nevertheless, closer inspection of the calibration plots reveals some poor calibration, particularly a tendency toward overestimation at the lower range of predicted risk, which should be considered when interpreting the model's predictive accuracy. However, after adding metabolites to age and sex, the classification of disease risk worsened, as shown by estimates of the absolute IDI (-0.00257, 95% CI -0.00487, -0.00121 when adding the Boruta-selected metabolites; -0.00211, 95% CI -0.00405, -0.00102 when adding 15 significant metabolites in Cox regression) and continuous NRI (-0.243, 95% CI -0.287, -0.188; -0.231, 95% CI -0.286, -0.187).

A model including the risk factors from FSRP slightly increased the C-index, but this was not significant (C-index 0.726, 0.706-0.746; **Table 5.1**, **Supplementary Figure 5.4** and **Supplementary Figure 5.5**). The metabolites did not add predictive information over the conventional risk factors (C-index 0.732, 0.712-0.751 when adding the Boruta-selected metabolites, and 0.728, 0.708-0.747 when adding 15 significant metabolites in Cox regression). The models were well calibrated (**Supplementary Figure 5.3 A**), but the reclassification showed no improvement from both IDI and NRI

after adding the metabolites to the conventional risk factors (**Table 5.1**). In addition, the prediction performances from RSF were not significantly different from those from the XGBoost, except for model 4 (C-statistic 0.656, 0.634-0.678 for RSF vs. 0.726, 0.706-0.746 for XGBoost; **Table 5.1** and **Supplementary Figure 5.1 A**).

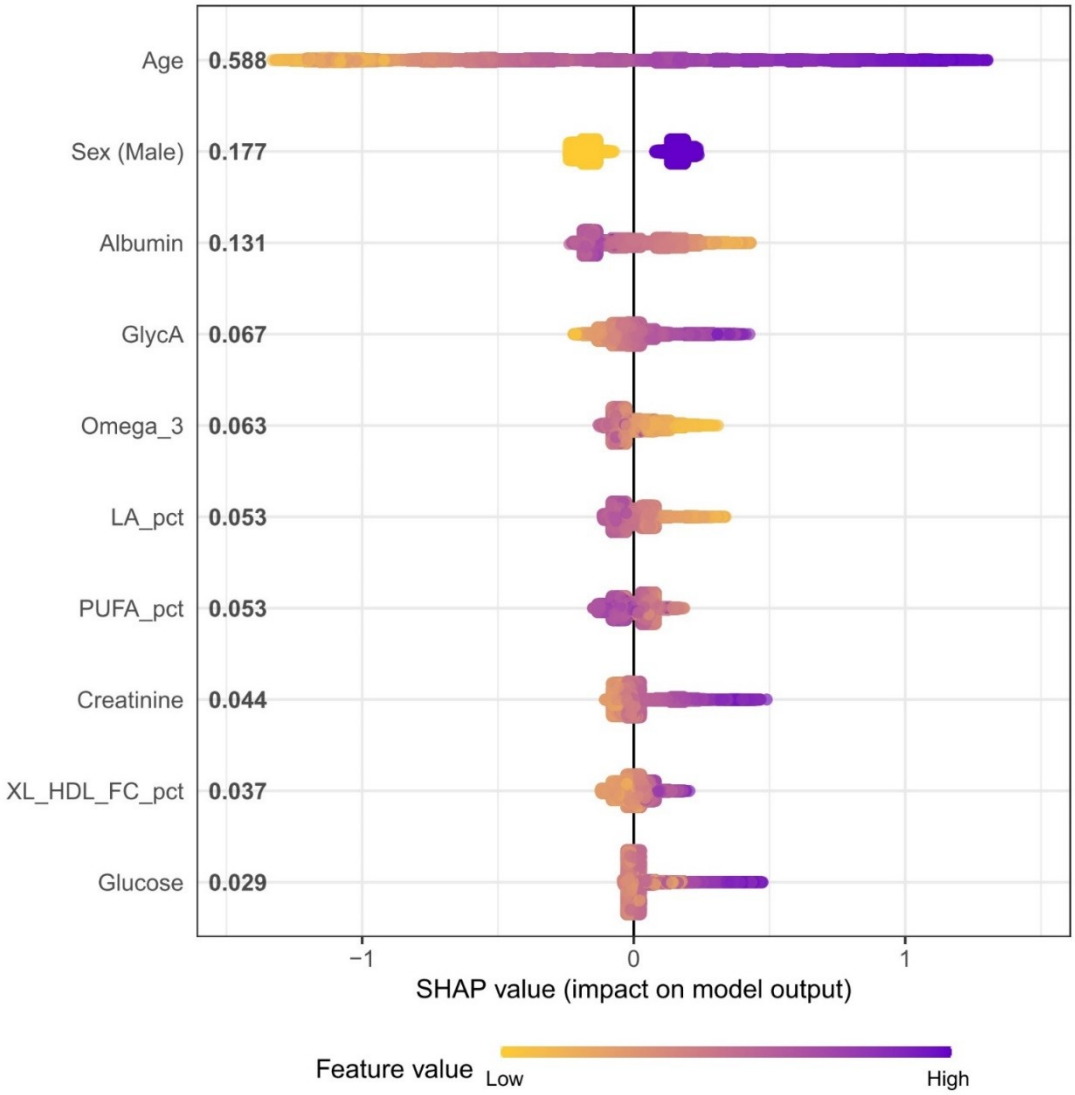


Figure 5.1 The SHAP summary plot of extreme gradient boosting model with age, sex, and metabolites selected by Boruta algorithm

The plot illustrates the global interpretability with how much each feature contributes, either positively or negatively, to the prediction of incident ischaemic stroke. A positive association with ischaemic stroke was depicted as SHAP values above 0, and a negative association with ischaemic stroke was depicted as SHAP values below 0 (i.e., the greater the SHAP value the greater the hazard ratio for ischaemic stroke. It also illustrates the local interpretability with each dot corresponding to a participant in the training dataset (n = 147,652). The dot's position on the x axis indicates the impact that feature has on the model's incident ischaemic stroke prediction for that participant.

When numerous dots are located at the same x position, they are aggregated to reveal density. The bold figures on the left side indicate the mean of absolute SHAP values of the features. Glyc, glycoprotein acetyls. Omega_3, omega-3 fatty acids. LA_pct, ratio of linoleic acid to total fatty acids. PUFA_pct, ratio of polyunsaturated fatty acids to total fatty acids. XL_HDL_FC_pct, free cholesterol to total lipids ratio in very large HDL.

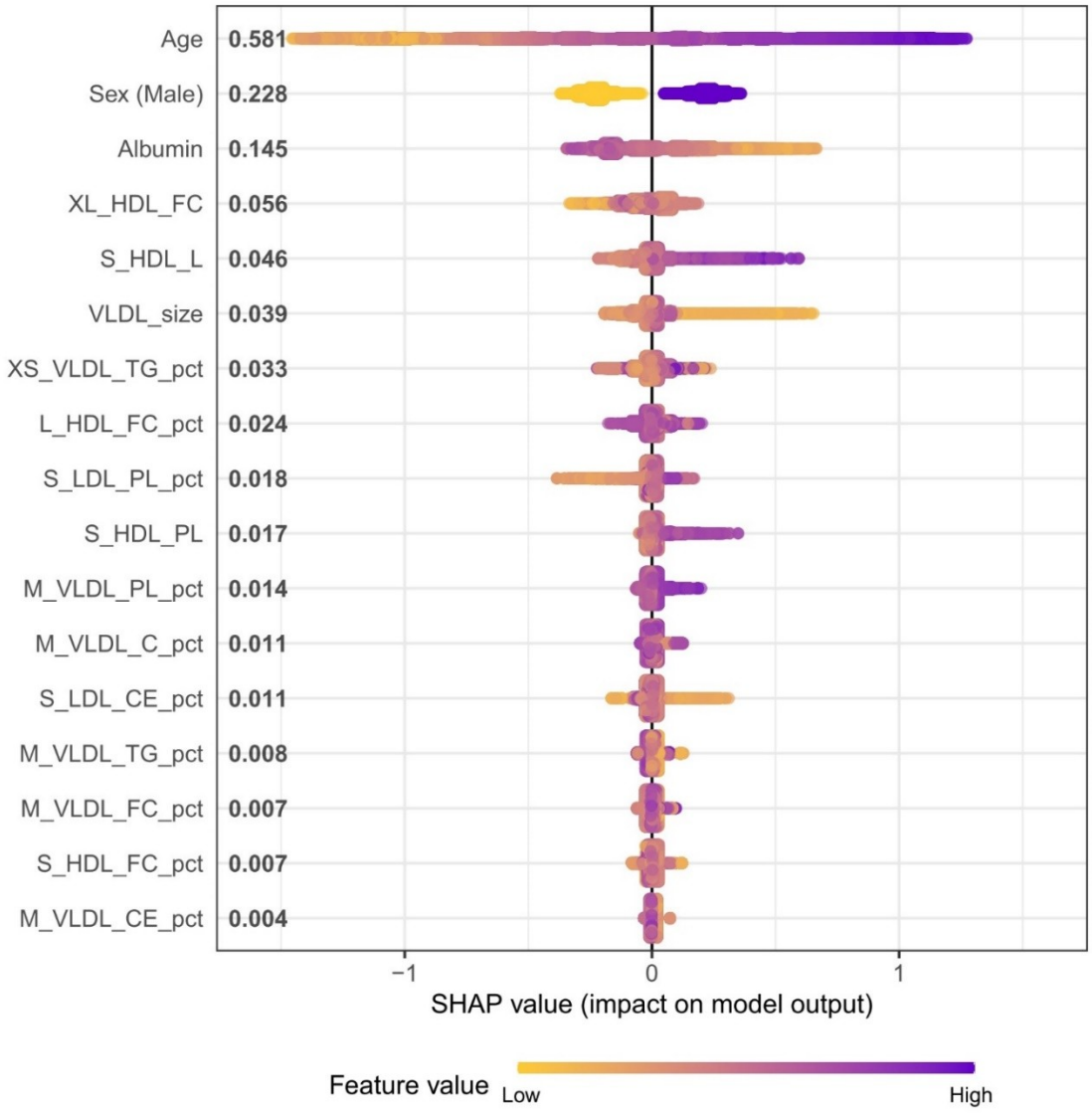
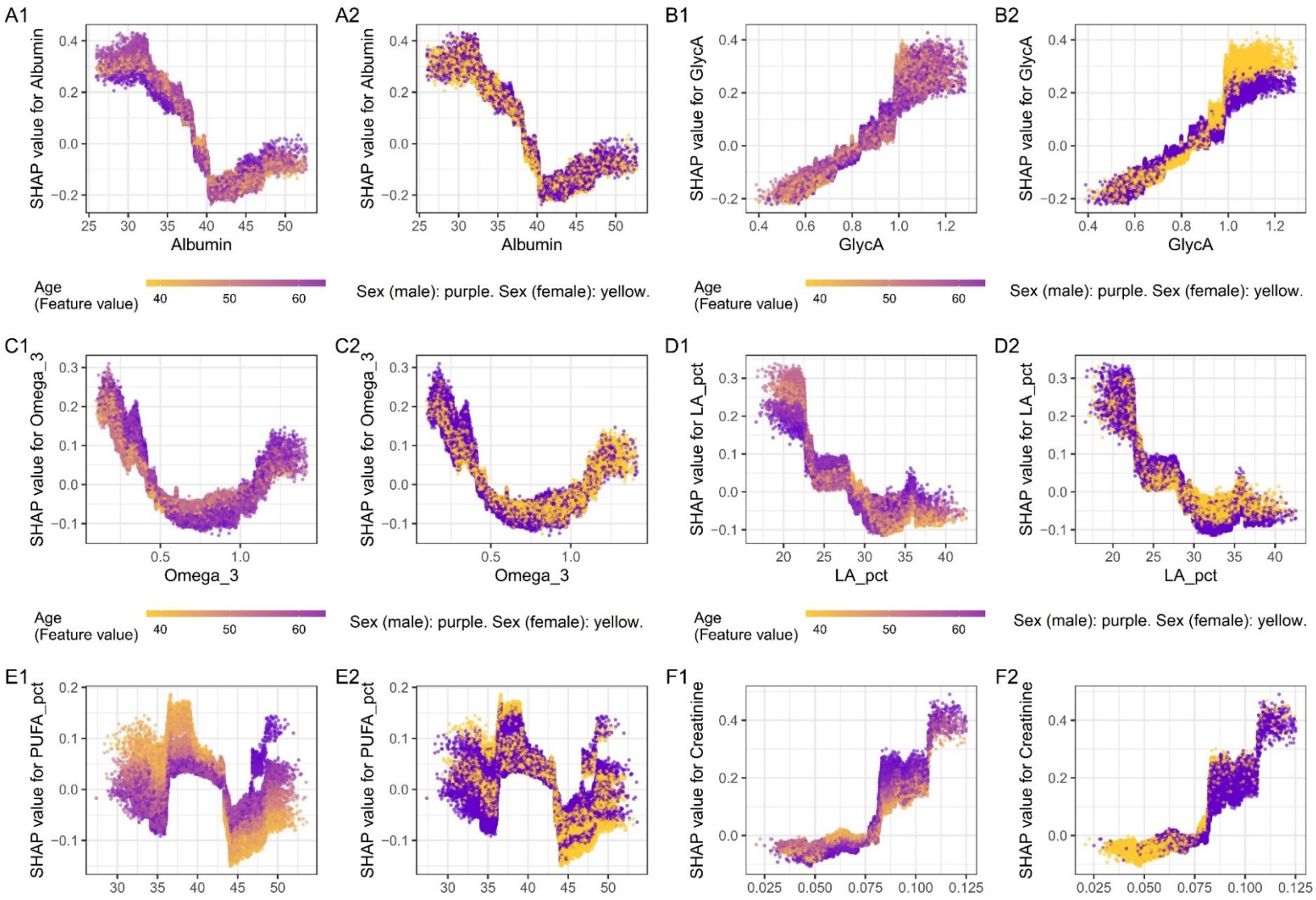


Figure 5.2 The SHAP summary plot of extreme gradient boosting model with age, sex, and significant metabolites in the training dataset.

The plot illustrates the global interpretability with how much each feature contributes, either positively or negatively, to the prediction of incident ischaemic stroke. A positive association with ischaemic stroke was depicted as SHAP values above 0, and a negative association with ischaemic stroke was depicted as SHAP values below 0 (i.e., the greater the SHAP value the greater the hazard ratio for ischaemic stroke. It also illustrates the local interpretability with each dot corresponding to a participant in the training dataset (n = 147,652). The dot's position on the x axis indicates the impact that

feature has on the model's incident ischaemic stroke prediction for that participant. When numerous dots are located at the same x position, they are aggregated to reveal density. The bold figures on the left side indicate the mean of absolute SHAP values of the features. Abbreviations are in line with **Supplementary Table 3.1**.

The XGBoost containing age, sex, and the 8 Boruta-selected metabolites, was used to evaluate non-linear associations and possible interaction effects because they were pivotal features for prediction of ischaemic stroke. The nonlinear relationships between these features and ischaemic stroke were depicted in SHAP dependence plots (**Figure 5.3**). The SHAP dependence plot of albumin versus its SHAP value shows a non-linear association (**Figure 5.3 A1**). Using LOESS to fit the values in the dependence plot, it shows that the occurrence of ischaemic stroke increased at a concentration of albumin of 38.4 g/l; above that value, the occurrence of ischaemic stroke was predicted to decrease among this specific group of participants. The occurrence of ischaemic stroke increased at a concentration of glycoprotein acetyls of 0.834 mmol/l (**Figure 5.3 B1**). Also, the dependence plots for omega-3 fatty acids, ratio of LA to total fatty acids, ratio of polyunsaturated fatty acids to total fatty acids, creatinine, free cholesterol to total lipids ratio in very large HDL, and glucose show non-linear relationships, in which the risk of ischaemic stroke is increased with omega-3 fatty acids below 0.432 mmol/l (**Figure 5.3 C1**), ratio of LA to total fatty acids below 28.6% (**Figure 5.3 D1**), creatinine from 0.0779 mmol/l onwards (**Figure 5.3 F1**), free cholesterol to total lipids ratio in very large HDL from 14.9% (**Figure 5.3 G1**), and glucose from 4.43 mmol/l (**Figure 5.3 H1**). The dependence plot of age versus its SHAP value showed an interaction with sex, which increased the impact of age on males over 50 (**Figure 5.3 I1**). There is some evidence of interaction effects by age (**Supplementary Figure 5.6 A2**, **C2**, **D2**, **E2**, **F2**, **G2**, and **H2**) and sex (**Supplementary Figure 5.6 A4**, **B4**, **C4**, **D4**, **E4**, **F4**, **G4** and **H4**), so I evaluated the model performance in subgroups (i.e. the older group, the male group, and the female group).


A1
SHAP value for Albumin
Albumin
A2
SHAP value for Albumin
Albumin
B1
SHAP value for GlycA
GlycA
B2
SHAP value for GlycA
GlycA
Age
(Feature value)
40
50
60
Sex (male): purple. Sex (female): yellow.
C1
SHAP value for Omega_3
Omega_3
C2
SHAP value for Omega_3
Omega_3
D1
SHAP value for LA_pct
LA_pct
D2
SHAP value for LA_pct
LA_pct
E1
SHAP value for PUFA_pct
E2
SHAP value for PUFA_pct
F1
SHAP value for Creatinine
F2
SHAP value for Creatinine

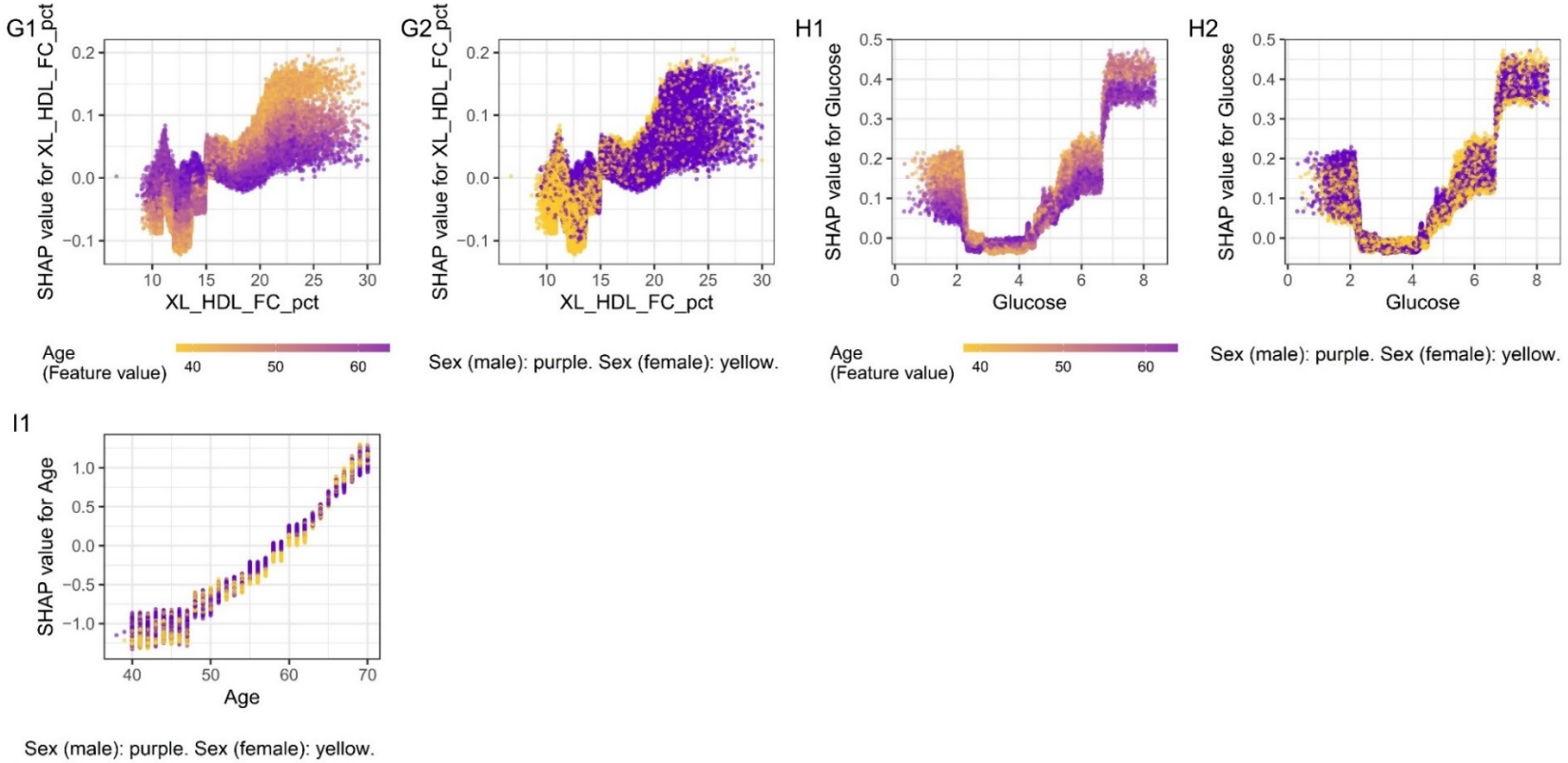


Figure 5.3 SHAP dependence plots were shown for the XGBoost-model 2 with age, sex, and metabolites selected by Boruta algorithm
The SHAP value of a participant was displayed as each point on the y-axis corresponding to the feature on the x-axis. Glyc, glycoprotein acetyls. Omega_3, omega-3 fatty acids. LA_pct, ratio of linoleic acid to total fatty acids. PUFA_pct, ratio of polyunsaturated fatty acids to total fatty acids. XL_HDL_FC_pct, free cholesterol to total lipids ratio in very large HDL.

5.3.2 Predictive value of metabolites for ischaemic stroke risk among those aged ≥ 50 years

Including only age and sex, the C-index was 0.679 (95% CI 0.655-0.702; model 1 in **Table 5.2** and **Supplementary Figure 5.1 B**) in the testing dataset. When age, sex and 249 metabolites were used to predict ischaemic stroke in XGBoost, the C-index did not materially change (0.684, 0.661-0.707; **Table 5.2**) in the testing dataset. Ten features were selected as pivotal features for prediction of incident ischaemic stroke by the Boruta algorithm: age, sex, DHA, ratio of PUFA to total FA, ratio of LA to total FA, glucose, lactate, creatinine, albumin, and glycoprotein acetyls. Adding the eight Boruta-selected metabolites to either age and sex or risk factors of FSRP did not significantly improve the discrimination in the old age group (0.677, 0.654-0.7 when adding Boruta-selected metabolites to age and sex; 0.684, 0.661-0.707 for FSRP; 0.687, 0.664-0.71 when adding Boruta-selected metabolites to FSRP). Calibration plots demonstrated that the models were generally well calibrated in the older age group (**Supplementary Figure 5.3 B**). The slope of the calibration lines ranged from 0.938 (95% CI 0.792-1.084) in model 1 to 1.308 (1.015-1.601) in model 2. This indicates that model 2 tends to underestimate risk at higher predicted probabilities, where observed risk exceeded predicted risk. At lower predicted probabilities, there was some evidence of slight overestimation. Despite this, overall calibration remained acceptable across models, with no evidence of severe miscalibration. These results suggest acceptable calibration with no strong systematic underestimation or overestimation. However, the degrees of correct reclassification were decreased in both absolute IDI (-0.00348, -0.00636 - -0.00188) and continuous NRI (-0.223, -0.272 - -0.167) after adding the Boruta-selected metabolites to the age and sex (**Table 5.2**). In addition, the prediction performances from RSF were not significantly different from those from the XGBoost, except for

model 3 (C-statistic 0.607, 0.583-0.631 for RSF vs. 0.684, 0.661-0.707 for XGBoost; **Table 5.2** and **Supplementary Figure 5.1 B**).

Table 5.2 Prediction of ischaemic stroke and components of each model among those aged ≥ 50 years (n = 137,310)

Predictors	Age, sex, and 249 metabolites	Age and sex	Age, sex, and metabolites selected by Boruta algorithm	FSRP[#]	FSRP[#] and metabolites selected by Boruta algorithm
Model		**1**	**2**	**3**	**4**
Extreme gradient boosting model					
C-index in the training dataset	0.801 (0.791, 0.81)	0.675 (0.663, 0.686)	0.714 (0.704, 0.725)	0.709 (0.699, 0.72)	0.732 (0.722, 0.743)
C-index in the testing dataset	0.684 (0.661, 0.707)	0.679 (0.655, 0.702)	0.677 (0.654, 0.7)	0.684 (0.661, 0.707)	0.687 (0.664, 0.71)
Slope of the calibration line	1.620 (1.279, 1.961)	0.938 (0.792, 1.084)	1.308 (1.015, 1.601)	1.200 (0.773, 1.626)	1.265 (0.873, 1.656)
Absolute integrated discrimination improvement			-0.00348 (-0.00636, -0.00188)		0.0000153 (-0.00018, 0.000228)
Continuous net reclassification improvement			-0.223 (-0.272, -0.167)		-0.00051 (-0.209, 0.240)
Random survival forests					
C-index in the training dataset		0.653 (0.641, 0.664)	0.661 (0.65, 0.673)	0.629 (0.618, 0.641)	0.678 (0.666, 0.689)
C-index in the testing dataset		0.675 (0.651, 0.699)	0.653 (0.629, 0.676)	0.607 (0.583, 0.631)	0.670 (0.646, 0.694)
Slope of the calibration line		1.029 (0.830, 1.228)	1.027 (0.609, 1.446)	0.919 (0.137, 1.701)	1.187 (0.987, 1.387)
Absolute integrated discrimination improvement			-0.00409 (-0.00725, -0.0021)		0.00023 (-0.00042, 0.000698)
Continuous net reclassification improvement			-0.213 (-0.273, -0.155)		0.133 (-0.046, 0.251)
Age	√	√	√	√	√
Sex (male)	√	√	√	√	√
Docosahexaenoic acid	√	-	√	-	√
Ratio of polyunsaturated fatty acids to total fatty acids	√	-	√	-	√
Ratio of linoleic acid to total fatty acids	√	-	√	-	√
Glucose	√	-	√	-	√
Lactate	√	-	√	-	√
Creatinine	√	-	√	-	√
Albumin	√	-	√	-	√
Glycoprotein acetyls	√	-	√	-	√
Systolic blood pressure	-	-	-	√	√
Blood pressure medication	-	-	-	√	√
Smoking status	-	-	-	√	√
Prevalent diabetes mellitus	-	-	-	√	√

Components in model 2 and 4 were selected from age, sex, and all metabolites using Boruta algorithm. FSRP, Framingham Stroke Risk Profile. C-index, concordance index.
Framingham Stroke Risk Profile: age, sex, smoking status, prevalent diabetes mellitus, blood pressure medication, and systolic blood pressure.

5.3.3 Predictive value of metabolites for ischaemic stroke risk by sex

Including only age, the C-indices were 0.661 (95% CI 0.632-0.691; model 1 in **Table 5.3** and **Supplementary Figure 5.1 C**) in men and 0.714 (0.683-0.746; model 1 in **Table 5.4** and **Supplementary Figure 5.1 D**) in women. When age and 249 metabolites were used to predict ischaemic stroke in XGBoost, the C-index did not significantly increase either in men (0.672, 0.645-0.699) or in women (0.717, 0.688-0.747). When adding pivotal features selected by the Boruta algorithm to age or risk factors of FSRP, the C-index did not improve either in men (0.665, 0.637-0.692 when adding Boruta-selected metabolites to age; 0.682, 0.654-0.709 for FSRP; 0.684, 0.657-0.711 when adding Boruta-selected metabolites to FSRP) or women (0.706, 0.675-0.737 when adding Boruta-selected metabolites to age; 0.727, 0.698-0.756 for FSRP; 0.721, 0.692-0.75 when adding Boruta-selected metabolites to FSRP). The prediction models demonstrated acceptable calibration in both sex-stratified groups. In men, the slope of the calibration line ranged from 0.818 (0.531-1.106) in model 1 to 1.503 (1.111-1.896) in model 3 and 1.425 (1.061, 1.789) in model 4 (**Supplementary Figure 5.3 C**). In women, the slopes ranged from 0.878 (0.582–1.174) in model 1 to 1.672 (1.186-2.158) in model 4 (**Supplementary Figure 5.3 D**). These results indicate reasonable alignment between predicted and observed risks in most models, although the slopes significantly greater than 1.0 in model 3 and model 4 for men and model 4 for women suggest a tendency toward slight underestimation of ischaemic stroke risk, particularly at higher predicted values. In addition, the prediction performances from RSF were not significantly different from those from the XGBoost, except for model 3 (C-statistic 0.623, 0.596-0.650) for RSF vs. 0.682, 0.654-0.709 for XGBoost in men; C-statistic 0.662, 0.632-0.693 for RSF vs. 0.727, 0.698-0.756 for XGBoost in women; **Table 5.3**, **Table 5.4**, **Supplementary Figure 5.1 C**, and **Supplementary Figure 5.1 D**).

Table 5.3 Prediction of ischaemic stroke and components of each model in men (n = 83,236)

Predictors	Age and 249 metabolites	Age	Age and metabolites selected by Boruta algorithm	FSRP#	FSRP# and metabolites selected by Boruta algorithm
Model		**1**	**2**	**3**	**4**
Extreme gradient boosting model					
C-index in the training dataset	0.796 (0.781, 0.804)	0.704 (0.69, 0.718)	0.733 (0.721, 0.746)	0.729 (0.716, 0.742)	0.751 (0.739, 0.764)
C-index in the testing dataset	0.672 (0.645, 0.699)	0.661 (0.632, 0.691)	0.665 (0.637, 0.692)	0.682 (0.654, 0.709)	0.684 (0.657, 0.711)
Slope of the calibration line	1.000 (0.517, 1.483)	0.818 (0.531, 1.106)	1.032 (0.709, 1.355)	1.503 (1.111, 1.896)	1.425 (1.061, 1.789)
Absolute integrated discrimination improvement			-0.000249 (-0.00503, -0.000951)		0.0000623 (0.000000494, 0.000133)
Continuous net reclassification improvement			-0.182 (-0.247, -0.122)		0.141 (0.0693, 0.224)
Random survival forests					
C-index in the training dataset		0.674 (0.661, 0.688)	0.675 (0.661, 0.689)	0.649 (0.634, 0.663)	0.702 (0.688, 0.715)
C-index in the testing dataset		0.656 (0.627, 0.686)	0.649 (0.621, 0.676)	0.623 (0.596, 0.650)	0.674 (0.648, 0.701)
Slope of the calibration line		0.864 (0.619, 1.109)	0.873 (0.353, 1.393)	1.080 (-0.483, 2.642)	1.049 (0.356, 1.742)
Absolute integrated discrimination improvement			-0.000933 (-0.00366, 0.00396)		0.000216 (-0.000159, 0.000895)
Continuous net reclassification improvement			-0.175 (-0.255, -0.0926)		0.0329 (-0.0919, 0.268)
Age	√	√	√	√	√
Ratio of linoleic acid to total fatty acids	√	-	√	-	√
Glucose	√	-	√	-	√
3-Hydroxybutyrate	√	-	√	-	√
Creatinine	√	-	√	-	√
Albumin	√	-	√	-	√
Glycoprotein acetyls	√	-	√	-	√
Systolic blood pressure	-	-	-	√	√
Blood pressure medication	-	-	-	√	√
Smoking status	-	-	-	√	√
Prevalent diabetes mellitus	-	-	-	√	√

Components in model 2 and 4 were selected from age, sex, and all metabolites using Boruta algorithm. FSRP, Framingham Stroke Risk Profile. C-index, concordance index.
Framingham Stroke Risk Profile: age, smoking status, prevalent diabetes mellitus, blood pressure medication, and systolic blood pressure.

Table 5.4 Prediction of ischaemic stroke and components of each model in women (n = 101,328)

Predictors	Age and 249 metabolites	Age	Age and metabolites selected by Boruta algorithm	FSRP#	FSRP# and metabolites selected by Boruta algorithm
Model		**1**	**2**	**3**	**4**
Extreme gradient boosting model					
C-index in the training dataset	0.798 (0.784, 0.811)	0.719 (0.703, 0.736)	0.766 (0.752, 0.78)	0.742 (0.727, 0.757)	0.786 (0.772, 0.799)
C-index in the testing dataset	0.717 (0.688, 0.747)	0.714 (0.683, 0.746)	0.706 (0.675, 0.737)	0.727 (0.698, 0.756)	0.721 (0.692, 0.75)
Slope of the calibration line	1.348 (0.761, 1.935)	0.878 (0.582, 1.174)	1.470 (0.791, 2.150)	1.126 (0.723, 1.528)	1.672 (1.186, 2.158)
Absolute integrated discrimination improvement			-0.00324 (-0.00667, -0.00152)		-0.00034 (-0.000624, -0.000123)
Continuous net reclassification improvement			-0.265 (-0.327, -0.189)		-0.249 (-0.336, -0.171)
Random survival forests					
C-index in the training dataset		0.684 (0.668, 0.7)	0.687 (0.671, 0.703)	0.654 (0.637, 0.67)	0.699 (0.683, 0.715)
C-index in the testing dataset		0.707 (0.676, 0.739)	0.687 (0.656, 0.718)	0.662 (0.632, 0.693)	0.703 (0.672, 0.733)
Slope of the calibration line		0.999 (0.567, 1.431)	1.104 (0.676, 1.532)	1.449 (0.554, 2.344)	1.079 (0.654, 1.505)
Absolute integrated discrimination improvement			-0.00195 (-0.00482, -0.000508)		0.0000879 (-0.000189, 0.000524)
Continuous net reclassification improvement			-0.219 (-0.296, -0.145)		-0.0948 (-0.179, 0.187)
Age	√	√	√	√	√
Ratio of omega-6 fatty acids to total fatty acids	√	-	√	-	√
Ratio of saturated fatty acids to total fatty acids	√	-	√	-	√
Creatinine	√	-	√	-	√
Albumin	√	-	√	-	√
Glycoprotein acetyls	√	-	√	-	√
Triglycerides in large LDL	√	-	√	-	√
Triglycerides to total lipids ratio in IDL	√	-	√	-	√
Systolic blood pressure	-	-	-	√	√
Blood pressure medication	-	-	-	√	√
Smoking status	-	-	-	√	√
Prevalent diabetes mellitus	-	-	-	√	√

Components in model 2 and 4 were selected from age, sex, and all metabolites using Boruta algorithm. FSRP, Framingham Stroke Risk Profile. C-index, concordance index.
Framingham Stroke Risk Profile: age, smoking status, prevalent diabetes mellitus, blood pressure medication, and systolic blood pressure.

5.4 Discussion

This large-scale prospective study comprehensively examined the predictive value of adding NMR metabolites to conventional risk factors for ischaemic stroke among 184,564 participants in UKB. Model performance was evaluated across three domains: discrimination using the concordance index, classification using the absolute integrated discrimination improvement (IDI) and continuous net reclassification improvement (NRI), and calibration using calibration slopes and visual agreement between predicted and observed risks. There was little evidence that NMR metabolites provide additional predictive information on incident ischaemic stroke over a median follow-up of 13.8 years when tested in the whole population, older adults, men, and women separately. Across all comparisons, discrimination changed minimally, classification often deteriorated, and calibration showed local misfit at risk extremes.

When age, sex and 249 metabolites were used to predict ischaemic stroke in XGBoost, 10 pivotal features for prediction of incident ischaemic stroke were selected by the Boruta algorithm, including age, sex, albumin, creatinine, glucose, ratio of PUFA to total FA, ratio of LA to total FA, omega-3 FA, glycoprotein acetyls, and free cholesterol to total lipids ratio in very large HDL. One of 8 metabolic features (i.e. albumin) in XGBoost model 2 (which accounted for non-linear associations) overlapped with the findings from the Cox regression in **Chapter 4.**, which indicated that both linear and nonlinear metabolic effects on incident ischaemic stroke were elicited in disease mechanisms. Among patients with acute ischaemic stroke from the Norfolk and Norwich Stroke and TIA Register (N = 9,979, median follow-up time = 5.5 years), Thuemmler et al.[215] also confirmed the non-linear relationship between albumin and long-term mortality risk. Four of 8 metabolic features in XGBoost model 2 overlapped with those in the top 20 important metabolites in an earlier analysis of Buergel et

al. based on the first release of the NMR data (N~120,000) using a neural network.[67] The overlapping metabolites include albumin, creatinine, glycoprotein acetyls, and glucose. The C-statistic of 0.732 (95% CI 0.712-0.751) from XGBoost overlapped that (0.754, 0.742-0.764) from a residual neural network for cerebral stroke. However, the C-statistic was achieved with fewer features (< 15) using the Boruta algorithm in the current study than that (168 metabolites + 8 predictors) in the residual neural network.[67] In line with previous studies,[46,67] the metabolites did not add considerable predictive information over conventional risk factors as captured by FSRP. Using the first release of the NMR data to evaluate the risk of CVD in UKB,[194] Jin et al. also found that metabolites did not provide additional predictive value over well-established CVD risk scores. Taken together, these findings show that, at the discrimination level, adding metabolites does not significantly improve ranking of cases versus non-cases beyond age, sex, and conventional risk factors.

From the perspective of epidemiology and classical biostatistics, it is of interest that the dependence plots for omega-3 FA, ratio of LA to total FA, ratio of PUFA to total FA, creatinine, free cholesterol to total lipids ratio in very large HDL and even a well-established risk factor such as glucose show non-linear relationship with incident ischaemic stroke. From the plot I can also infer the cut-off point of each predictor in which the risk increases in incident ischaemic stroke because the SHAP value indicated the association of ischaemic stroke with each predictor (**Figure 5.3**). The estimated cut-off points of albumin (38.4 g/l) and creatinine (0.0779 mmol/l) are close to those of the normal clinical thresholds for serum albumin value (39-55 g/l)[216] and creatinine value (for adult men, 0.053-0.106 mmol/L; for adult women, 0.044-0.097 mmol/L).[217] The threshold points for ischaemic stroke are of interest for those metabolites whose normal values are not known (e.g. ratio of LA to total FA, ratio of PUFA to total FA, and free cholesterol to total lipids ratio in very large HDL).

This study has new contributions to literature. First, I used an expanded sample sizes of 249 metabolites from the first release of the NMR data[66,67] (N~120,000) to the second release (N~274,000; **Figure 4.1**). Second, I applied and compared two machine learning algorithms (i.e. XGBoost and RSF) to evaluate the predictive value of metabolites. Notably, I identified a reduced set of eight metabolic features (alongside predictors in FSRP) via the Boruta algorithm that maintained a comparable C-statistic of 0.732 (0.712, 0.751; **Table 5.1**), comparable to a much larger input feature set (168 metabolites + 8 predictors) in Buergel et al.[67] using a neural network (C-statistic = 0.754, 95% CI 0.742-0.764). This efficiency in model with comparable accuracy underscores the practical benefit of feature selection techniques like the Boruta algorithm in high-dimensional settings. Third, I provide evidence of non-linear associations between selected metabolites (e.g. albumin, creatinine, and omega-3 FA; **Figure 5.3**) and incident ischaemic stroke, as revealed by SHAP value-based dependence plots from the XGBoost model. This observation aligns with and extends prior findings by Thuemmler et al.[215] regarding the prognostic importance of albumin, and contributes new insights into potential threshold effects for under-characterised lipid ratios (e.g. ratio of LA to total FA, ratio of PUFA to total FA, and free cholesterol to total lipids ratio in very large HDL; **Figure 5.3**), which lack established clinical reference ranges. Forth, I explicitly evaluated discrimination, calibration, and classification showing that negative IDI, NRI and tail-area miscalibration explain why apparently informative metabolite features did not yield meaningful gains in prediction.

There are also limitations to this study. First, the NMR metabolic coverage of this study is relatively lipid focused and narrow compared to 217,920 compounds in The Human Metabolome Database 5.0 in 2022.[30,218] Although a growth of NMR metabolic coverage is anticipated, it restricts its prediction merit and clinical application to date. Second, the

participants in UKB were recruited on a volunteer basis and were healthier than the general UK population.[84,184] This may have influenced our findings, although one may argue that studying the relevance of predictors of disease is less likely to be bias by the non-representativeness of the selection of participants than studies of the burden of disease in populations.[184] Third, fasting before the blood test was not required, and levels of TG, total cholesterol, and LDL cholesterol would be influenced.[219] However, the fasting/postprandial state accounted for less than 2% of the total variance in the NMR spectroscopy-based plasma metabolite quantification.[220] Forth, the use of a random split of the dataset into training (80%) and testing (20%) subsets for model development and internal validation is a weak form of internal validation. While this approach has been commonly used in machine learning workflows,[197,202,221,222] and is superior to a non-random split, the two datasets would be closely similar, because they varied only by chance (such that the testing dataset would probably show similar performance as in the training dataset); and different splits led to different results notably in relatively small data sets.[197] Furthermore, it is unclear how much data should be used as the training dataset and how much should be set aside to evaluate the model.[197] According to the TRIPOD recommendations,[197] resampling-based techniques such as bootstrapping or cross-validation are preferred for internal validation, as they provide more stable estimates of model performance and reduce the dependency on a single random split. However, due to the large sample size in UKB, for this analysis with already several levels of complexity, it was decided that random split as a method of internal validation was sufficient to avoid further complexity of other methods like bootstrapping or cross-validation.

5.5 Conclusion

Given the vast global burden of ischaemic stroke, it is crucial to identify individuals at higher risk of developing ischaemic stroke to allow early interventions to delay or prevent its

onset. This large-scale prospective study demonstrates that the improvement in predictive accuracy when NMR metabolites were added to the conventional risk factors model was little, making it urgent to discover other biomarkers for better prediction. This finding indicates that compared to the algorithm currently used for assessment of stroke risk (i.e. FSRP), ischaemic stroke risk prediction in national screening programmes or clinical practice is, at least using the methods applied here and within the limitations already described, unlikely to be improved by the addition of NMR metabolites. Future studies should prioritise evaluating the advantages of metabolic profiling to ischaemic stroke risk in non-White ethnic groups in high-income populations, populations in other parts of the world, and discover the predictive value of a broader range of metabolic biomarkers, as well as other types of biomarkers (e.g. gut microbiome and proteomics).

Chapter 6. Assessment of the causality of the associations between metabolites and ischaemic stroke: A Mendelian randomisation study

Chapter summary

This chapter assesses whether circulating metabolites play a causal role in ischaemic stroke using Mendelian randomisation (MR). First, a new genome-wide association study was conducted in the updated UK Biobank metabolomics dataset (about 230,000 White British participants) for 47 metabolites selected from earlier observational and machine-learning analyses. This expanded scan identified 6,137 genome-wide significant variant-metabolite associations, nearly half of which were novel, and yielded substantially stronger genetic instruments than previous work.

Two-sample MR combined these instruments with large stroke genetics from the GIGASTROKE consortium and applied a suite of pleiotropy-robust estimators. Consistent evidence across multiple methods supported higher genetically predicted levels of cholesterol to total lipids ratio and cholesteryl esters to total lipids ratio in chylomicrons and extremely large VLDL as causal risk factors for ischaemic stroke. A genetically predicted level of triglycerides to total lipids ratio in very large VLDL showed suggestive inverse association. Multivariable MR that accounts for correlated exposures singled out the cholesteryl ester ratio in chylomicrons and extremely large VLDL as having a direct effect. Reverse analyses found no evidence that genetic liability to ischaemic stroke alters metabolite levels.

Widespread horizontal pleiotropy was detected and addressed with multiple sensitivity frameworks, but residual bias cannot be ruled out. Key limitations include restriction to White British ancestry, single-time-point metabolite measures, and modest

sample overlap. Findings highlight remnant-cholesterol pathways as potential therapeutic targets and motivate validation in diverse cohorts and intervention studies.

6.1 Background

Observational studies have generated estimates of the associations of metabolites with ischaemic stroke.[46,47] However, a series of biases and limitations (e.g. reverse causation, measurement error, and residual confounding) are of concern to observational epidemiology, and causal inference can often be challenging as a result, especially in the absence of randomised controlled trials.[223] There have been numerous examples of seemingly strong observational associations between disease risk and pharmacological, physiological or behavioural measures, which were not affirmed as the expected health benefits in subsequent randomised controlled trials,[224] including circulating biomarkers (e.g. HDL cholesterol),[225] pharmacological agents (e.g. hormone replacement therapy),[226] and nutritional factors (e.g. vitamin C,[227] vitamin E,[228,229] beta carotene,[228] folate and homocysteine[230]). For example, four prospective American studies[16] showed that a 1-mg/dl (0.026 mmol/L) increment in HDL cholesterol was associated with 3.7-4.7% decrement in cardiovascular disease mortality rates. Nevertheless, in a 12-month randomised, double-blind study among 15,067 participants with high cardiovascular risk, the torcetrapib plus atorvastatin therapy elevated HDL cholesterol by 72.1%, but it also led to an increased risk of cardiovascular events (HR 1.25, 95% CI 1.09-1.44) and all-cause mortality (1.58, 1.14-2.19) compared to the atorvastatin-only therapy.[225] Considering these obstacles and leveraging on the success of human genetic research, statistical methods employing genetic data provide the opportunity to delineate the causal inference in the presence of such limitations. According to Mendel's Law of Segregation dictating that "parental alleles are randomly assigned to offspring", Mendelian randomisation (MR) is an instrumental variable analysis and employs the random assignment of alleles via meiotic cell division and conception as the principle for a natural experiment.[39,223,231] According to this principle,

if a genetic variant influences the level of, or mirrors the biological effect of, a metabolite, and that metabolite is related to ischaemic stroke, then, assuming causality, the genetic variant should also be related to ischaemic stroke. In these analyses, the possibility of reverse causation can be prevented since germ-line genetic variants are determined at conception and temporally precede the metabolic measurement. Offspring genetic variants are less likely to be vulnerable to confounders in the population and can be used as proxy markers for metabolites of interest.[232,233] Moreover, the quantitative evaluation of causal inference derived from an MR framework permits the re-estimation of observational effects attributable to the metabolite being instrumented for by genetic variants.[234] Taken together, these components give a natural experiment that can be applied either to question, or to corroborate, the observational results in the absence of randomised controlled trials.[235]

In the face of the theoretical evolution of MR, developments in the detection of robust associations between risk factors of interest and genetic variation are now rapidly boosting the potential of MR to ascertain the causal pathways of ischaemic stroke. An MR study can be conducted using genetic instruments, ischaemic stroke, and metabolic levels from participants measured in the same sample (i.e. single-sample MR or one-sample MR) or from two samples (i.e. two-sample MR) in which data on the metabolites and ischaemic stroke have been measured in different or only partially overlapping samples. Although one-sample MR benefits from harmonization and subgroup analyses, the effects are more likely to be biased towards the observational associations since both the gene-exposure and the gene-outcome associations are derived from the same sample.[236] Instead, two-sample MR has often the advantage of increased power because the estimates can be obtained from publicly available GWAS summary data of consortia, including many studies, and thus higher sample sizes.[236]

A new discovery GWAS is warranted because current genetic instruments explain only a limited portion of the variance in nuclear magnetic resonance (NMR) metabolites, which restricts the power of MR to detect causal effects.[237] Although Harshfield et al.[41] recently harnessed the power of UKB (N~120,000) to assess the relationship of NMR metabolites with ischaemic stroke using MR, their study found no causal evidence, likely due to insufficiently strong instruments and limited power to detect small effect sizes. In contrast, I used the second release of the UKB NMR data, which more than doubles the sample size available compared with the first release analysed by Harshfield et al. This expanded dataset provides the opportunity to conduct a new GWAS with larger sample sizes that have greater power to identify novel and stronger genetic variants associated with NMR metabolites, thereby improving instrument strength and enhancing the power of MR to uncover subtle causal relationships with ischaemic stroke.[238]

Here, I conduct a large-scale integrative genetic and metabolomic analysis to explore putative causal inference between NMR metabolites and ischaemic stroke. The targeted 47 metabolites were selected from the findings of Cox regression (40 metabolites significantly associated with incident ischaemic stroke in the fully adjusted Cox regression; **Table 4.3**; **Chapter 4.**) and XGBoost (8 Boruta-selected metabolites while albumin overlapped with one of prior 40 metabolites; **Table 5.1**; **Chapter 5.**). The current chapter comprises two parts: first, a GWAS was conducted to explore robust genetic instruments of metabolites in the updated NMR metabolomic data of UKB involving over 270,000 participants; second, two-sample MR was used to examine causal inference between metabolites and ischaemic stroke.

6.2 Methods

To provide a systematic evaluation of causality between metabolites and ischaemic stroke, my analytic strategy followed a multi-step framework. First, I conducted a new GWAS of NMR metabolites in the second release of UKB to generate updated, stronger genetic instruments. Second, I applied two-sample MR using the inverse-variance weighted method as the primary approach, given its efficiency under standard instrumental variable assumptions. Third, to account for potential violations of MR assumptions, I implemented several robust MR methods that provide more reliable inference when some assumptions of the inverse-variance weighted method are not met. Fourth, I employed multivariable MR to address the challenge of high correlations among metabolites and to disentangle direct from indirect causal effects. Fifth, I performed secondary analyses with reverse MR to explore metabolic consequences of genetic liability to ischaemic stroke. Finally, I adopted a structured framework to integrate results across all these approaches, classifying associations as causal, suggestive, or null depending on consistency and robustness across methods.

6.2.1 Genome-wide association study of metabolites

Published GWAS summary statistics for each metabolite in the first release of the NMR data (N~120,000) in UKB have been available from the MRC Integrative Epidemiology Unit OpenGWAS platform.[42] Since genetic effects are usually of small magnitudes, and further increases in sample sizes can facilitate discovery of smaller effects, I performed GWAS for the 47 metabolites in the updated dataset with more than double the number of participants (N = 274,355) to determine single nucleotide polymorphisms (SNPs) as instrumental variables of the following two-sample MR (**Figure 6.1**). Participants who had mismatched genetic and self-reported sex (n = 211), were not of white British ancestry (n = 43,178), had sex chromosome aneuploidy (i.e.

putatively carrying sex chromosome configurations that are not either XX or XY, n = 239), had ten or more third-degree relatives identified (i.e. excessive relatives, n = 107), were identified as outliers in heterozygosity and missing rates (i.e. the genotypes for these samples are of poor quality, n = 418), or had missing data for fasting time (n = 5) were successively excluded in the current analyses.[239] To minimize bias from population stratification, individuals not classified as white British (defined by UKB Field 22006 "Genetic ethnic grouping",[85] denoting participants who self-identified as "White British" according to Field 21000 "Ethnic background" and have very similar genetic ancestry according to a principal components analysis of the genotypes) were excluded. This filtering step is consistent with the methods reported by Bycroft et al.[85] and current GWAS guide using REGENIE on the UKB Research Analysis Platform,[240] which demonstrate that fine-scale structure within European ancestry can lead to confounding and inflated false-positive rates if not properly controlled. Therefore, despite potential sample size gains, inclusion of white non-British participants without stratified analysis would compromise GWAS validity. Thus I only retained the genetically homogeneous "White British" subset for GWAS.

The descriptive characteristics in the analysis sample were assessed using the "CBCgrps" package[143] and shown as median and the interquartile range (IQR) for continuous variables and proportions for categorical variables.

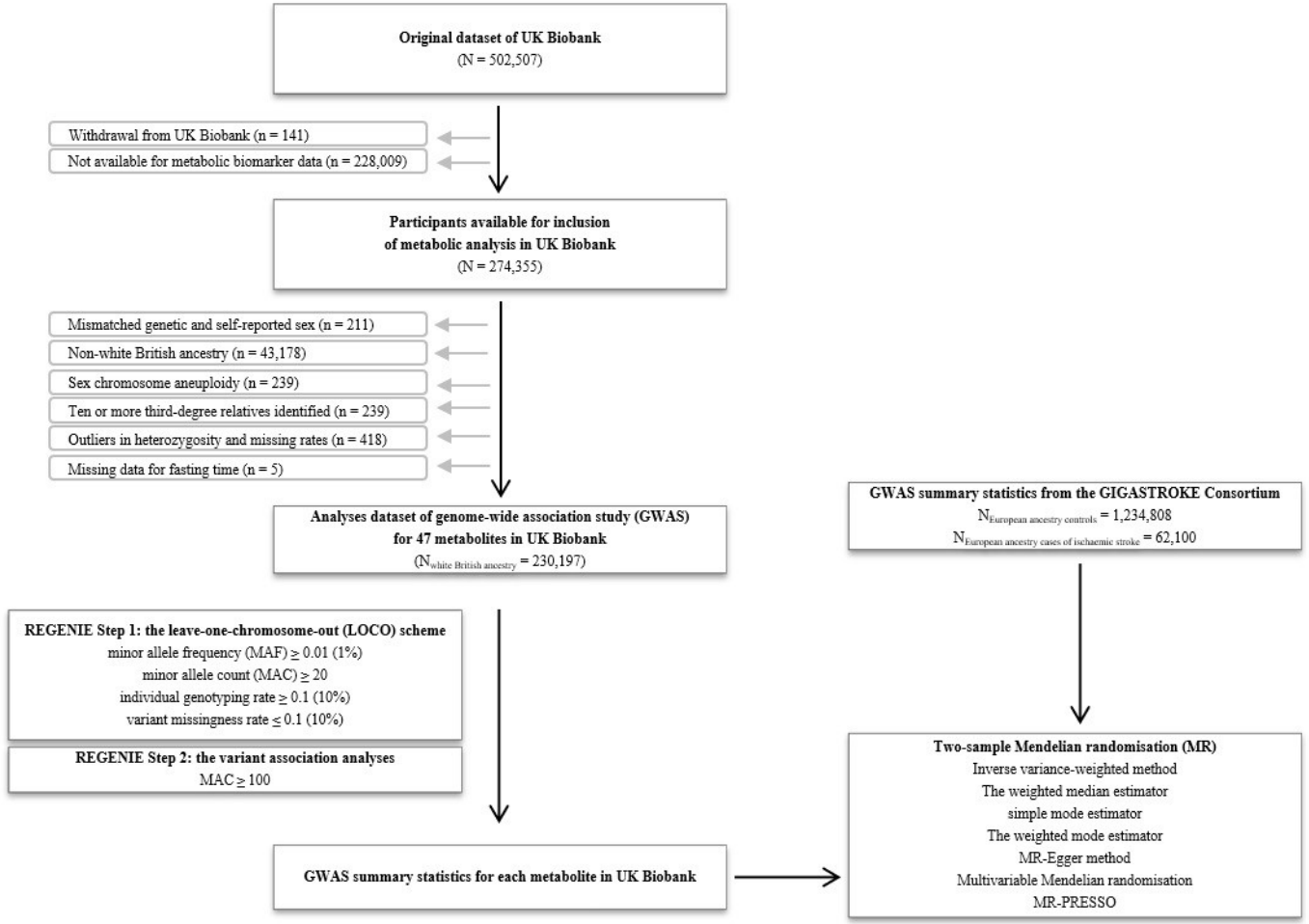


Figure 6.1 A flowchart describing the genetic study.
GWAS, genome-wide association study. MR-PRESSO, Mendelian randomisation pleiotropy residual sum and outlier.

REGENIE v3.3,[241] a two-step whole-genome regression model accounting for relatedness and population structure, was applied on the remaining dataset (N = 230,197) to perform GWAS for the 47 metabolites selected from the epidemiological analyses and the machine learning algorithm. In the first step, the leave-one-chromosome-out (LOCO) scheme was carried out to fit a whole-genome regression model for individual metabolite predictions based on genetic data. The stringent quality-control filters were conducted with PLINK 2.0 alpha[242] to procure a set of high-quality genotyped variants based on minor allele frequency (MAF) by keeping variants with MAF ≥ 0.01 (1%), minor allele count (MAC) by retaining variants with MAC ≥ 20, individual genotyping rate by preserving individuals with genotyping rate ≥ 0.1 (10%), and variant missingness rate by retaining variants with missingness rate ≤ 0.1 (10%). The LOCO metabolite predictions from step 1 were applied as offsets in step 2, and the standard linear regression was used to carry out the variant association analyses. The variant association analyses were limited to variants with MAC ≥ 100 to minimize spurious associations.[243] The rank inverse normal transformation was applied to each metabolite, so that they are distributed normally and expressed in SD units, and covariates included age, sex, fasting time, genotype measurement batch, and the first twenty genetic principal components to control for population stratification. Results were considered genome-wide significant at $p < 5.0 \times 10^{-8}$.[244,245] Based on the European 1000 Genomes LD reference panel,[246] biallelic SNPs with LD $r^2 < 0.001$ and a physical distance threshold of 10,000 kb were retained as the genetic variants independently associated with each metabolite.[237,247] Thus, genomic loci were defined using a 10,000 kilobase (kb) window around the lead SNP in the discovery genome-wide association analyses. In addition, the genetic correlation matrix among the metabolites (249 x 249) was derived using cross-trait LD Score regression, a method that estimates genetic

correlations from GWAS summary statistics and is robust to sample overlap and confounding due to shared population structure.[248]

To evaluate test statistic inflation in GWAS, the genomic inflation factor (λ_{GC}) was estimated for each metabolite.[249,250] The λ_{GC} was calculated as the ratio of the observed median of the chi-squared test statistics for association (χ^2) to the expected median under the null hypothesis.[249] For a chi-squared distribution with 1 degree of freedom, the null median corresponds to the 50th percentile of the χ^2 distribution, denoted as $\chi^2_{0.5,\ df=1}$.

$$\lambda_{GC} = \frac{median\left(\chi^2_{observed}\right)}{\chi^2_{0.5,\ df=1}}$$

The SNP-based heritability of each metabolite was estimated using LD Score Regression, a method that distinguishes polygenicity from confounding in genome-wide association summary statistics.[251,252] The LD Score Regression leverages the relationship between LD and test statistics (i.e. chi-squared statistics for association between each SNP and the metabolic level) to infer the proportion of phenotypic variance attributable to common variants across the genome.[251] As part of this analysis, the LD Score Regression intercept was used to quantify the extent of inflation in the GWAS test statistics due to confounding factors, such as population stratification or cryptic relatedness.[252] An intercept value near 1.0 indicates little evidence of confounding, while values substantially greater than 1.1 may suggest meaningful inflation beyond polygenic signal (e.g. population stratification or cryptic relatedness).[252] Unlike the genomic inflation factor (λ_{GC}), which conflates true polygenic signal with confounding, the LD Score Regression intercept provides a more robust and interpretable estimate of non-polygenic inflation by explicitly modelling the contribution of LD.[252] Summary statistics for each metabolite were derived from the GWAS results of my thesis. The LD Score Regression was performed using precomputed LD scores and regression weights

from European-ancestry samples.[251] The analysis was conducted using the software, ldsc.py script (v1.0.1),[251] with LD scores and weights from the European 1000 Genomes LD reference panel.[246] In addition, the genetic correlation matrix among the metabolites (249 x 249) was derived using cross-trait LD Score regression, a method that estimates genetic correlations from GWAS summary statistics and is robust to sample overlap and confounding due to shared population structure.[248,253]

To classify SNP-gene associations as *cis* or *trans*, the genomic distance between each SNP and the nearest gene transcription start site was determined. Gene annotations were processed using the R package "GenomicRanges",[254] and transcription start site coordinates were computed based on strand orientation, using the start position for genes on the forward strand and the end position for those on the reverse strand. For each SNP, the nearest transcription start site was calculated, and variants located within 1 megabase (i.e. 1Mb; 1,000,000 bases) of the nearest transcription start site were classified as *cis*. Variants beyond this window were considered *trans*.[255] This 1 megabase threshold is a widely used convention to differentiate local from distal regulatory effects in genetic association studies.[255-257]

To define known variants, I constructed two reference sets from harmonized GWAS Catalog data[258] updated until August 2025 with European ancestry and harmonized genomic coordinates to the GRCh37/hg19 reference assembly using the UCSC LiftOver tool and the R package "rtracklayer":[259] (1) lead variants previously reported for each metabolite based on their reference SNP identifiers, and (2) a positional set containing unique combinations of chromosome numbers and base-pair positions in the GRCh37 reference genome. This ensured that all variants were represented consistently in a single genomic build before assessing their novelty.

Novelty was determined using a hierarchical matching approach:

(1) If a variant's reference SNP identifier matched any previously reported variant, it was classified as known.

(2) If no identifier match was found, I compared chromosome numbers and base-pair positions; variants located at the same genomic positions as previously reported variants were classified as known.

(3) If no exact identifier or positional match was found, I examined variants within ± 1 megabase (Mb) and considered those in strong linkage disequilibrium ($r^2 \geq 0.40$) with any known variant to be known as well. The choice of the linkage disequilibrium (LD) threshold of $r^2 \geq 0.40$ was in line with the interactive Cross-Phenotype Analysis of GWAS database (iCPAGdb),[260] which adopted this threshold as a practical decision to reduce multiple counting of regions with broad association signals in moderate LD, while maintaining sensitivity and specificity in detecting cross-phenotype associations.

Variants failing all three overlap checks were classified as novel.

6.2.2 Mendelian randomisation

Two-sample MR was conducted to evaluate the causal relationship of metabolites, from the findings of conventional statistical methods and machine learning, with ischaemic stroke using the "TwoSampleMR" package version 0.5.6.[42] I applied the GWAS summary statistics with units of OR from the GIGASTROKE Consortium as the outcome dataset,[261] containing 1,234,808 European ancestry controls and 62,100 European ancestry cases of ischaemic stroke (30% of controls and 4% of cases from UKB were included in GIGASTROKE). The GWAS summary statistics for each metabolite in UKB from this thesis were applied as the exposure dataset. Outcome and exposure datasets were harmonised to

assure that the effect size of each SNP on ischaemic stroke and the metabolite were registered on the same allele.

Estimates of the proportion of variance in NMR metabolites explained by the genetic instruments (R^2) were calculated, and F-statistics (F) were used to evaluate the strength of the instruments.[82,262,263]

$$R^2 = 2\beta^2 *(EAF)*(1-EAF)/ (2\beta^2*(EAF)*(1-EAF) + (se(\beta)) 2*2N*EAF*(1-EAF))$$

$$F = R^2*(N-1-k)/ ((1- R^2)*k)$$

In the formulas, β is the effect size for the genetic effect on the metabolites. se(β) is the standard error of the effect size. EAF is the effect allele frequency. N is the sample size of the GWAS. k is the number of SNPs included in the instrumental variable.

The instrumental variable estimates may be vulnerable to finite sample bias (i.e. weak instrument bias) despite the fact that instrumental variable techniques are asymptotically unbiased against confounding.[264] The bias of the estimator from the instrumental variable is 10% of the bias of the observational estimator if the F-statistic is 10.[262] And the F-statistics greater than 10 are typically assumed to imply sufficient strength to assure the validity of instrumental variable methods, although it is useful to note that the F-statistic is also sensitive to sample size variations.[262,265]

MR analyses rely on three core assumptions: (1) the relevance assumption, which requires that the genetic variants are robustly associated with the exposure; (2) the independence assumption, which requires that the variants are not associated with confounders of the exposure-outcome relationship; and (3) the exclusion restriction assumption, which requires that the variants influence the outcome only through the exposure.[266]

The inverse variance-weighted method was applied as the primary MR analysis, combining individual ratio estimates from each genetic variant into a single more efficient estimate of the causal effect of each metabolite on ischaemic stroke.[267,268] Heterogeneity across the estimates for each instrument in the inverse variance-weighted method with Cochran's Q statistic was applied to test global pleiotropy.[269] The inverse variance-weighted method requires all three MR assumptions (i.e. exclusion restriction, independence, and relevance) to hold.[39,270] As this method has no breakdown point, the estimator is often biased if there is even one invalid instrumental variable.[271] Bias would be avoided only when the average pleiotropic effects of genetic variants is zero (i.e. balanced pleiotropy) and the pleiotropic effects are independent of the relation between the genetic variants and the metabolite.[272] It is less possible that all variants are valid instrumental variables since a great number of variants are used in the current study (> 50).[268] Therefore, it is important to apply other robust estimation methods regardless of their lower statistical efficiency (i.e., lower power to identify a causal relationship).[268]

To assess robustness of the inverse-variance weighted findings, four alternative methods were conducted, which require different assumptions: (1) MR-Egger method (which allows directional pleiotropy to all instrumental variables under the assumption that direct pleiotropic effects are independent of the instrument strengths);[273] (2) the weighted median estimator (which takes the median of the ratio estimates and enables the use of up to 50% of invalid instrumental variables, and is less influenced by outliers than other models);[271] (3) the simple mode estimator (which clusters the SNPs into groups according to similarity of causal effects, and carries out the causal effect estimate according to the cluster that has the largest number of SNPs, with an assumption that the mode of the bias terms for the included instruments is zero);[274] and (4) the weighted mode estimator (which applies the same

procedures as the simple mode estimator but gives weights to each SNP).[274] Because the MR-Egger intercept gives a poor test to detect global pleiotropy and was never designed for the standard heterogeneity statistic,[269] the Rucker's Q statistics of the MR-Egger method were instead applied to evaluate the heterogeneity of the MR analysis.[42,269] Additional sensitivity analyses with the Mendelian randomisation pleiotropy residual sum and outlier (MR-PRESSO) method were conducted to identify and correct outliers and contained 3 elements: (a) the MR-PRESSO global test: detection of horizontal pleiotropy; (b) the MR-PRESSO outlier test: accounting for horizontal pleiotropy by removing outliers; and (c) the MR-PRESSO distortion test: evaluating significant differences in the causal estimates before and after accounting for outliers.[275]

Regarding the assessment of MR assumptions in this study, the relevance assumption was formally evaluated by calculating the proportion of variance (R^2) explained by the instruments and the F-statistics, with $F > 10$ used as the threshold for strong instruments. The independence and exclusion restriction assumptions were partly assessed using sensitivity analyses. Heterogeneity among SNP-specific estimates was examined using Cochran's Q-statistic, while directional pleiotropy was evaluated using the MR-Egger intercept. In addition, the four robust MR methods described above, such as MR-Egger, weighted median, and mode-based estimators, were employed to relax assumptions and account for the possibility of invalid instruments. These complementary approaches collectively strengthened confidence in the validity of the MR assumptions.

Causal language was used to describe MR estimates in results text since MR models evaluated the effects of each metabolite on ischaemic stroke via genetically predicted exposure values based on standard instrumental variable assumptions.[276] The false discovery rate (FDR) correction with the Benjamini-Hochberg method was used to adjust for multiple testing when

identifying significant relationships for each MR method.[145] The MR reporting complied with the Strengthening the Reporting of Observational Studies in Epidemiology Using Mendelian Randomization (STROBE-MR) Statement (**Supplementary Table 6.1**).[277]

6.2.3 Multivariable Mendelian randomisation

To address challenges in estimating direct causal effects from high-dimensional and highly correlated metabolites, in which genetic instruments are likely to share a large number of variants, the Multivariable Mendelian Randomisation with Constrained Maximum Likelihood and Sum of Single Effects (MVMR-cML-SuSiE) framework, recently developed by Chan et al.,[253] was employed. This approach integrates the Multivariable Mendelian Randomisation with Constrained Maximum Likelihood (MVMR-cML)[278] with the Sum of Single Effects (SuSiE) model, [279] a Bayesian variable selection method originally designed for genetic fine-mapping. The MVMR-cML-SuSiE improves statistical power and interpretability in the presence of multicollinearity among exposures, a common feature in omics datasets such as metabolomics.[253] MVMR-cML-SuSiE operates in the following steps: (i) Marginal screening of metabolites using univariable Mendelian randomisation cML with model averaging (UVMR-cML-MA)[280] to identify candidate metabolites. In this step, the causal effect of each metabolite on the outcome is estimated while allowing for the presence of pleiotropic (invalid) instruments. The constrained maximum likelihood framework models both the causal and direct effects of SNPs, while model averaging accounts for uncertainty in the number and identity of invalid instruments by averaging estimates across a set of models weighted by likelihood-based criteria.[280] This approach enhances robustness to horizontal pleiotropy and reduces bias introduced by relying on a single model specification;[280] (ii) Removal of invalid instruments, which is defined as a genetic variant with a nonzero pleiotropic effect on the outcome that is not mediated through the specific metabolite,[253] via

univariable Mendelian Randomisation using Constrained Maximum Likelihood with Bayesian Information Criterion (UVMR-cML-BIC);[253] (iii) Identification of signal clusters using an iterative application of the SuSiE algorithm[278] on the cML profile likelihood. This cluster-based approach preserves sets of correlated exposures, avoiding arbitrary selection of individual traits, and provides posterior inclusion probabilities (PIPs) for each metabolite.[253] In order to implement the iterative SuSiE algorithm within the MVMR-cML-SuSiE framework, I adopted a 249 × 249 genetic correlation matrix of the metabolites,[248,253] estimated using cross-trait LD Score regression.

6.2.4 Secondary analyses

A reverse-MR framework was applied to investigate the metabolic profile of genetic liability to ischaemic stroke, with the aim of both assessing the possibility of reverse causality and evaluating the metabolic traits which portray the progress of ischaemic stroke.

6.2.5 Framework for determining causality between metabolites and ischaemic stroke

The evidence for causality was evaluated based on the above results of MR, including inverse variance-weighted method, MR-Egger, weighted median estimator, the simple mode estimator, weighted mode estimator, MR-PRESSO, and MVMR-cML-SuSiE. Evidence for causality (**Supplementary Figure 6.1**) was achieved when (1) at least one of the aforementioned performed methods for the specific association presented a significant p value after adjusting for multiple testing; and (2) the direction of the effect estimate was concordant for all the methods. Suggestive evidence for causality was achieved when at least one method had a significant p value after adjusting for multiple testing, but the direction of the effect estimates differed between methods. This framework was adapted from Markozannes et al.,[281] who proposed evidence-grading criteria in their systematic review of MR studies on cancer. I

adopted their approach because it provides a transparent and reproducible way to summarize causal inference across multiple MR methods.

6.3 Results

6.3.1 Genome-wide association study in UK Biobank

The 230,197 eligible White British participants had a median age of 58 (IQR 51-63) years, 46.1% were men, had fasting time of 3 (2-4) hours, BMI of 26.8 (24.2-29.9) kg/m^2, and SBP 137 (125-150) mmHg (**Table 6.1**). Nearly half (45.0%) reported drinking alcohol at least thrice/week. The proportion for current smokers was 10.1%. The prevalence of blood pressure medication and lipid modifying medication use was 21.3% and 20.7%, respectively.

The REGENIE regression GWAS results between genetic variants and 47 metabolites are shown in **Supplementary Figure 6.2**, and genome-wide significant associations are highlighted in **Supplementary Table 6.2**. For example, multiple independent variants showed strong associations with cholesteryl esters to total lipids ratio in chylomicrons and extremely large VLDL, including rs117733303 (beta = -0.266, SE = 0.011, $p < 5 \times 10^{-8}$), rs118039278 (beta = 0.222, SE = 0.005, $p < 5 \times 10^{-8}$), and rs11591147 (beta = -0.131, SE = 0.011, $p < 5 \times 10^{-8}$). All of these variants were classified as cis-acting variants, with distances less than 1 megabase (Mb) to the transcription start sites of nearby genes, consistent with a direct regulatory influence.

Across the 47 metabolites, the number of independent genome-wide significant associations varied substantially (**Supplementary Table 6.3**), ranging from 20 for phenylalanine to 191 for cholesterol in very large HDL. This wide distribution highlights the differential genetic architecture underlying these metabolites. Importantly, a total of 6,137 significant associations were identified, of which 3,025 (49.3%) represented novel variants not

reported in previous studies. At the metabolite level, the proportion classified as novel varied widely, ranging from 3.9% ($N_{novel\ SNPs}$ = 4) for the ratio of polyunsaturated fatty acids to total fatty acids to 81.7% ($N_{novel\ SNPs}$ = 156) for cholesterol in very large HDL. Among these, several metabolites exhibited a high proportion of novel associations, reflecting gaps in the current understanding of the metabolic genetic architecture. For example, cholesterol in very large HDL had 191 significant associations, of which 156 (81.7%) were newly identified. Similarly, free cholesterol in very large HDL showed 130 novel associations out of 160 significant variants (81.3%), while cholesterol in very small VLDL yielded 98 novel variants among 125 significant signals (78.4%). In addition, cholesteryl esters to total lipids ratio in chylomicrons and extremely large VLDL demonstrated 50 significant associations, of which 15 (30.0%) were novel, while cholesterol to total lipids ratio in chylomicrons and extremely large VLDL had 51 significant associations with 21 (41.2%) being novel. In contrast, several metabolites exhibited fewer novel associations despite a relatively high number of significant findings. For example, omega-3 fatty acids had 118 significant associations, but only 20 (16.9%) were novel; concentration of HDL particles showed 141 significant associations, with 17 (12.1%) being novel; and ratio of polyunsaturated fatty acids to total fatty acids had 103 significant variants, of which only 4 (3.9%) were novel. Overall, across the 47 metabolites, a total of 6,137 genome-wide significant SNPs were identified, of which 3,025 (49.3%) were classified as novel associations based on stringent criteria that considered reference SNP identifiers, genomic position, and LD overlap with previously reported variants (**Supplementary Table 6.3**). These findings highlight that nearly half of the observed associations represent previously unreported genetic signals, providing new insights into the genetic regulation of metabolite levels.

In terms of independent signals, the number of unique genome-wide significant SNPs per chromosome also showed substantial variability (**Supplementary Table 6.4**). Chromosome 1 contributed the highest number of independent associations (231 SNPs), followed by chromosome 6 (228 SNPs), and chromosome 11 (162 SNPs), indicating potential hotspots for metabolic regulation. Moreover, rs72836561 was the top variant with pleiotropic effects (**Supplementary Table 6.5**), being associated with 36 different metabolites, such as concentration of HDL particles, average diameter for VLDL particles, ratio of PUFA to total fatty acids (FA), ratio of linoleic acid (LA) to total FA, total lipids in chylomicrons and extremely large VLDL, and free cholesterol in chylomicrons and extremely large VLDL. The mapped gene *CD300LG*, where rs72836561 resides, has been implicated in lipid metabolism, with the Arg82Cys polymorphism previously associated with altered lipoprotein levels, intramyocellular lipid accumulation, and impaired glucose uptake.[282] These associations underscore the broad regulatory role that certain variants may exert across multiple metabolic pathways. Together, these observations reinforce the polygenic nature of circulating metabolites, particularly lipoprotein subclasses, and highlight specific chromosomes and variants that may contribute disproportionately to metabolic variation. These results also guided the prioritization of loci for further fine-mapping and functional follow-up.

Table 6.1 Characteristics of participants in genome-wide association analyses in UK Biobank (n = 230,197)

Exposures	
HDL_P, mmol/l	0.015 (0.014-0.017)
VLDL_size, nm	38.6 (37.7-39.6)
Omega_3, mmol/l	0.5 (0.4-0.6)
PUFA_pct, %	42.6 (39.9-44.7)
LA_pct, %	29 (26.7-31.1)
Phe, mmol/l	0.046 (0.040-0.053)
Glucose, mmol/l	3.6 (3.1-4)
Creatinine, mmol/l	0.066 (0.058-0.075)
Albumin, g/l	39.4 (37.3-41.5)
GlycA, mmol/l	0.8 (0.7-0.9)
XXL_VLDL_L, mmol/l	0.2 (0.1-0.3)
XXL_VLDL_FC, mmol/l	0.02 (0.01-0.04)
XS_VLDL_C, mmol/l	0.2 (0.1-0.2)
XS_VLDL_CE, mmol/l	0.1 (0.1-0.1)
XL_HDL_L, mmol/l	0.1 (0.1-0.2)
XL_HDL_C, mmol/l	0.1 (0.1-0.1)
XL_HDL_FC, mmol/l	0.022 (0.018-0.026)
M_HDL_L, mmol/l	1 (0.9-1.2)
M_HDL_PL, mmol/l	0.5 (0.4-0.6)
S_HDL_P, mmol/l	0.010 (0.009-0.011)
S_HDL_L, mmol/l	1.2 (1.1-1.3)
S_HDL_PL, mmol/l	0.66 (0.61-0.73)
S_HDL_C, mmol/l	0.45 (0.41-0.49)
S_HDL_CE, mmol/l	0.33 (0.30-0.36)
XXL_VLDL_C_pct, %	26.1 (22.9-31.6)
XXL_VLDL_CE_pct, %	14.4 (12.1-18.2)
XL_VLDL_C_pct, %	27.8 (24-32.8)
XL_VLDL_CE_pct, %	16 (12.9-20.1)
XL_VLDL_TG_pct, %	53.5 (48.4-57.7)
M_VLDL_PL_pct, %	22.4 (21.1-23.5)
M_VLDL_C_pct, %	30.4 (25.5-34.8)
M_VLDL_CE_pct, %	16.5 (12.9-20)
M_VLDL_FC_pct, %	13.8 (12.6-14.9)
M_VLDL_TG_pct, %	47.3 (41.8-53.3)
S_VLDL_C_pct, %	38.2 (34.9-41.2)
S_VLDL_CE_pct, %	23.8 (21.8-25.5)
S_VLDL_FC_pct, %	14.4 (12.9-15.7)
S_VLDL_TG_pct, %	38.2 (33.9-42.7)
XS_VLDL_C_pct, %	51.8 (48.5-54.4)
XS_VLDL_CE_pct, %	35.7 (32.7-38.1)
XS_VLDL_TG_pct, %	19 (16.8-21.9)

M_LDL_CE_pct, %	49.3 (48-50.4)
S_LDL_PL_pct, %	31.2 (30.1-32.3)
S_LDL_CE_pct, %	45.9 (44.6-47.2)
XL_HDL_FC_pct, %	14.6 (12.7-16.9)
L_HDL_FC_pct, %	10.2 (9.5-10.7)
S_HDL_FC_pct, %	9.9 (9.6-10.2)
Covariates	
Age, years	58 (51-63)
Male, n (%)	106120 (46.1)
Fasting time, hours	3 (2-4)
Others	
Body mass index, kg/m^2	26.8 (24.2-29.9)
Alcohol intake, n (%)	
Daily or almost daily	48131 (20.9)
Three or four times a week	55528 (24.1)
Once or twice a week	61429 (26.7)
One to three times a month	25612 (11.1)
Special occasions only	24434 (10.6)
Non-drinker	15063 (6.5)
Smoking status, n (%)	
Never	125333 (54.4)
Previous	81530 (35.4)
Current	23334 (10.1)
Physical activity based on International Physical Activity Questionnaire, n (%)	
Low	44247 (19.2)
Moderate	96029 (41.7)
High	89921 (39.1)
Systolic blood pressure, mmHg	137 (125-150)
Diastolic blood pressure, mmHg	82 (75-89)
Blood pressure medication, n (%)	49061 (21.3)
Lipid modifying medication, n (%)	47682 (20.7)

Continuous variables are presented as the median with interquartile ranges, and categorical variables are presented as frequencies with percentages. Abbreviations are in line with **Supplementary Table 3.1**.

The genomic inflation and heritability estimates for 47 circulating metabolites were summarized using both genomic control (λ_{GC}) and LD Score Regression (**Table 6.2** and **Supplementary Figure 6.3**). The λ_{GC} values ranged from 1.041 to 1.168, with the highest inflation observed for cholesterol in very large HDL (XL_HDL_C, λGC = 1.168), total lipids in very large HDL (XL_HDL_L, λGC = 1.159), and free cholesterol to total lipids ratio in very large HDL (XL_HDL_FC_pct, λGC = 1.154), which indicated low to moderate levels of inflation that were typical for polygenic traits analyzed in large-scale studies.[250] Correspondingly, LD Score Regression intercepts were mildly elevated for some metabolites, such as cholesterol in very large HDL (XL_HDL_C, intercept = 1.096, standard error (SE) = 0.015), average diameter for VLDL particles (VLDL_size, 1.095, SE = 0.016), and free cholesterol to total lipids ratio in very large HDL (XL_HDL_FC_pct, 1.092, SE = 0.014), suggesting that the observed inflation may be attributed to polygenicity rather than confounding.

6.3.2 Genetic instruments for metabolites in UK Biobank

Total observed SNP-based heritability (h^2) varied widely across metabolites. The highest heritability was observed for cholesterol in very large HDL (XL_HDL_C, $h^2 = 0.288$, SE = 0.048), followed by total lipids in very large HDL (XL_HDL_L, $h^2 = 0.278$, SE = 0.056), and free cholesterol to total lipids ratio in very large HDL (XL_HDL_FC_pct, $h^2 = 0.261$, SE = 0.062), indicating strong genetic contribution to these metabolites. In contrast, the lowest heritability was found for phenylalanine (Phe, $h^2 = 0.038$, SE = 0.009), glucose ($h^2 = 0.062$, SE = 0.007), and cholesteryl esters to total lipids ratio in chylomicrons and extremely large VLDL (XXL_VLDL_CE_pct, $h^2 = 0.055$, SE = 0.009), suggesting a comparatively smaller proportion of variance explained by common SNPs.

Table 6.2 Genomic control lambda (λ_{GC}), Linkage Disequilibrium Score Regression intercept, and heritability (h^2) estimates for 47 metabolites

ID	Metabolite	λ_{GC}	Linkage Disequilibrium Score Regression			
			Intercept	Standard error	Total observed h^2	Standard error
1	HDL_P	1.112	1.069	0.014	0.154	0.020
2	VLDL_size	1.144	1.095	0.016	0.220	0.031
3	Omega_3	1.104	1.057	0.012	0.226	0.095
4	PUFA_pct	1.132	1.055	0.014	0.159	0.016
5	LA_pct	1.121	1.054	0.011	0.150	0.030
6	Phe	1.041	1.020	0.008	0.038	0.009
7	Glucose	1.072	1.042	0.009	0.062	0.007
8	Creatinine	1.126	1.065	0.012	0.145	0.008
9	Albumin	1.087	1.056	0.012	0.084	0.006
10	GlycA	1.136	1.075	0.012	0.153	0.014
11	XXL_VLDL_L	1.126	1.067	0.015	0.175	0.021
12	XXL_VLDL_FC	1.128	1.066	0.015	0.189	0.024
13	XS_VLDL_C	1.077	1.052	0.014	0.142	0.034
14	XS_VLDL_CE	1.08	1.050	0.014	0.142	0.032
15	XL_HDL_L	1.159	1.089	0.014	0.278	0.056
16	XL_HDL_C	1.168	1.096	0.015	0.288	0.048
17	XL_HDL_FC	1.139	1.074	0.014	0.237	0.039
18	M_HDL_L	1.115	1.078	0.013	0.168	0.025
19	M_HDL_PL	1.106	1.073	0.014	0.158	0.023
20	S_HDL_P	1.091	1.041	0.012	0.115	0.014
21	S_HDL_L	1.098	1.045	0.013	0.126	0.015
22	S_HDL_PL	1.098	1.046	0.013	0.131	0.015
23	S_HDL_C	1.089	1.041	0.012	0.111	0.013
24	S_HDL_CE	1.087	1.037	0.012	0.115	0.017
25	XXL_VLDL_C_pct	1.055	1.031	0.010	0.063	0.012
26	XXL_VLDL_CE_pct	1.05	1.030	0.009	0.055	0.009
27	XL_VLDL_C_pct	1.124	1.076	0.015	0.168	0.016
28	XL_VLDL_CE_pct	1.124	1.075	0.015	0.168	0.017
29	XL_VLDL_TG_pct	1.105	1.067	0.014	0.144	0.014
30	M_VLDL_PL_pct	1.103	1.066	0.013	0.156	0.022
31	M_VLDL_C_pct	1.138	1.090	0.016	0.202	0.024
32	M_VLDL_CE_pct	1.143	1.092	0.016	0.210	0.025
33	M_VLDL_FC_pct	1.118	1.076	0.015	0.179	0.022
34	M_VLDL_TG_pct	1.131	1.084	0.015	0.194	0.023
35	S_VLDL_C_pct	1.107	1.075	0.015	0.164	0.024
36	S_VLDL_CE_pct	1.088	1.073	0.014	0.135	0.021
37	S_VLDL_FC_pct	1.127	1.075	0.015	0.204	0.031

38	S_VLDL_TG_pct	1.114	1.074	0.015	0.176	0.026
39	XS_VLDL_C_pct	1.143	1.083	0.015	0.228	0.032
40	XS_VLDL_CE_pct	1.148	1.088	0.016	0.235	0.029
41	XS_VLDL_TG_pct	1.138	1.085	0.015	0.224	0.031
42	M_LDL_CE_pct	1.078	1.044	0.013	0.106	0.015
43	S_LDL_PL_pct	1.086	1.055	0.013	0.124	0.018
44	S_LDL_CE_pct	1.088	1.056	0.013	0.130	0.020
45	XL_HDL_FC_pct	1.154	1.092	0.014	0.261	0.062
46	L_HDL_FC_pct	1.124	1.063	0.013	0.233	0.045
47	S_HDL_FC_pct	1.119	1.056	0.014	0.260	0.079

Abbreviations are in line with **Supplementary Table 3.1**.

6.3.3 Mendelian randomisation

The numbers of genetic instruments from the current study were generally higher than previous research (**Table 6.3**).[41] The highest number of genetic instruments was 157, which accounted for 9.3% of variance in cholesterol in very large HDL, and the lowest number of genetic instruments was 17, which explained 1.1% variance in phenylalanine. The corresponding F-statistics for these instruments ranged from 63.5 to 205.0, suggesting that there was little evidence of weak instrument bias. In addition, the number of genetic instruments for ischaemic stroke used in bidirectional MR was 23, which accounted for 0.08% of variance with F-statistic at 44.5.

Table 6.3 The proportions of variance in metabolites explained by the genetic instruments (R^2) and F-statistics

ID	Metabolite	N_{SNP}	R^2	F statistic	Former N_{SNP}*
1	HDL_P	113	5.5%	117.8	59
2	VLDL_size	143	8.5%	149.8	75
3	Omega_3	88	5.5%	152.0	42
4	PUFA_pct	89	3.9%	104.4	46
5	LA_pct	88	3.0%	80.0	38
6	Phe	17	1.1%	150.4	9
7	Glucose	50	1.7%	79.4	23
8	Creatinine	132	3.5%	63.5	57
9	Albumin	75	2.5%	78.2	27
10	GlycA	104	4.5%	103.1	57
11	XXL_VLDL_L	124	6.7%	132.4	59
12	XXL_VLDL_FC	123	8.4%	171.4	63
13	XS_VLDL_C	103	6.9%	165.1	55
14	XS_VLDL_CE	92	6.8%	182.2	57
15	XL_HDL_L	145	8.6%	149.2	83
16	XL_HDL_C	157	9.3%	150.9	88
17	XL_HDL_FC	126	8.3%	165.3	82
18	M_HDL_L	119	7.0%	145.7	68
19	M_HDL_PL	115	6.9%	149.1	63
20	S_HDL_P	80	3.7%	110.0	44
21	S_HDL_L	82	4.3%	127.0	38
22	S_HDL_PL	90	4.2%	112.3	49
23	S_HDL_C	80	4.0%	119.2	45
24	S_HDL_CE	73	4.0%	132.6	34
25	XXL_VLDL_C_pct	37	3.2%	205.0	21
26	XXL_VLDL_CE_pct	37	2.7%	172.8	20
27	XL_VLDL_C_pct	119	6.8%	140.2	65
28	XL_VLDL_CE_pct	123	6.4%	128.6	67
29	XL_VLDL_TG_pct	100	6.6%	163.9	56
30	M_VLDL_PL_pct	112	7.9%	177.1	63
31	M_VLDL_C_pct	141	8.8%	156.8	73
32	M_VLDL_CE_pct	145	8.9%	155.0	72
33	M_VLDL_FC_pct	128	7.9%	155.0	71
34	M_VLDL_TG_pct	136	8.2%	150.3	74
35	S_VLDL_C_pct	120	7.1%	146.5	63
36	S_VLDL_CE_pct	103	7.1%	171.6	63
37	S_VLDL_FC_pct	131	8.0%	152.4	60
38	S_VLDL_TG_pct	125	7.7%	153.6	66
39	XS_VLDL_C_pct	141	8.3%	147.9	74
40	XS_VLDL_CE_pct	148	8.5%	144.7	72
41	XS_VLDL_TG_pct	145	8.4%	145.5	75

ID	Metabolite	N_{SNP}	R^2	F statistic	Former N_{SNP}*
42	M_LDL_CE_pct	69	4.2%	147.1	41
43	S_LDL_PL_pct	83	6.1%	179.6	45
44	S_LDL_CE_pct	84	6.6%	193.7	42
45	XL_HDL_FC_pct	142	8.6%	151.8	79
46	L_HDL_FC_pct	120	7.5%	156.3	69
47	S_HDL_FC_pct	120	7.5%	156.0	73

Abbreviations are in line with **Supplementary Table 3.1**.

*The number of single nucleotide polymorphisms (SNPs) as genetic instruments was used by Harshfield et al. (2023) from the first release of the nuclear magnetic resonance data (N~120,000).

For results of inverse variance-weighted MR, there were no significant associations between metabolites and ischaemic stroke (all FDR-corrected p > 0.05), and significant heterogeneity was observed for all metabolites by Cochran's Q test (**Supplementary Table 6.6**, **Supplementary Table 6.7**, **Supplementary Figure 6.4**, and **Supplementary Figure 6.5**). MR-Egger method revealed 22 associations with directional pleiotropy (FDR-corrected p for intercept < 0.05), in which there were 17 significant associations, including 13 positive and 4 inverse associations. Of these, the strongest positive association was for genetically predicted levels of phospholipids to total lipids ratio in medium VLDL (OR = 1.23, 95% CI 1.11-1.37). Conversely, the strongest inverse association was for genetically predicted levels of triglycerides to total lipids ratio in very large VLDL (OR = 0.81, 95% CI 0.73-0.89). However, significant heterogeneity was still observed for all metabolites by the Rucker's Q statistics (**Supplementary Table 6.7**). For example, there were groups of specific variants with strong negative causal estimates, compared to the overall positive effects, on ischaemic stroke which immensely contributed to the observed heterogeneity for genetically predicted levels of phospholipids to total lipids ratio in medium VLDL and phospholipids to total lipids ratio in small LDL (**Supplementary Figure 6.5** ID-30 and ID-43**)**.

I therefore considered the results of other methods, as they gave robust and consistent results in the presence of heterogeneity (**Supplementary Table 6.6**).[268,271] Based on the weighted median method and weighted mode method, higher genetically predicted levels of cholesterol to total lipids ratio in chylomicrons and extremely large VLDL (OR = 1.17, 95% CI 1.06-1.29 in the weighted median method; 1.16, 1.07-1.25 in the weighted mode method) and cholesteryl esters to total lipids ratio in chylomicrons and extremely large VLDL (1.20, 1.08-1.32; 1.18, 1.08-1.30) were estimated to increase risk of ischaemic stroke, and higher genetically predicted levels of triglycerides to total lipids ratio in very large VLDL (0.89, 0.83-

0.96; 0.89, 0.84-0.95) were estimated to decrease risk of ischaemic stroke. According to the simple mode method, higher levels of genetically predicted levels of average diameter for VLDL particles were estimated to increase risk of ischaemic stroke (1.30, 1.12-1.52).

Across the observational study (**Chapter 4.**) and 5 different MR models (**Supplementary Figure 6.4**), evidence was directionally consistent for effects of genetically predicted levels of cholesterol to total lipids ratio in chylomicrons and extremely large VLDL and cholesteryl esters to total lipids ratio in chylomicrons and extremely large VLDL on higher risk of ischaemic stroke, and an effect of TG to total lipids ratio in very large VLDL on lower risk of ischaemic stroke. After the application of almost twofold larger metabolic data over the first release, the effect estimates became more precise and their directions remained consistent with those by Harshfield et al. (**Supplementary Figure 6.6**).[41] However, the estimates of average diameter for VLDL particles appeared less consistent for risk of ischaemic stroke across models. Higher average diameter for VLDL particles was associated with lower risk of ischaemic stroke based on simple mode estimators, but not on other methods.

The MR-PRESSO global test showed evidence of horizontal pleiotropy in all tests (**Supplementary Table 6.8**). The MR-PRESSO outlier test revealed false-positive relationships in 0.34% of putatively causal relationships (n = 17 of 5,017 total tests) using a threshold of $p < 0.05$. The MR-PRESSO distortion test was applied to evaluate the extent to which horizontal pleiotropy influenced causal effect estimates. Among the results with valid distortion coefficients, the median distortion was 11.2%, with a 5th to 95th percentile range from -127.8% to 204.7% (**Supplementary Table 6.9**). These indicated that while the majority of causal estimates were only modestly affected by the presence of pleiotropic outliers, a subset of results experienced substantial distortion, highlighting the importance of accounting

for horizontal pleiotropy in MR analyses. The MR-PRESSO outlier-corrected results showed no significant association (**Supplementary Table 6.10**).

6.3.4 Multivariable Mendelian randomisation

To disentangle the causal effects of highly correlated metabolites on ischaemic stroke, I also applied the recently developed MVMR-cML-SuSiE framework, which integrates constrained maximum likelihood estimation with Bayesian fine-mapping principles from the SuSiE model to improve robustness to pleiotropy and correlated exposures.[253] In Step 1, I applied univariable Mendelian randomisation with constrained maximum likelihood and model averaging (UVMR-cML-MA) to each of the 47 metabolites. All 47 metabolites showed significant associations with ischaemic stroke and were retained for downstream analysis. In Step 2, 2783 SNPs remained after harmonization and removal of invalid instruments using UVMR-cML-BIC across 47 metabolites, out of an initial pool of 7473 genome-wide significant SNPs identified in the univariable screening step. This joint instrument set was used in Step 3 to fit an iterative SuSiE algorithm on the profile likelihood, identifying one dominant signal cluster. Five metabolites surpassed the PIP threshold of 1/47 (2.13%; **Supplementary Table 6.11**), indicating their likely contribution to the single causal cluster, including total lipids in medium HDL (36.14%), phospholipids in medium HDL (25.93%), concentration of HDL particles (25.03%), cholesteryl esters to total lipids ratio in chylomicrons and extremely large VLDL (PIP = 6.82%), and free cholesterol to total lipids ratio in very large HDL (2.99%). To validate these cluster findings using a more conventional MR framework, the multivariable Mendelian randomisation with inverse variance-weighted estimation[283] was applied on the five selected metabolites. Only cholesteryl esters to total lipids ratio in chylomicrons and extremely large VLDL demonstrated a statistically significant direct association with ischaemic stroke (beta = 0.063, SE = 0.015, $p = 3.4 \times 10^{-5}$;

Supplementary Table 6.12), while the other four metabolites did not meet nominal significance.

6.3.5 Secondary analyses

Based on the reverse MR, there was little evidence that higher genetic liability to ischaemic stroke was estimated to influence the genetically predicted levels of individual metabolites (all FDR-corrected $p > 0.05$ for all MR methods; **Supplementary Table 6.11 Results of posterior inclusion probabilities estimated via the iterative Sum of Single Effects (SuSiE) algorithm**, **Supplementary Table 6.14**, and **Supplementary Figure 6.7**).

6.4 Discussion

Based on a variety of pleiotropy-robust MR methods[268] and the single largest NMR metabolite GWAS data to date, evidence for causality of the associations with ischaemic stroke was supported for cholesterol to total lipids ratio in chylomicrons and extremely large VLDL and cholesteryl esters to total lipids ratio in chylomicrons and extremely large VLDL (**Supplementary Figure 6.1**). The genetically predicted levels of these two metabolites exhibited consistent direction of effect across multiple MR methods and significant results after multiple testing correction. Among these, the latter further demonstrated a statistically significant direct effect on ischaemic stroke in the MVMR-cML-SuSiE and the confirmatory multivariable MR with inverse-variance weighted estimation. The significant association was also observed for genetically predicted TG to total lipids ratio in very large VLDL, though this association was more suggestive, given model discrepancies in directionality. Little evidence supported that genetic liability to ischaemic stroke would influence the levels of the metabolites in this study, thereby reducing concerns of reverse causality.

The initial GWAS using REGENIE regression identified multiple independent variants significantly associated with each of 47 metabolites, including cis-acting variants located within 1 Mb of transcription start sites, suggesting possible direct regulatory effects. Heritability estimates and genomic inflation factors revealed that several metabolic traits, such as cholesterol in very large HDL, total lipids in very large HDL, and free cholesterol to total lipids ratio in very large HDL, are highly polygenic, with SNP-based heritability estimates exceeding 25%, consistent with strong genetic influence.[284] These findings highlight notable differences in polygenic architecture and genetic contribution among the 47 metabolites. The observed inflation is likely driven by true polygenic effects rather than confounding, particularly given the heritable biology and strong genetic underpinnings of these metabolites.[285] These results highlight the strong genetic contribution to lipid metabolism.[285]

The identification of novel genetic variants associated with cholesteryl esters to total lipids ratio in chylomicrons and extremely large VLDL provides important insights into the biological mechanisms linking the metabolite to the risk of ischaemic stroke (**Supplementary Table 6.2** and **Supplementary Table 6.3**), such as rs117857195 in the mapped gene *LPA*, rs1800789 in the mapped genes *PLRG1* and *FGB*, rs9901675 in the mapped gene *CD68*, rs3770586 in the mapped gene *ABCB11*, rs13081418 in the mapped gene *NBEAL2*, rs3803286 in the mapped gene *TRAF3*, rs1247304, and rs72913032. Several variants map to loci with well-established roles in lipid metabolism, vascular inflammation, and thrombosis. Variants annotated to *LPA* (e.g. rs117857195) implicate lipoprotein(a) biology, which has been strongly associated with increased risk of atherosclerosis and ischaemic stroke through both lipid deposition and prothrombotic pathways.[286] Variants near the *FGB* gene (e.g. rs1800789 in the fibrinogen beta chain gene) are associated with fibrinogen regulation,[287,288] suggesting a potential link between elevated fibrinogen levels, altered coagulation, and stroke

susceptibility. Variants mapped to *CD68* (e.g. rs9901675) point to a role for macrophage activation and foam cell formation,[289] which are critical processes in plaque development and progression. Signals mapped to *ABCB11* (e.g. rs3770586), a bile salt export pump, link bile acid transport to cholesterol homeostasis,[290] a pathway known to modulate plasma lipids and lipoprotein profiles. Variants associated with *NBEAL2* (e.g. rs13081418) are involved in platelet α-granule biogenesis,[291] which may influence platelet activation and thrombus formation. And variants mapped to *TRAF3* (e.g. rs3803286) suggest a potential role of vascular inflammatory signaling pathways in atherosclerotic processes.[292] Collectively, these findings suggest that cholesteryl-ester-enriched remnant particles from chylomicrons and very large VLDL are likely to be retained within the arterial wall, promoting lipid accumulation, endothelial dysfunction, and enhanced thrombotic potential.[293] Together with multivariable MR results showing a direct association between genetically predicted levels of cholesteryl esters to total lipids ratio in chylomicrons and extremely large VLDL and ischaemic stroke, these observations support a mechanism involving the interplay of remnant cholesterol metabolism, macrophage-driven plaque biology, bile acid transport, platelet function, and vascular inflammation, highlighting potential therapeutic targets for stroke prevention. More studies are needed to validate these findings before they can be proposed as drug targets.

The suggestive causal association of TG to total lipids ratio in very large VLDL with lower ischaemic stroke risk is counterintuitive because extensive evidence from observational studies has demonstrated that hypertriglyceridemia is a risk factor for ischaemic stroke.[294] Numerous metabolites are products and substrates of enzymatic reactions, so understanding causal relationship of the ratio of substrate to product can indeed shed light on the biological mechanism that cannot be discerned when delving into only a single metabolite.[295] Nevertheless, these metabolite ratios have not been extensively explored in previous

studies.[24,46,47] As an example, the much smaller two-sample MR study by Si et al.[82] selected summary statistics for only 15 directly measured lipids from 24,925 individuals with European ancestry[38] and summary statistics for ischaemic stroke from the International Stroke Genetics Consortium with 29,633 participants.[296] There was no causal relationship between the 15 lipid traits and ischaemic stroke after multiple testing correction, although the authors used two consistent and nominally significant results ($p < 0.05$) in two different MR methods to claim the protective causal effect of TG in very large, and chylomicrons and extremely large VLDL on ischaemic stroke. It has been suggested that the substantial TG transported by large lipoproteins (i.e. VLDL) cannot infiltrate the arterial wall, which protect individuals from atherogenesis.[171-173] This proposition gained further support from the fact that more TG could be carried by very large VLDL than LDL cholesterol,[173] which helped lessen the load of LDL cholesterol and reduce the risk of ischaemic stroke.

However, it must be emphasized that considerable pleiotropy was observed, indicating that some genetic variants were related to various metabolites through multiple pathways.[297] Using MR-PRESSO, considerable distortions in the causal estimates were evident in the presence of significant horizontal pleiotropy for all metabolites. This may result from three reasons. First, there were notable correlations between metabolites. Second, specific variants with outlying causal estimates hugely contributed to the observed heterogeneity (**Supplementary Figure 6.5**).[298] Third, the variance among the variant-specific causal estimates was beyond that anticipated by chance, but various genetic variants shared the load of heterogeneity (over-dispersion).[298] This introduces challenges to evaluate whether the aforementioned causal inference is authentic or spurious, so the results need to be judiciously interpreted. Compared to traditional MR methods, MVMR-cML-SuSiE offers substantially improved robustness to pleiotropy and correlated metabolites.[253] Inverse variance-weighted

method and MR-Egger method are highly sensitive to multicollinearity, often resulting in biased or inefficient estimates.[299] MR-Egger method further suffers from low power and inflated Type I error rates.[299] In contrast, MVMR-cML-SuSiE simultaneously handles invalid instruments and multicollinearity, ensuring valid inference even in complex settings.[253] By decomposing the profile likelihood into single-effect components and evaluating PIPs for each metabolite, this method avoids arbitrary variable selection and a "tunnel vision" perspective.[253] Importantly, it highlights the novel finding that genetically predicted levels of cholesteryl esters in chylomicrons and extremely large VLDL as a promising modifiable pathway in the aetiology of ischaemic stroke. Furthermore, the GWAS using the REGENIE model substantially increased the number of genetic instruments compared to previous studies (**Table 6.3**), including multiple cis-acting variants with regulatory potential. This enhanced instrumental strength enabled more precise effect estimates and contributed to the improved power of the MR analyses. Finally, reverse MR analyses found no evidence that genetic liability to ischaemic stroke influences metabolic levels, supporting the unidirectionality of the observed causal pathways. Together, these novel findings advance our understanding of the metabolic determinants of ischaemic stroke, highlight specific lipid subfractions as potential therapeutic targets, and underscore the value of integrating multivariable and fine-mapping MR approaches in high-dimensional metabolomics research.

There are some noteworthy limitations. First, the analyzed population was from White British ancestry, so results might not be generalisable to other populations. Second, the exposures were baseline NMR metabolites, so I did not examine how the genetic variant affected the distribution of the metabolites across the life course and whether the metabolites had an influence on the different stages of ischaemic stroke. If the genetic relationships with the metabolites change over time, then MR estimates derived from genetic relationships with

the metabolites measured at a single time point can be inaccurate.[300] Compared to clinical trials that often give evidence based on a rather short intervention course, genotype-attributable differences in metabolic levels discerned in MR analyses represent the long-lasting alterations in metabolic exposures and tend to be larger than those from clinical trials. Third, MR relies on key assumptions, including the absence of horizontal pleiotropy, meaning that genetic instruments should not affect the outcome through pathways other than the exposure of interest. Although MVMR-cML-SuSiE and sensitivity analyses were applied to detect and minimise pleiotropic bias, these approaches rely on their own assumptions and cannot completely exclude violations. Future work would benefit from systematically screening genetic instruments using resources such as PhenoScanner to identify variants associated with potential confounders and to strengthen the validity of causal inference. Forth, the samples for providing genetic instruments for ischaemic stroke included individuals from UKB, the same sample from which genetic association estimates were calculated for metabolic traits (~19% overlap). This may introduce bias in the direction of the observational association, however, given the robust genetic instruments for metabolic traits, bias from sample overlap is quite small.[40,301] This gained further support from the consistent results of sensitivity analyses by weighting the genetic instruments according to a sub-sample of the UKB and excluding cases included in GIGASTROKE.[261]

Clinically robust conclusions and suggestions are not obtained from one cohort study but rather are driven by concordant results of various cohorts and, critically, by various study designs which pick up diverse sources of bias.[302] Triangulation of evidence using randomised controlled trials would be helpful to re-estimate the effects tested here in other scenarios, in the distinct life stages, in other populations, and using different methods to measure metabolites. Despite the fact that UKB possesses the biggest dataset to date, replication of this

research in another big cohort study and employment of various measurements will permit more robust metabolic characterisation of ischaemic stroke. In terms of the clinical implications of my work, the findings illustrate the new insights into the metabolic pathways underlying ischaemic stroke. This could help provide the development of novel therapies or dietary interventions since controlling concentrations of certain metabolites might help abate the risk of ischaemic stroke. Furthermore, these findings could be used by clinicians to predict which persons have higher chances to develop ischaemic stroke and provide individualised treatment strategies.

6.5 Conclusion

The results of this large MR study using new high-throughput metabolic data from UKB support that elevated genetically predicted levels of cholesteryl esters to total lipids ratio in chylomicrons and extremely large VLDL and cholesterol to total lipids ratio in chylomicrons and extremely large VLDL increased the risk of ischaemic stroke, and suggest that genetically predicted levels of TG to total lipids ratio in very large VLDL was estimated to decrease ischaemic stroke. Such causal effects were not bidirectional, as there was little evidence that genetic liability to ischaemic stroke would increase or decrease the metabolites in this study. To translate these findings into clinical practice, further research is needed, particularly intervention trials, to validate the causal mechanisms and assess the potential for metabolite-targeted prevention or treatment strategies. Given the pervasive challenge of horizontal pleiotropy in metabolic genetics, my analyses combined multiple approaches to mitigate bias, including pleiotropy-robust univariable estimators (MR-Egger, weighted median, mode-based methods), outlier detection and correction (MR-PRESSO), and multivariable MR using the cML-SuSiE framework to handle correlated exposures and invalid instruments. Nonetheless, additional steps could further strengthen causal inference. First,

prioritising instruments that are cis to biologically relevant genes for each metabolite and using fine-mapping can reduce pleiotropy by anchoring variants to mechanistic loci. Second, within-family MR or sibling-based designs, where available, can attenuate confounding from population structure and dynastic effects. Third, extending multivariable MR to include additional metabolic classes and inflammatory markers may help partition direct from mediated pathways. Finally, improving instrument strength through larger discovery GWAS and meta-analyses across ancestries will increase power and reduce weak-instrument bias, while drug-target MR and recall-by-genotype or challenge studies can triangulate biological plausibility. Collectively, these methodological refinements, alongside external validation in independent cohorts and prospective trials, will help determine which metabolite targets are genuinely causal and clinically actionable for ischaemic stroke prevention.

Chapter 7. Discussion

Chapter summary

This chapter synthesises the main findings of the thesis, highlighting contributions, strengths, limitations, and implications.

The thesis had three aims: (1) to evaluate associations between nuclear magnetic resonance metabolites and incident ischaemic stroke, (2) to assess the predictive value of metabolites beyond the Framingham Stroke Risk Profile, and (3) to test causal relationships using Mendelian randomisation. Prospective analyses in UK Biobank identified VLDLs, LDLs, HDLs, phenylalanine, and albumin associated with ischaemic stroke risk, independent of conventional confounders. Machine learning approaches revealed non-linear associations and thresholds but showed little improvement in discrimination or reclassification compared with established predictors. Mendelian randomisation, leveraging stronger genetic instruments from an expanded genome-wide association study, provided evidence for causal roles of lipid ratios in chylomicrons and very large VLDL, while identifying nearly half of significant associations as novel.

Strengths include the scale of UK Biobank, linkage to health records, advanced analytic methods, and triangulation across study designs. Limitations include volunteer bias, limited ethnic diversity, and the lipid-focused scope of nuclear magnetic resonance profiling.

These findings extend knowledge of the metabolic architecture of ischaemic stroke, suggesting potential biological pathways but limited clinical prediction benefit. Replication in diverse populations and interventional studies will be needed to inform translation.

7.1 Main findings

Reliable identification of metabolic biomarkers of ischaemic stroke risk provides important evidence for effective preventive and therapeutic strategies, but the current literature is limited by absence of robust replications, small sample sizes, and lack of causal evidence. This thesis addresses these gaps in knowledge, and the three major objectives were:

(1) To evaluate the association of individual metabolites with incident ischaemic stroke using conventional statistical methods in UKB and compare the findings with those from a nested case-control study within the CKB and a consortium of 7 studies on metabolomics and stroke.

(2) To assess to what extent metabolites can jointly predict incident ischaemic stroke with machine learning algorithms and evaluate the additive predictive value of metabolites over risk factors in FSRP.

(3) To ascertain the causal relationship of above metabolic findings with incident ischaemic stroke using MR.

The characteristics of the metabolic exposures and their associations with potential confounding factors were investigated in **Chapter 3.**. Prospective associations of plasma NMR metabolites with ischaemic stroke were addressed in **Chapter 4.**. The potential benefit of adding metabolic profiling to risk factors in machine learning to predict ischaemic stroke was addressed in **Chapter 5.**. The Boruta algorithm was also applied to find the relevant metabolites for ischaemic stroke. Finally, based on the identified metabolites, the third objective to investigate causality of observed associations was addressed in **Chapter 6.**.

7.1.1 Description of NMR metabolites by key socio-demographic and lifestyle characteristics

The analyses contained 249 NMR metabolites from the baseline plasma samples of the UKB participants. Measurements of biomarkers quantified by both NMR and clinical chemistry were largely comparable (**Chapter 3.**). The NMR metabolites were highly correlated with each other and with selected lifestyle and socio-demographic characteristics at baseline, which manifested the physiological and biochemical nature of each metabolite as well as the pathophysiological and epidemiological associations between them. The levels of most metabolites were associated with a variety of demographic factors, behavioural factors, and pre-existing medical conditions (e.g. age, sex, smoking status, alcohol intake frequency, dietary variables, prevalent diabetes, and SBP), which recapitulated the current literature. Importantly, these determinants were thought to be potential confounders of the associations between NMR metabolites and ischaemic stroke risk investigated throughout the thesis.

7.1.2 Epidemiological associations of individual metabolites with incident ischaemic stroke

The prospective study was conducted among 184,564 participants with available NMR metabolic measurements and without CVD at baseline, of which 3,002 had an incident ischaemic stroke event during a median follow-up of 13.8 years. Incident ischaemic stroke was related to 22 VLDLs, 3 LDLs, 13 HDLs, aromatic amino acid phenylalanine, and fluid balance-related albumin. These associations were independent of numerous confounders, such as ethnicity, lipid modifying medication, smoking status, alcohol intake, dietary supplement use and 13 dietary variables, physical activity, BMI, prevalent diabetes and SBP. After controlling for the same covariates as conducted in other cohorts, the directions of associations in UKB were more consistent

with the consortium of 7 studies on metabolomics and stroke (80.8% directionally consistent) than with CKB (30.2%; **Chapter 4.**).

7.1.3 Predictive value of NMR metabolites for ischaemic stroke

The additional benefit of adding NMR metabolites to FSRP, which is an established algorithm for the prediction of ischaemic stroke, was examined using machine-learning approaches including XGBoost and RSF (**Chapter 5.**). The machine learning dealt with highly correlated metabolites and took the complicated interactions between metabolites and ischaemic stroke into account. It also contributed insight into the predictive value of specific metabolites that have been overlooked in the conventional analyses under a linear assumption (e.g. creatinine, glucose, omega-3 FA, and glycoprotein acetyls). Four Boruta-selected metabolites in XGBoost overlapped with those in the top 20 important metabolites in an earlier analysis using a neural network,[67] including albumin, creatinine, glycoprotein acetyls, and glucose. In addition, the estimated cut-off points of metabolic levels for ischaemic stroke can be inferred from the SHAP dependence plots. However, compared with FSRP, there was little evidence of improvement in prediction of ischaemic stroke after adding the NMR metabolites, which was in accordance with former literature.[46,67]

7.1.4 Causal relationships between the metabolites and ischaemic stroke in Mendelian randomisation analysis

The mega-scale two-sample MR analyses using the most updated high-throughput NMR metabolic data from UKB and summary statistics of ischaemic stroke from the GIGASTROKE Consortium substantially enhanced the strength of the genetic instrumental variables (**Chapter 6.**). Compared to the existing literature with limited sample sizes,[41,82] this work greatly improved instrumental strength. Over 6100 SNP-

metabolite associations were identified, nearly half of which were novel. The average F-statistics ranged from 63.5 to 205.0, indicating strong instruments and low risk of weak instrument bias. Heritability estimates for several lipoprotein traits exceeded 25 percent, and genomic inflation statistics remained modest, suggesting valid and powerful instruments. MR analyses demonstrated that higher genetically predicted levels of cholesteryl esters to total lipids ratios and cholesterol to total lipids ratios in chylomicrons and extremely large VLDL were associated with increased risk of ischaemic stroke. These associations were consistent across various MR approaches. In contrast, higher genetically predicted levels of TG to total lipids ratio in very large VLDL were associated with a reduced risk of ischaemic stroke. The Multivariable Mendelian Randomisation with Constrained Maximum Likelihood and Sum of Single Effects (MVMR-cML-SuSiE) method confirmed a direct effect of genetically predicted levels of cholesteryl esters to total lipids ratio in chylomicrons and extremely large VLDL on ischaemic stroke. Genetic variants underlying these associations mapped to biologically relevant genes including *LPA*, *FGB*, *CD68*, *ABCB11*, *NBEAL2*, and *TRAF3*, implicating mechanisms involving lipoprotein metabolism, inflammation, platelet function, and coagulation. Reverse MR analysis showed limited evidence that genetic liability to stroke influenced metabolite levels.

7.2 Strengths and limitations

This thesis has several major strengths. It is the largest observational study with this depth of phenotyping that has contributed to the understanding of the link between NMR metabolites and ischaemic stroke with the application of a multi-disciplinary approach combining conventional, molecular and genetic epidemiology. To describe metabolic profiles and their impacts on ischaemic stroke risk, the updated metabolic data of UKB allowed better precision of estimates and enhanced ability to assess for

bias (such as confounding and reverse causality). Moreover, the linkage to nationwide death registries and electronic health records allowed reliable definitions of ischaemic stroke events and restricted the loss of follow-up. The results of two machine learning approaches that accounted for multicollinearity and interactions among metabolites, as well as non-linear associations, genuinely reflected the additive predictive information of all metabolites over conventional risk factors. The findings from randomised controlled trials (RCTs) were viewed as the top in the hierarchy of causal evidence, they suffered from a number of drawbacks such as small sample size, short follow-up time, and difficult blinding. Blinding can be particularly difficult for lifestyle or biomarker-modifying interventions (e.g. diet, exercise, or metabolic supplements), where participants and investigators are often aware of treatment allocation, potentially introducing expectation bias. This thesis provides a valuable original source of research by triangulating results in the context of various study designs, including observational studies, machine learning, and MR. Furthermore, the findings from the MR analysis were less susceptible to reverse causality and confounding bias, which situate their ranking of causal evidence between interventional trials and conventional observational epidemiology.[266]

Before application in clinical scenarios, considerable challenges remain. The unrepresentative makeup of UKB with 5.5% initial response rate (e.g. lower current smokers, 11% in UKB vs. 19% in the UK population.[303,304] However, the generalisability of the relationships between exposures and outcome is supported by a sufficiently great number of participants with a wide range of factors that allowed for the assessment of confounding and effect modification in important population subgroups.[184,305,306] Furthermore, while the NMR measurement is cheaper and more robust than MS-based alternatives, its sensitivity is lower. Also, the relatively narrow and lipid focused

coverage of metabolic biomarkers limited explorations of other possible mechanisms to ischaemic stroke. In addition, fasting before the blood test was not required, and levels of triglycerides, total cholesterol, and LDL cholesterol would be influenced.[219] However, the fasting/postprandial state accounted for less than 2% of the total variance in the NMR spectroscopy-based plasma metabolite quantification,[220] and fasting time was taken into account in the Cox regression models and GWAS. Finally, although MR is a valuable method of assessing causality of observational associations, the interpretation of its results is challenging especially when the instrumental variables might have pleiotropic effects and the exposures are highly correlated.

7.3 Implications for population health

The wealth of electronic health data linkage, metabolic and genetic data in UKB gave an unparalleled resource for illustrating early molecular signatures of ischaemic stroke. NMR measurements enabled a broader scope of metabolites beyond the range of the clinical chemistry methods. In **Chapter 4.,** early metabolic biomarkers that associated with ischaemic stroke were identified, and these biomarkers therefore allowed risk stratification improvement. The 5% (phenylalanine) to 9% (free cholesterol in very large HDL) higher risks and 5% (triglycerides to total lipids ratio in small VLDL) to 13% (albumin) lower risks of ischaemic stroke demonstrated the potential impact of NMR metabolites on ischaemic stroke. Some associations were known for lipoprotein particle concentration (i.e. concentration of HDL particles) and fluid balance (i.e. albumin), but associations had not previously been shown for numerous circulating detailed lipoprotein levels. These findings imply that the high-throughput NMR approach provides an avenue for discovering novel biomarkers of ischaemic stroke and elucidating its pathological mechanisms. Although metabolites captured the genetic, lifestyle and environmental effects on ischaemic stroke, **Chapter 5.** demonstrated that

the circulating NMR metabolites did not add predictive value over conventional risk factors. These findings indicate that the prediction of ischaemic stroke as part of national screening programmes is unlikely to be improved by the addition of currently measured NMR metabolites.

The essentially random allocation of alleles that were related to circulating metabolic levels, and the relationship between genotypes and risk of ischaemic stroke, were applied to provide an unbiased estimate of the strength of causal associations between metabolites and ischaemic stroke (**Chapter 6.)**. This approach suggested that the 1% higher genetically predicted cholesterol to total lipids ratio in chylomicrons and extremely large VLDL and cholesteryl esters to total lipids ratio in chylomicrons and extremely large VLDL produced 16% and 20% higher risks of ischaemic stroke, respectively. Conversely, 1% higher genetically predicted TG to total lipids ratio in very large VLDL led to 11% lower risk of ischaemic stroke. The causal roles of metabolite ratios indicated that indirect measurement of lipoprotein concentrations was more informative than direct measurement with regard to ischaemic stroke. The results also suggested that maintaining adequate proportions of lipids might be one approach for preventing ischaemic stroke. The MR method served as a valuable low-cost and time-efficient first step in estimating both the efficacy and possible beneficial response of an intervention before implementing a RCT. Because therapeutic interventions for ischaemic stroke may result in better efficacy when implemented in the early stages of disease, these specific metabolites, which were measured up to a decade prior to formal diagnoses, could strengthen personalised intervention strategies and inform trial recruitment. However, these results alone cannot establish causality; replication in further MR analyses and complementary evidence from other study designs are required. For example, an RCT of lipid-modifying interventions that specifically alter

cholesteryl ester to total lipid ratios in chylomicrons and extremely large VLDL would be needed to confirm causality and test clinical applicability. Finally, the non-significant associations between ischaemic stroke and the NMR metabolites observed in this thesis need further investigation in different cohorts to determine causality.

7.4 Future research

Recent decades have seen technological advances which enable the integration of various molecular levels of data in the omics era. Measurements of NMR metabolites and protein biomarkers have been underway among the whole UKB cohort. The information inferred from the abundant omics quantification will become a prism for understanding both the pathophysiology and the prevention of ischaemic stroke. Although UKB is the biggest cohort with NMR metabolic data, the lack of ethnic diversity makes it hard to investigate associations in other ancestral subgroups. The replication of MR findings in different ancestries will provide evidence for the generalisability of ischaemic stroke related metabolites. Studies in different geographical and ethnic populations will complement analyses of the UKB, expanding generalisability of the findings. In addition, including both healthy individuals from primary care and high-risk patients from secondary care will improve representativeness and allow for more robust identification of metabolic signatures of ischaemic stroke risk. This dual approach would not only strengthen causal inference but also enhance the clinical translation of findings, supporting both early risk stratification in the general population and targeted prevention in high-risk groups. Future research would also be important to evaluate the etiological subtypes of ischaemic stroke, such as cardioembolism, large artery atherosclerosis, small vessel disease, or other determined cause (e.g. mitochondrial disorder, genetic mutation).[4] Furthermore, future studies would benefit from introducing large studies in diverse populations with multiple types

of biomarkers (including the metabolites, proteomics and gut microbiome) in the same individuals. If the above issues are addressed, these procedures will contribute to crucial biomarkers at multiple healthcare stages to address the risk of ischaemic stroke, with the opportunities to transform ischaemic stroke risk management and improve the life expectancy.

7.5 Conclusion

The associations of circulating NMR metabolites with risk of ischaemic stroke have been robustly and reliably investigated across mid-to-later life in this large-scale prospective study. Independent of conventional risk factors, phenylalanine, VLDLs, LDLs, HDLs, and albumin remained significantly associated with risk of ischaemic stroke. However, compared with the established FSRP, adding NMR metabolites did not show improvement in the prediction of ischaemic stroke. Three VLDL ratios showed evidence of causal relationships with ischaemic stroke, which indicated that indirect measurement of lipoprotein concentrations may be of greater value than their direct measurement. Future studies should focus on prediction and causality using a wider range of metabolites on ischaemic stroke subtypes in more diverse populations.

Reference

1. 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* 2021; **20**(10): 795-820.
2. Hankey GJ. Stroke. *Lancet* 2017; **389**(10069): 641-54.
3. Maida CD, Norrito RL, Daidone M, Tuttolomondo A, Pinto A. Neuroinflammatory Mechanisms in Ischemic Stroke: Focus on Cardioembolic Stroke, Background, and Therapeutic Approaches. *Int J Mol Sci* 2020; **21**(18).
4. Adams HP, Jr., Bendixen BH, Kappelle LJ, et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke* 1993; **24**(1): 35-41.
5. Cannistraro RJ, Badi M, Eidelman BH, Dickson DW, Middlebrooks EH, Meschia JF. CNS small vessel disease: A clinical review. *Neurology* 2019; **92**(24): 1146-56.
6. Campbell BCV, De Silva DA, Macleod MR, et al. Ischaemic stroke. *Nat Rev Dis Primers* 2019; **5**(1): 70.
7. Lisabeth LD, Brown DL, Hughes R, Majersik JJ, Morgenstern LB. Acute stroke symptoms: comparing women and men. *Stroke* 2009; **40**(6): 2031-6.
8. Patil S, Rossi R, Jabrah D, Doyle K. Detection, Diagnosis and Treatment of Acute Ischemic Stroke: Current and Future Perspectives. *Front Med Technol* 2022; **4**: 748949.
9. Ding Q, Liu S, Yao Y, Liu H, Cai T, Han L. Global, Regional, and National Burden of Ischemic Stroke, 1990-2019. *Neurology* 2022; **98**(3): e279-e90.
10. O'Donnell MJ, Chin SL, Rangarajan S, et al. Global and regional effects of potentially modifiable risk factors associated with acute stroke in 32 countries (INTERSTROKE): a case-control study. *Lancet* 2016; **388**(10046): 761-75.
11. Diener HC, Hankey GJ. Primary and Secondary Prevention of Ischemic Stroke and Cerebral Hemorrhage: JACC Focus Seminar. *J Am Coll Cardiol* 2020; **75**(15): 1804-18.
12. Organization WH. Stroke, Cerebrovascular accident. https://www.emro.who.int/health-topics/stroke-cerebrovascular-accident/index.html (accessed 17th January 2024).
13. Kopin L, Lowenstein C. Dyslipidemia. *Ann Intern Med* 2017; **167**(11): Itc81-itc96.
14. Rodriguez-Colon SM, Mo J, Duan Y, et al. Metabolic syndrome clusters and the risk of incident stroke: the atherosclerosis risk in communities (ARIC) study. *Stroke* 2009; **40**(1): 200-5.
15. Mihaylova B, Emberson J, Blackwell L, et al. The effects of lowering LDL cholesterol with statin therapy in people at low risk of vascular disease: meta-analysis of individual data from 27 randomised trials. *Lancet* 2012; **380**(9841): 581-90.
16. Gordon DJ, Probstfield JL, Garrison RJ, et al. High-density lipoprotein cholesterol and cardiovascular disease. Four prospective American studies. *Circulation* 1989; **79**(1): 8-15.
17. Mackness MI, Abbott C, Arrol S, Durrington PN. The role of high-density lipoprotein and lipid-soluble antioxidant vitamins in inhibiting low-density lipoprotein oxidation. *Biochem J* 1993; **294 (Pt 3)**(Pt 3): 829-34.
18. Woolf N. Pathology of atherosclerosis. *Br Med Bull* 1990; **46**(4): 960-85.

19. Parmar P, Krishnamurthi R, Ikram MA, et al. The Stroke Riskometer(TM) App: validation of a data collection tool and stroke risk predictor. *Int J Stroke* 2015; **10**(2): 231-44.
20. Hippisley-Cox J, Coupland C, Brindle P. Derivation and validation of QStroke score for predicting risk of ischaemic stroke in primary care and comparison with other risk scores: a prospective open cohort study. *Bmj* 2013; **346**: f2573.
21. Wolf PA, D'Agostino RB, Belanger AJ, Kannel WB. Probability of stroke: a risk profile from the Framingham Study. *Stroke* 1991; **22**(3): 312-8.
22. Dufouil C, Beiser A, McLure LA, et al. Revised Framingham Stroke Risk Profile to Reflect Temporal Trends. *Circulation* 2017; **135**(12): 1145-59.
23. Sacco RL, Benson RT, Kargman DE, et al. High-density lipoprotein cholesterol and ischemic stroke in the elderly: the Northern Manhattan Stroke Study. *Jama* 2001; **285**(21): 2729-35.
24. Bragg F, Trichia E, Aguilar-Ramirez D, Bešević J, Lewington S, Emberson J. Predictive value of circulating NMR metabolic biomarkers for type 2 diabetes risk in the UK Biobank study. *BMC Med* 2022; **20**(1): 159.
25. Montaner J, Ramiro L, Simats A, et al. Multilevel omics for the discovery of biomarkers and therapeutic targets for stroke. *Nat Rev Neurol* 2020; **16**(5): 247-64.
26. Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat Rev Mol Cell Biol* 2016; **17**(7): 451-9.
27. Schrimpe-Rutledge AC, Codreanu SG, Sherrod SD, McLean JA. Untargeted Metabolomics Strategies-Challenges and Emerging Directions. *J Am Soc Mass Spectrom* 2016; **27**(12): 1897-905.
28. Shah SH, Kraus WE, Newgard CB. Metabolomic profiling for the identification of novel biomarkers and mechanisms related to common cardiovascular diseases: form and function. *Circulation* 2012; **126**(9): 1110-20.
29. Hagenbeek FA, Pool R, van Dongen J, et al. Heritability estimates for 361 blood metabolites across 40 genome-wide association studies. *Nat Commun* 2020; **11**(1): 39.
30. Wishart DS, Guo A, Oler E, et al. HMDB 5.0: the Human Metabolome Database for 2022. *Nucleic Acids Res* 2022; **50**(D1): D622-d31.
31. Sidorov E, Sanghera DK, Vanamala JKP. Biomarker for Ischemic Stroke Using Metabolome: A Clinician Perspective. *J Stroke* 2019; **21**(1): 31-41.
32. Posti JP, Dickens AM, Orešič M, Hyötyläinen T, Tenovuo O. Metabolomics Profiling As a Diagnostic Tool in Severe Traumatic Brain Injury. *Front Neurol* 2017; **8**: 398.
33. Au A. Metabolomics and Lipidomics of Ischemic Stroke. *Adv Clin Chem* 2018; **85**: 31-69.
34. Lindon JC, Nicholson JK. Spectroscopic and statistical techniques for information recovery in metabonomics and metabolomics. *Annu Rev Anal Chem (Palo Alto Calif)* 2008; **1**: 45-69.
35. Soininen P, Kangas AJ, Würtz P, Suna T, Ala-Korpela M. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circ Cardiovasc Genet* 2015; **8**(1): 192-206.
36. Würtz P, Kangas AJ, Soininen P, Lawlor DA, Davey Smith G, Ala-Korpela M. Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Large-Scale Epidemiology: A Primer on -Omic Technologies. *Am J Epidemiol* 2017; **186**(9): 1084-96.

37. Aguilar-Ramirez D, Herrington WG, Alegre-Díaz J, et al. Adiposity and NMR-measured lipid and metabolic biomarkers among 30,000 Mexican adults. *Commun Med (Lond)* 2022; **2**(1): 143.
38. Kettunen J, Demirkan A, Würtz P, et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. *Nat Commun* 2016; **7**: 11122.
39. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003; **32**(1): 1-22.
40. Smith ML, Bull CJ, Holmes MV, et al. Distinct metabolic features of genetic liability to type 2 diabetes and coronary artery disease: a reverse Mendelian randomization study. *EBioMedicine* 2023; **90**: 104503.
41. Harshfield EL, Markus HS. Association of Baseline Metabolomic Profiles With Incident Stroke and Dementia and With Imaging Markers of Cerebral Small Vessel Disease. *Neurology* 2023; **101**(5): e489-e501.
42. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife* 2018; **7**.
43. Bocher O, Singh A, Huang Y, et al. Disentangling the consequences of type 2 diabetes on targeted metabolite profiles using causal inference and interaction QTL analyses. *PLoS Genet* 2024; **20**(12): e1011346.
44. Burgess S. Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. *Int J Epidemiol* 2014; **43**(3): 922-9.
45. Rannikmäe K, Ngoh K, Bush K, et al. Accuracy of identifying incident stroke cases from linked health care data in UK Biobank. *Neurology* 2020; **95**(6): e697-e707.
46. Holmes MV, Millwood IY, Kartsonaki C, et al. Lipids, Lipoproteins, and Metabolites and Risk of Myocardial Infarction and Stroke. *J Am Coll Cardiol* 2018; **71**(6): 620-32.
47. Vojinovic D, Kalaoja M, Trompet S, et al. Association of Circulating Metabolites in Plasma or Serum and Risk of Stroke. *Neurology* 2021; **96**(8): e1110.
48. Li W, Shao C, Zhou H, et al. Multi-omics research strategies in ischemic stroke: A multidimensional perspective. *Ageing Res Rev* 2022; **81**: 101730.
49. Di Angelantonio E, Sarwar N, Perry P, et al. Major lipids, apolipoproteins, and risk of vascular disease. *Jama* 2009; **302**(18): 1993-2000.
50. Willey JZ, Xu Q, Boden-Albala B, et al. Lipid profile components and risk of ischemic stroke: the Northern Manhattan Study (NOMAS). *Arch Neurol* 2009; **66**(11): 1400-6.
51. Sun D, Tiedt S, Yu B, et al. A prospective study of serum metabolites and risk of ischemic stroke. *Neurology* 2019; **92**(16): e1890-e8.
52. Marklund M, Wu JHY, Imamura F, et al. Biomarkers of Dietary Omega-6 Fatty Acids and Incident Cardiovascular Disease and Mortality. *Circulation* 2019; **139**(21): 2422-36.
53. Ament Z, Patki A, Chaudhary N, et al. Nucleosides Associated With Incident Ischemic Stroke in the REGARDS and JHS Cohorts. *Neurology* 2022; **98**(21): e2097-e107.
54. Floegel A, Kühn T, Sookthai D, et al. Serum metabolites and risk of myocardial infarction and ischemic stroke: a targeted metabolomic approach in two German prospective cohorts. *Eur J Epidemiol* 2018; **33**(1): 55-66.

55. Julkunen H, Cichońska A, Tiainen M, et al. Atlas of plasma NMR biomarkers for health and disease in 118,461 individuals from the UK Biobank. *Nat Commun* 2023; **14**(1): 604.
56. Guo Y, Chen SF, Zhang YR, et al. Circulating metabolites associated with incident myocardial infarction and stroke: A prospective cohort study of 90 438 participants. *J Neurochem* 2022; **162**(4): 371-84.
57. Niu R, Wang H, Peng R, et al. Associations of Plasma Metabolites With Risks of Incident Stroke and Its Subtypes in Chinese Adults. *J Am Heart Assoc* 2024; **13**(12): e033201.
58. Kijpaisalratana N, Ament Z, Patki A, et al. Plasma Metabolites and Life's Simple 7 in REGARDS. *Stroke* 2024; **55**(5): 1191-9.
59. Jung JY, Lee HS, Kang DG, et al. 1H-NMR-based metabolomics study of cerebral infarction. *Stroke* 2011; **42**(5): 1282-8.
60. Jiang Z, Sun J, Liang Q, et al. A metabonomic approach applied to predict patients with cerebral infarction. *Talanta* 2011; **84**(2): 298-304.
61. Liu P, Li R, Antonov AA, et al. Discovery of Metabolite Biomarkers for Acute Ischemic Stroke Progression. *J Proteome Res* 2017; **16**(2): 773-9.
62. Markley JL, Brüschweiler R, Edison AS, et al. The future of NMR-based metabolomics. *Curr Opin Biotechnol* 2017; **43**: 34-40.
63. Ambale-Venkatesh B, Yang X, Wu CO, et al. Cardiovascular Event Prediction by Machine Learning: The Multi-Ethnic Study of Atherosclerosis. *Circ Res* 2017; **121**(9): 1092-101.
64. Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. eDoctor: machine learning and the future of medicine. *J Intern Med* 2018; **284**(6): 603-19.
65. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence* 2020; **2**(1): 56-67.
66. Zhang C, Li M, Yang M, et al. Plasma metabolites, systolic blood pressure, lifestyle, and stroke risk: A prospective cohort study. *Int J Stroke* 2024: 17474930241293408.
67. Buergel T, Steinfeldt J, Ruyoga G, et al. Metabolomic profiles predict individual multidisease outcomes. *Nat Med* 2022.
68. Natekin A, Knoll A. Gradient boosting machines, a tutorial. *Front Neurorobot* 2013; **7**: 21.
69. Grinsztajn L, Oyallon E, Varoquaux G. Why do tree-based models still outperform deep learning on tabular data? *ArXiv* 2022; **abs/2207.08815**.
70. Garriga R, Mas J, Abraha S, et al. Machine learning model to predict mental health crises from electronic health records. *Nat Med* 2022; **28**(6): 1240-8.
71. Nielsen D. Tree Boosting With XGBoost - Why Does XGBoost Win "Every" Machine Learning Competition? : NTNU; 2016.
72. Kossen J, Band N, Lyle C, Gomez AN, Rainforth T, Gal Y. Self-attention between datapoints: Going beyond individual input-output pairs in deep learning. *Advances in Neural Information Processing Systems* 2021; **34**: 28742-56.
73. Doupe P, Faghmous J, Basu S. Machine Learning for Health Services Researchers. *Value Health* 2019; **22**(7): 808-15.
74. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, California, USA: Association for Computing Machinery; 2016. p. 785–94.

75. Zhang T, Cao Y, Zhao J, Yao J, Liu G. Assessing the causal effect of genetically predicted metabolites and metabolic pathways on stroke. *J Transl Med* 2023; **21**(1): 822.
76. Zheng R, Lind L. A combined observational and Mendelian randomization investigation reveals NMR-measured analytes to be risk factors of major cardiovascular diseases. *Sci Rep* 2024; **14**(1): 10645.
77. Lin D, Zhu Y, Tian Z, et al. Causal associations between gut microbiota, gut microbiota-derived metabolites, and cerebrovascular diseases: a multivariable Mendelian randomization study. *Front Cell Infect Microbiol* 2023; **13**: 1269414.
78. He M, Xu C, Yang R, Liu L, Zhou D, Yan S. Causal relationship between human blood metabolites and risk of ischemic stroke: a Mendelian randomization study. *Front Genet* 2024; **15**: 1333454.
79. Qiao J, Zhang M, Wang T, Huang S, Zeng P. Evaluating Causal Relationship Between Metabolites and Six Cardiovascular Diseases Based on GWAS Summary Statistics. *Front Genet* 2021; **12**: 746677.
80. Hu X, Ren H, Cao Y. The association between trimethylamine N-oxide levels and ischemic stroke occurrence: a meta-analysis and Mendelian randomization study. *BMC Neurol* 2023; **23**(1): 413.
81. Si S, Hou L, Chen X, et al. Exploring the Causal Roles of Circulating Remnant Lipid Profile on Cardiovascular and Cerebrovascular Diseases: Mendelian Randomization Study. *J Epidemiol* 2022; **32**(5): 205-14.
82. Si S, Hou L, Chen X, et al. Exploring the causal roles of circulating remnant lipid profile on cardiovascular and cerebrovascular diseases: Mendelian randomization study. *J Epidemiol* 2021.
83. Reith C, Landray M, Devereaux PJ, et al. Randomized clinical trials--removing unnecessary obstacles. *N Engl J Med* 2013; **369**(11): 1061-5.
84. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015; **12**(3): e1001779.
85. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018; **562**(7726): 203-9.
86. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016; **48**(10): 1279-83.
87. Huang J, Howie B, McCarthy S, et al. Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat Commun* 2015; **6**: 8111.
88. Auton A, Brooks LD, Durbin RM, et al. A global reference for human genetic variation. *Nature* 2015; **526**(7571): 68-74.
89. Marchini J. Software IMPUTE 4. https://jmarchini.org/software/ (accessed 24th January 2024).
90. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* 2012; **44**(8): 955-9.
91. UCSC genome annotation database. http://hgdownload.cse.ucsc.edu/goldenpath/hg19/database/ (accessed 25th January 2024).
92. Woodfield R, Grant I, Sudlow CL. Accuracy of Electronic Health Record Data for Identifying Stroke Cases in Large-Scale Epidemiological Studies: A Systematic Review from the UK Biobank Stroke Outcomes Group. *PLoS One* 2015; **10**(10): e0140533.

93. Iqbal G, Johnson MR, Szczepura A, Wilson S, Gumber A, Dunn JA. UK ethnicity data collection for healthcare statistics: the South Asian perspective. *BMC Public Health* 2012; **12**: 243.
94. Craig CL, Marshall AL, Sjöström M, et al. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc* 2003; **35**(8): 1381-95.
95. Stekhoven DJ, Bühlmann P. MissForest--non-parametric missing value imputation for mixed-type data. *Bioinformatics* 2012; **28**(1): 112-8.
96. Wright MN, Ziegler A. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software* 2017; **77**(1): 1 - 17.
97. van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med* 1999; **18**(6): 681-94.
98. Troyanskaya O, Cantor M, Sherlock G, et al. Missing value estimation methods for DNA microarrays. *Bioinformatics* 2001; **17**(6): 520-5.
99. Waljee AK, Mukherjee A, Singal AG, et al. Comparison of imputation methods for missing laboratory data in medicine. *BMJ Open* 2013; **3**(8).
100. Menni C, Kastenmüller G, Petersen AK, et al. Metabolomic markers reveal novel pathways of ageing and early development in human populations. *Int J Epidemiol* 2013; **42**(4): 1111-9.
101. van den Akker EB, Trompet S, Barkey Wolf JJH, et al. Metabolic Age Based on the BBMRI-NL (1)H-NMR Metabolomics Repository as Biomarker of Age-related Disease. *Circ Genom Precis Med* 2020; **13**(5): 541-7.
102. Lau CE, Manou M, Markozannes G, et al. NMR metabolomic modeling of age and lifespan: A multicohort analysis. *Aging Cell* 2024: e14164.
103. Panyard DJ, Yu B, Snyder MP. The metabolomics of human aging: Advances, challenges, and opportunities. *Sci Adv* 2022; **8**(42): eadd6155.
104. Rist MJ, Roth A, Frommherz L, et al. Metabolite patterns predicting sex and age in participants of the Karlsruhe Metabolomics and Nutrition (KarMeN) study. *PLoS One* 2017; **12**(8): e0183228.
105. Darst BF, Koscik RL, Hogan KJ, Johnson SC, Engelman CD. Longitudinal plasma metabolomics of aging and sex. *Aging (Albany NY)* 2019; **11**(4): 1262-82.
106. Li Y, Wang M, Liu X, et al. Circulating metabolites may illustrate relationship of alcohol consumption with cardiovascular disease. *BMC Med* 2023; **21**(1): 443.
107. Allen NE, Arnold M, Parish S, et al. Approaches to minimising the epidemiological impact of sources of systematic and random variation that may affect biochemistry assay data in UK Biobank. *Wellcome Open Res* 2020; **5**: 222.
108. Nightingale Health Metabolic Biomarkers: Phase 1 Release. https://biobank.ndph.ox.ac.uk/showcase/ukb/docs/nmrm_companion_doc.pdf (accessed 20th March 2024).
109. Koton S, Schneider AL, Rosamond WD, et al. Stroke incidence and mortality trends in US communities, 1987 to 2011. *Jama* 2014; **312**(3): 259-68.
110. Howard VJ, Madsen TE, Kleindorfer DO, et al. Sex and Race Differences in the Association of Incident Ischemic Stroke With Risk Factors. *JAMA Neurol* 2019; **76**(2): 179-86.
111. Shiozawa M, Kaneko H, Itoh H, et al. Association of Body Mass Index with Ischemic and Hemorrhagic Stroke. *Nutrients* 2021; **13**(7).
112. Kannel WB, McGee DL. Diabetes and cardiovascular disease. The Framingham study. *Jama* 1979; **241**(19): 2035-8.

113. O'Donnell MJ, Xavier D, Liu L, et al. Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. *Lancet* 2010; **376**(9735): 112-23.
114. Dehghan M, Mente A, Teo KK, et al. Relationship between healthy diet and risk of cardiovascular disease among patients on drug therapies for secondary prevention: a prospective cohort study of 31 546 high-risk individuals from 40 countries. *Circulation* 2012; **126**(23): 2705-12.
115. Jenkins DJA, Spence JD, Giovannucci EL, et al. Supplemental Vitamins and Minerals for CVD Prevention and Treatment. *J Am Coll Cardiol* 2018; **71**(22): 2570-84.
116. Rustan AC, Drevon CA. Fatty Acids: Structures and Properties. eLS; 2005.
117. Jin D. Metabolomics and coronary heart disease risk in UK Biobank; 2023.
118. Waring WS, Evans LE, Kirkpatrick CT. Glycolysis inhibitors negatively bias blood glucose measurements: potential impact on the reported prevalence of diabetes mellitus. *J Clin Pathol* 2007; **60**(7): 820-3.
119. Kim HS. Blood Glucose Measurement: Is Serum Equal to Plasma? *Diabetes Metab J* 2016; **40**(5): 365-6.
120. Haeckel R, Brinck U, Colic D, et al. Comparability of blood glucose concentrations measured in different sample systems for detecting glucose intolerance. *Clin Chem* 2002; **48**(6 Pt 1): 936-9.
121. Biomarker assay quality procedures: approaches used to minimise systematic and random errors (and the wider epidemiological implications). https://biobank.ndph.ox.ac.uk/showcase/ukb/docs/biomarker_issues.pdf (accessed 13th March 2024).
122. Ritchie SC, Surendran P, Karthikeyan S, et al. Quality control and removal of technical variation of NMR metabolic biomarker data in ~120,000 UK Biobank participants. *Sci Data* 2023; **10**(1): 64.
123. Tikkanen E, Jägerroos V, Holmes MV, et al. Metabolic Biomarker Discovery for Risk of Peripheral Artery Disease Compared With Coronary Artery Disease: Lipoprotein and Metabolite Profiling of 31 657 Individuals From 5 Prospective Cohorts. *J Am Heart Assoc* 2021; **10**(23): e021995.
124. Aguilar Ramirez D. Nuclear magnetic resonance spectroscopy-measured biomarkers and cardiometabolic disease in the Mexico City prospective study; 2021.
125. Kujala UM, Mäkinen VP, Heinonen I, et al. Long-term leisure-time physical activity and serum metabolome. *Circulation* 2013; **127**(3): 340-8.
126. Bhale AS, Venkataraman K. Leveraging knowledge of HDLs major protein ApoA1: Structure, function, mutations, and potential therapeutics. *Biomed Pharmacother* 2022; **154**: 113634.
127. Ginsberg HN. Lipoprotein physiology. *Endocrinol Metab Clin North Am* 1998; **27**(3): 503-19.
128. Freedman DS, Otvos JD, Jeyarajah EJ, et al. Sex and age differences in lipoprotein subclasses measured by nuclear magnetic resonance spectroscopy: the Framingham Study. *Clin Chem* 2004; **50**(7): 1189-200.
129. Link JC, Reue K. Genetic Basis for Sex Differences in Obesity and Lipid Metabolism. *Annu Rev Nutr* 2017; **37**: 225-45.
130. Bell JA, Santos Ferreira DL, Fraser A, et al. Sex differences in systemic metabolites at four life stages: cohort study with repeated metabolomics. *BMC Med* 2021; **19**(1): 58.
131. Ambrose JA, Barua RS. The pathophysiology of cigarette smoking and cardiovascular disease: an update. *J Am Coll Cardiol* 2004; **43**(10): 1731-7.

132. Craig WY, Palomaki GE, Haddow JE. Cigarette smoking and serum lipid and lipoprotein concentrations: an analysis of published data. *Bmj* 1989; **298**(6676): 784-8.
133. Reaven G, Tsao PS. Insulin resistance and compensatory hyperinsulinemia: the key player between cigarette smoking and cardiovascular disease? *J Am Coll Cardiol* 2003; **41**(6): 1044-7.
134. Heitzer T, Ylä-Herttuala S, Luoma J, et al. Cigarette smoking potentiates endothelial dysfunction of forearm resistance vessels in patients with hypercholesterolemia. Role of oxidized LDL. *Circulation* 1996; **93**(7): 1346-53.
135. Frei B, Forte TM, Ames BN, Cross CE. Gas phase oxidants of cigarette smoke induce lipid peroxidation and changes in lipoprotein properties in human blood plasma. Protective effects of ascorbic acid. *Biochem J* 1991; **277 (Pt 1)**(Pt 1): 133-8.
136. Brien SE, Ronksley PE, Turner BJ, Mukamal KJ, Ghali WA. Effect of alcohol consumption on biological markers associated with risk of coronary heart disease: systematic review and meta-analysis of interventional studies. *Bmj* 2011; **342**: d636.
137. Würtz P, Cook S, Wang Q, et al. Metabolic profiling of alcohol consumption in 9778 young adults. *Int J Epidemiol* 2016; **45**(5): 1493-506.
138. Wang Y, Crawford MA, Chen J, et al. Fish consumption, blood docosahexaenoic acid and chronic diseases in Chinese rural populations. *Comp Biochem Physiol A Mol Integr Physiol* 2003; **136**(1): 127-40.
139. Berr C, Akbaraly T, Arnaud J, Hininger I, Roussel AM, Barberger Gateau P. Increased selenium intake in elderly high fish consumers may account for health benefits previously ascribed to omega-3 fatty acids. *J Nutr Health Aging* 2009; **13**(1): 14-8.
140. Meschia JF, Bushnell C, Boden-Albala B, et al. Guidelines for the primary prevention of stroke: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* 2014; **45**(12): 3754-832.
141. Arnett DK, Blumenthal RS, Albert MA, et al. 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation* 2019; **140**(11): e596-e646.
142. Vojinovic D, Radjabzadeh D, Kurilshikov A, et al. Relationship between gut microbiota and circulating metabolites in population-based cohorts. *Nat Commun* 2019; **10**(1): 5813.
143. Zhang Z, Gayle AA, Wang J, Zhang H, Cardinal-Fernández P. Comparing baseline characteristics between groups: an introduction to the CBCgrps package. *Ann Transl Med* 2017; **5**(24): 484.
144. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018; **392**(10159): 1923-94.
145. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)* 1995; **57**(1): 289-300.
146. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 1994; **81**(3): 515-26.
147. Zhang Z. Survival analysis in the presence of competing risks. *Ann Transl Med* 2017; **5**(3): 47.

148. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association* 1999; **94**(446): 496-509.
149. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb)* 2005; **95**(3): 221-7.
150. McQuitty LL. Similarity Analysis by Reciprocal Pairs for Discrete and Continuous Data. *Educational and Psychological Measurement* 1966; **26**(4): 825-31.
151. Amin N, Liu J, Bonnechere B, et al. Interplay of Metabolome and Gut Microbiome in Individuals With Major Depressive Disorder vs Control Individuals. *JAMA Psychiatry* 2023; **80**(6): 597-609.
152. Ringnér M. What is principal component analysis? *Nat Biotechnol* 2008; **26**(3): 303-4.
153. Zou H, Hastie T. Regularization and Variable Selection Via the Elastic Net. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 2005; **67**(2): 301-20.
154. Chen Z, Chen J, Collins R, et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int J Epidemiol* 2011; **40**(6): 1652-66.
155. Mehio-Sibai A, Feinleib M, Sibai TA, Armenian HK. A positive or a negative confounding variable? A simple teaching aid for clinicians and students. *Ann Epidemiol* 2005; **15**(6): 421-3.
156. Xu WH, Dong C, Rundek T, Elkind MS, Sacco RL. Serum albumin levels are associated with cardioembolic and cryptogenic ischemic strokes: Northern Manhattan Study. *Stroke* 2014; **45**(4): 973-8.
157. Ronit A, Kirkegaard-Klitbo DM, Dohlmann TL, et al. Plasma Albumin and Incident Cardiovascular Disease: Results From the CGPS and an Updated Meta-Analysis. *Arterioscler Thromb Vasc Biol* 2020; **40**(2): 473-82.
158. Saedi AA, Feehan J, Phu S, Duque G. Current and emerging biomarkers of frailty in the elderly. *Clin Interv Aging* 2019; **14**: 389-98.
159. Belayev L, Pinard E, Nallet H, et al. Albumin therapy of transient focal cerebral ischemia: in vivo analysis of dynamic microvascular responses. *Stroke* 2002; **33**(4): 1077-84.
160. Park HP, Nimmagadda A, DeFazio RA, Busto R, Prado R, Ginsberg MD. Albumin therapy augments the effect of thrombolysis on local vascular dynamics in a rat model of arteriolar thrombosis: a two-photon laser-scanning microscopy study. *Stroke* 2008; **39**(5): 1556-62.
161. Belayev L, Liu Y, Zhao W, Busto R, Ginsberg MD. Human albumin therapy of acute ischemic stroke: marked neuroprotective efficacy at moderate doses and with a broad therapeutic window. *Stroke* 2001; **32**(2): 553-60.
162. Kato S, Chernyavsky S, Tokita JE, et al. Relationship between proteinuria and venous thromboembolism. *Journal of Thrombosis and Thrombolysis* 2010; **30**(3): 281-5.
163. Lionaki S, Derebail VK, Hogan SL, et al. Venous thromboembolism in patients with membranous nephropathy. *Clin J Am Soc Nephrol* 2012; **7**(1): 43-51.
164. He H, Guo J. Serum albumin: a risk of stroke? *The American Journal of Emergency Medicine* 2017; **35**(1): 186-7.
165. Fuchs JE, Huber RG, von Grafenstein S, et al. Dynamic regulation of phenylalanine hydroxylase by simulated redox manipulation. *PLoS One* 2012; **7**(12): e53005.

166. Neurauter G, Schröcksnadel K, Scholl-Bürgi S, et al. Chronic immune stimulation correlates with reduced phenylalanine turnover. *Curr Drug Metab* 2008; **9**(7): 622-7.
167. Ribas GS, Sitta A, Wajner M, Vargas CR. Oxidative stress in phenylketonuria: what is the evidence? *Cell Mol Neurobiol* 2011; **31**(5): 653-62.
168. Bullock R, Zauner A, Woodward J, Young HF. Massive persistent release of excitatory amino acids following human occlusive stroke. *Stroke* 1995; **26**(11): 2187-9.
169. Kagiyama T, Glushakov AV, Sumners C, et al. Neuroprotective action of halogenated derivatives of L-phenylalanine. *Stroke* 2004; **35**(5): 1192-6.
170. Liu M, Zhou K, Li H, et al. Potential of serum metabolites for diagnosing post-stroke cognitive impairment. *Mol Biosyst* 2015; **11**(12): 3287-96.
171. Borén J, Williams KJ. The central role of arterial retention of cholesterol-rich apolipoprotein-B-containing lipoproteins in the pathogenesis of atherosclerosis: a triumph of simplicity. *Curr Opin Lipidol* 2016; **27**(5): 473-83.
172. Nordestgaard BG, Zilversmit DB. Large lipoproteins are excluded from the arterial wall in diabetic cholesterol-fed rabbits. *J Lipid Res* 1988; **29**(11): 1491-500.
173. Rosenson RS, Davidson MH, Hirsh BJ, Kathiresan S, Gaudet D. Genetics and causality of triglyceride-rich lipoproteins in atherosclerotic cardiovascular disease. *J Am Coll Cardiol* 2014; **64**(23): 2525-40.
174. Kontush A, Chantepie S, Chapman MJ. Small, dense HDL particles exert potent protection of atherogenic LDL against oxidative stress. *Arterioscler Thromb Vasc Biol* 2003; **23**(10): 1881-8.
175. Rohatgi A, Westerterp M, von Eckardstein A, Remaley A, Rye KA. HDL in the 21st Century: A Multifunctional Roadmap for Future HDL Research. *Circulation* 2021; **143**(23): 2293-309.
176. Kontush A, Therond P, Zerrad A, et al. Preferential sphingosine-1-phosphate enrichment and sphingomyelin depletion are key features of small dense HDL3 particles: relevance to antiapoptotic and antioxidative activities. *Arterioscler Thromb Vasc Biol* 2007; **27**(8): 1843-9.
177. Kim DS, Li YK, Bell GA, et al. Concentration of Smaller High-Density Lipoprotein Particle (HDL-P) Is Inversely Correlated With Carotid Intima Media Thickening After Confounder Adjustment: The Multi Ethnic Study of Atherosclerosis (MESA). *J Am Heart Assoc* 2016; **5**(5).
178. Bonnechère B, Liu J, Thompson A, Amin N, van Duijn C. Does ethnicity influence dementia, stroke and mortality risk? Evidence from the UK Biobank. *Front Public Health* 2023; **11**: 1111321.
179. Wallace J, McWilliams JM, Lollo A, Eaton J, Ndumele CD. Residual Confounding in Health Plan Performance Assessments: Evidence From Randomization in Medicaid. *Ann Intern Med* 2022; **175**(3): 314-24.
180. Vardarajan B, Kalia V, Manly J, et al. Differences in plasma metabolites related to Alzheimer's disease, APOE ε4 status, and ethnicity. *Alzheimers Dement (N Y)* 2020; **6**(1): e12025.
181. Vantaku V, Donepudi SR, Piyarathna DWB, et al. Large-scale profiling of serum metabolites in African American and European American patients with bladder cancer reveals metabolic pathways associated with patient survival. *Cancer* 2019; **125**(6): 921-32.
182. MacMahon S, Peto R, Cutler J, et al. Blood pressure, stroke, and coronary heart disease. Part 1, Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet* 1990; **335**(8692): 765-74.

183. Jin D, Trichia E, Islam N, Lewington S, Lacey B. Associations of circulating fatty acids with incident coronary heart disease: a prospective study of 89,242 individuals in UK Biobank. *BMC Cardiovasc Disord* 2023; **23**(1): 365.
184. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol* 2017; **186**(9): 1026-34.
185. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine* 2015; **12**(3): e1001779.
186. Keyes KM, Westreich D. UK Biobank, big data, and the consequences of non-representativeness. *Lancet* 2019; **393**(10178): 1297.
187. Leo B. Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Statistical Science* 2001; **16**(3): 199-231.
188. Eloyan A, Rose S. Considerations Across Three Cultures: Parametric Regressions, Interpretable Algorithms, and Complex Algorithms. *Obs Stud* 2021; **7**(1): 191-6.
189. Efron B. Prediction, Estimation, and Attribution. *Journal of the American Statistical Association* 2020; **115**(530): 636-55.
190. Freeman NLB, Sperger J, El-Zaatari H, et al. Beyond Two Cultures: Cultural Infrastructure for Data-driven Decision Support. *Obs Stud* 2021; **7**(1): 77-94.
191. Kosorok MR, Laber EB. Precision Medicine. *Annu Rev Stat Appl* 2019; **6**: 263-86.
192. World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions. *Lancet Glob Health* 2019; **7**(10): e1332-e45.
193. Truelsen T, Lindenstrøm E, Boysen G. Comparison of probability of stroke between the Copenhagen City Heart Study and the Framingham Study. *Stroke* 1994; **25**(4): 802-7.
194. Jin D, Trichia E, Islam N, Lewington S, Lacey B. Predictive value of metabolic profiling in cardiovascular risk scores: analysis of 75 000 adults in UK Biobank. *J Epidemiol Community Health* 2023; **77**(12): 802-8.
195. Hemant I, Udaya BK, Eugene HB, Michael SL. Random survival forests. *The Annals of Applied Statistics* 2008; **2**(3): 841-60.
196. Lundberg SM, Erion GG, Lee S-I. Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:180203888* 2018.
197. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015; **162**(1): W1-73.
198. Andersen PK, Gill RD. Cox's Regression Model for Counting Processes: A Large Sample Study. *The Annals of Statistics* 1982; **10**(4): 1100-20.
199. Kursa MB, Rudnicki WR. Feature Selection with the Boruta Package. *Journal of Statistical Software* 2010; **36**(11): 1 - 13.
200. Orfanoudaki A, Chesley E, Cadisch C, et al. Machine learning provides evidence that stroke risk is not linear: The non-linear Framingham stroke risk score. *PLoS One* 2020; **15**(5): e0232414.
201. Bergstra J, Bengio Y. Random search for hyper-parameter optimization. *J Mach Learn Res* 2012; **13**(null): 281–305.
202. Oikonomou EK, Van Dijk D, Parise H, et al. A phenomapping-derived tool to personalize the selection of anatomical vs. functional testing in evaluating chest pain (ASSIST). *Eur Heart J* 2021; **42**(26): 2536-48.

203. Degenhardt F, Seifert S, Szymczak S. Evaluation of variable selection methods for random forests and omics data sets. *Brief Bioinform* 2019; **20**(2): 492-503.
204. Kohavi R, John GH. Wrappers for feature subset selection. *Artificial Intelligence* 1997; **97**(1): 273-324.
205. Sorayaie Azar A, Babaei Rikan S, Naemi A, et al. Application of machine learning techniques for predicting survival in ovarian cancer. *BMC Med Inform Decis Mak* 2022; **22**(1): 345.
206. Pencina MJ, D'Agostino RB, Sr., Demler OV. Novel metrics for evaluating improvement in discrimination: net reclassification and integrated discrimination improvement for normal variables and nested models. *Stat Med* 2012; **31**(2): 101-13.
207. Longato E, Vettoretti M, Di Camillo B. A practical perspective on the concordance index for the evaluation and selection of prognostic time-to-event models. *J Biomed Inform* 2020; **108**: 103496.
208. Harrell FE, Jr., Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA* 1982; **247**(18): 2543-6.
209. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med* 2004; **23**(13): 2109-23.
210. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med* 2011; **30**(1): 11-21.
211. Chambless LE, Cummiskey CP, Cui G. Several methods to assess improvement in risk prediction models: extension to survival analysis. *Stat Med* 2011; **30**(1): 22-38.
212. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, California, USA: Curran Associates Inc.; 2017. p. 4768–77.
213. Cleveland WS. Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association* 1979; **74**(368): 829-36.
214. Golub GH, Heath M, Wahba G. Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter. *Technometrics* 1979; **21**(2): 215-23.
215. Thuemmler RJ, Pana TA, Carter B, et al. Serum Albumin and Post-Stroke Outcomes: Analysis of UK Regional Registry Data, Systematic Review, and Meta-Analysis. *Nutrients* 2024; **16**(10).
216. Pupim LB, Martin CJ, Ikizler TA. Chapter 10 - Assessment of Protein and Energy Nutritional Status. In: Kopple JD, Massry SG, Kalantar-Zadeh K, eds. Nutritional Management of Renal Disease (Third Edition): Academic Press; 2013: 137-58.
217. Hosten AO. BUN and Creatinine. In: Walker HK, Hall WD, Hurst JW, eds. Clinical Methods: The History, Physical, and Laboratory Examinations. Boston: Butterworths
Copyright © 1990, Butterworth Publishers, a division of Reed Publishing.; 1990.
218. Emwas AH. The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research. *Methods Mol Biol* 2015; **1277**: 161-93.
219. Miller M, Stone NJ, Ballantyne C, et al. Triglycerides and cardiovascular disease: a scientific statement from the American Heart Association. *Circulation* 2011; **123**(20): 2292-333.

220. Li-Gao R, Hughes DA, le Cessie S, et al. Assessment of reproducibility and biological variability of fasting and postprandial plasma metabolite concentrations using 1H NMR spectroscopy. *PLoS One* 2019; **14**(6): e0218549.
221. Unterhuber M, Kresoja KP, Rommel KP, et al. Proteomics-Enabled Deep Learning Machine Algorithms Can Enhance Prediction of Mortality. *J Am Coll Cardiol* 2021; **78**(16): 1621-31.
222. Chen Y, Huang S, Chen T, et al. Machine Learning for Prediction and Risk Stratification of Lupus Nephritis Renal Flare. *Am J Nephrol* 2021; **52**(2): 152-60.
223. Brown IJ, Elliott P. Recent Findings from Mendelian Randomization Studies of Cardiovascular Disease. *Current Cardiovascular Risk Reports* 2010; **4**(6): 429-36.
224. Davey Smith G, Ebrahim S. Epidemiology--is it time to call it a day? *Int J Epidemiol* 2001; **30**(1): 1-11.
225. Barter PJ, Caulfield M, Eriksson M, et al. Effects of torcetrapib in patients at high risk for coronary events. *N Engl J Med* 2007; **357**(21): 2109-22.
226. Beral V, Banks E, Reeves G. Evidence from randomised trials on the long-term effects of hormone replacement therapy. *Lancet* 2002; **360**(9337): 942-4.
227. MRC/BHF Heart Protection Study of antioxidant vitamin supplementation in 20,536 high-risk individuals: a randomised placebo-controlled trial. *Lancet* 2002; **360**(9326): 23-33.
228. The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. *N Engl J Med* 1994; **330**(15): 1029-35.
229. de Gaetano G. Low-dose aspirin and vitamin E in people at cardiovascular risk: a randomised trial in general practice. Collaborative Group of the Primary Prevention Project. *Lancet* 2001; **357**(9250): 89-95.
230. Bønaa KH, Njølstad I, Ueland PM, et al. Homocysteine lowering and cardiovascular events after acute myocardial infarction. *N Engl J Med* 2006; **354**(15): 1578-88.
231. Smith GD, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. *Int J Epidemiol* 2004; **33**(1): 30-42.
232. Smith GD, Timpson N, Ebrahim S. Strengthening causal inference in cardiovascular epidemiology through Mendelian randomization. *Ann Med* 2008; **40**(7): 524-41.
233. Timpson NJ, Wade KH, Smith GD. Mendelian randomization: application to cardiovascular disease. *Curr Hypertens Rep* 2012; **14**(1): 29-37.
234. Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol* 2000; **29**(4): 722-9.
235. Hingorani A, Humphries S. Nature's randomised trials. *Lancet* 2005; **366**(9501): 1906-8.
236. Burgess S, Davey Smith G, Davies NM, et al. Guidelines for performing Mendelian randomization investigations: update for summer 2023. *Wellcome Open Res* 2019; **4**: 186.
237. Richardson TG, Leyden GM, Wang Q, et al. Characterising metabolomic signatures of lipid-modifying therapies through drug target mendelian randomisation. *PLoS Biol* 2022; **20**(2): e3001547.
238. Altman DG, Bland JM. Statistics notes: Absence of evidence is not evidence of absence. *BMJ* 1995; **311**(7003): 485.
239. Welsh S. Genotyping of 500,000 UK Biobank participants Version 2.0 2017. https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/ukb_dna_processing.pdf.

240. GWAS guide using Alzheimer's disease. https://dnanexus.gitbook.io/uk-biobank-rap/science-corner/gwas-using-alzheimers-disease (accessed 18th May 2025).
241. Mbatchou J, Barnard L, Backman J, et al. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet* 2021; **53**(7): 1097-103.
242. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015; **4**: 7.
243. Sun BB, Chiou J, Traylor M, et al. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* 2023; **622**(7982): 329-38.
244. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996; **273**(5281): 1516-7.
245. Jannot AS, Ehret G, Perneger T. P < 5 × 10(-8) has emerged as a standard of statistical significance for genome-wide association studies. *J Clin Epidemiol* 2015; **68**(4): 460-5.
246. Abecasis GR, Altshuler D, Auton A, et al. A map of human genome variation from population-scale sequencing. *Nature* 2010; **467**(7319): 1061-73.
247. Yuan S, Chen J, Dan L, et al. Homocysteine, folate, and nonalcoholic fatty liver disease: a systematic review with meta-analysis and Mendelian randomization investigation. *Am J Clin Nutr* 2022; **116**(6): 1595-609.
248. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet* 2015; **47**(11): 1236-41.
249. Devlin B, Roeder K. Genomic control for association studies. *Biometrics* 1999; **55**(4): 997-1004.
250. Georgiopoulos G, Evangelou E. Power considerations for λ inflation factor in meta-analyses of genome-wide association studies. *Genet Res (Camb)* 2016; **98**: e9.
251. Finucane HK, Bulik-Sullivan B, Gusev A, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* 2015; **47**(11): 1228-35.
252. Bulik-Sullivan BK, Loh PR, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 2015; **47**(3): 291-5.
253. Chan LS, Malakhov MM, Pan W. A novel multivariable Mendelian randomization framework to disentangle highly correlated exposures with application to metabolomics. *Am J Hum Genet* 2024; **111**(9): 1834-47.
254. Lawrence M, Huber W, Pagès H, et al. Software for computing and annotating genomic ranges. *PLoS Comput Biol* 2013; **9**(8): e1003118.
255. Battle A, Brown CD, Engelhardt BE, Montgomery SB. Genetic effects on gene expression across human tissues. *Nature* 2017; **550**(7675): 204-13.
256. Dutta D, He Y, Saha A, Arvanitis M, Battle A, Chatterjee N. Aggregative trans-eQTL analysis detects trait-specific target gene sets in whole blood. *Nat Commun* 2022; **13**(1): 4323.
257. Wang L, Babushkin N, Liu Z, Liu X. Trans-eQTL mapping in gene sets identifies network effects of genetic variants. *Cell Genom* 2024; **4**(4): 100538.
258. Sollis E, Mosaku A, Abid A, et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res* 2023; **51**(D1): D977-d85.
259. Lawrence M, Gentleman R, Carey V. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* 2009; **25**(14): 1841-2.

260. Wang L, Balmat TJ, Antonia AL, et al. An atlas connecting shared genetic architecture of human diseases and molecular phenotypes provides insight into COVID-19 susceptibility. *Genome Med* 2021; **13**(1): 83.
261. Mishra A, Malik R, Hachiya T, et al. Stroke genetics informs drug discovery and risk prediction across ancestries. *Nature* 2022; **611**(7934): 115-23.
262. Burgess S, Thompson SG. Avoiding bias from weak instruments in Mendelian randomization studies. *Int J Epidemiol* 2011; **40**(3): 755-64.
263. Yarmolinsky J, Bonilla C, Haycock PC, et al. Circulating Selenium and Prostate Cancer Risk: A Mendelian Randomization Analysis. *J Natl Cancer Inst* 2018; **110**(9): 1035-8.
264. Nelson CR, Startz R. The Distribution of the Instrumental Variables Estimator and Its t-Ratio When the Instrument is a Poor One. *The Journal of Business* 1990; **63**(1): S125-S40.
265. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med* 2008; **27**(8): 1133-63.
266. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *Bmj* 2018; **362**: k601.
267. Burgess S, Foley CN, Zuber V. Inferring Causal Relationships Between Risk Factors and Outcomes from Genome-Wide Association Study Data. *Annu Rev Genomics Hum Genet* 2018; **19**: 303-27.
268. Slob EAW, Burgess S. A comparison of robust Mendelian randomization methods using summary data. *Genet Epidemiol* 2020; **44**(4): 313-29.
269. Bowden J, Hemani G, Davey Smith G. Invited Commentary: Detecting Individual and Global Horizontal Pleiotropy in Mendelian Randomization-A Job for the Humble Heterogeneity Statistic? *Am J Epidemiol* 2018; **187**(12): 2681-5.
270. Zheng J, Baird D, Borges MC, et al. Recent Developments in Mendelian Randomization Studies. *Curr Epidemiol Rep* 2017; **4**(4): 330-45.
271. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol* 2016; **40**(4): 304-14.
272. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan N, Thompson J. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stat Med* 2017; **36**(11): 1783-802.
273. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015; **44**(2): 512-25.
274. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol* 2017; **46**(6): 1985-98.
275. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet* 2018; **50**(5): 693-8.
276. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014; **23**(R1): R89-98.
277. Skrivankova VW, Richmond RC, Woolf BAR, et al. Strengthening the Reporting of Observational Studies in Epidemiology Using Mendelian Randomization: The STROBE-MR Statement. *Jama* 2021; **326**(16): 1614-21.

278. Wang G, Sarkar A, Carbonetto P, Stephens M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J R Stat Soc Series B Stat Methodol* 2020; **82**(5): 1273-300.
279. Lin Z, Xue H, Pan W. Robust multivariable Mendelian randomization based on constrained maximum likelihood. *Am J Hum Genet* 2023; **110**(4): 592-605.
280. Xue H, Shen X, Pan W. Constrained maximum likelihood-based Mendelian randomization robust to both correlated and uncorrelated pleiotropic effects. *Am J Hum Genet* 2021; **108**(7): 1251-69.
281. Markozannes G, Kanellopoulou A, Dimopoulou O, et al. Systematic review of Mendelian randomization studies on risk of cancer. *BMC Med* 2022; **20**(1): 41.
282. Støy J, Kampmann U, Mengel A, et al. Reduced CD300LG mRNA tissue expression, increased intramyocellular lipid content and impaired glucose metabolism in healthy male carriers of Arg82Cys in CD300LG: a novel genometabolic cross-link between CD300LG and common metabolic phenotypes. *BMJ Open Diabetes Res Care* 2015; **3**(1): e000095.
283. Sanderson E, Davey Smith G, Windmeijer F, Bowden J. An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. *Int J Epidemiol* 2019; **48**(3): 713-27.
284. Teslovich TM, Musunuru K, Smith AV, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010; **466**(7307): 707-13.
285. Willer CJ, Schmidt EM, Sengupta S, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet* 2013; **45**(11): 1274-83.
286. Tsimikas S. A Test in Context: Lipoprotein(a): Diagnosis, Prognosis, Controversies, and Emerging Therapies. *J Am Coll Cardiol* 2017; **69**(6): 692-711.
287. Lee SH, Kim MK, Park MS, et al. beta-Fibrinogen Gene -455 G/A Polymorphism in Korean Ischemic Stroke Patients. *J Clin Neurol* 2008; **4**(1): 17-22.
288. Ken-Dror G, Khanna T, Hills E, et al. Mendelian randomization assessing causal relationship between fibrinogen levels and ischemic stroke. *J Stroke Cerebrovasc Dis* 2025; **34**(2): 108199.
289. Moore KJ, Sheedy FJ, Fisher EA. Macrophages in atherosclerosis: a dynamic balance. *Nat Rev Immunol* 2013; **13**(10): 709-21.
290. Lam P, Soroka CJ, Boyer JL. The bile salt export pump: clinical and experimental aspects of genetic and acquired cholestatic liver disease. *Semin Liver Dis* 2010; **30**(2): 125-33.
291. Deppermann C, Cherpokova D, Nurden P, et al. Gray platelet syndrome and defective thrombo-inflammation in Nbeal2-deficient mice. *J Clin Invest* 2013; **123**(8): 3331-42.
292. Gissler MC, Stachon P, Wolf D, Marchini T. The Role of Tumor Necrosis Factor Associated Factors (TRAFs) in Vascular Inflammation and Atherosclerosis. *Front Cardiovasc Med* 2022; **9**: 826630.
293. Mamo JC, Proctor SD, Smith D. Retention of chylomicron remnants by arterial tissue; importance of an efficient clearance mechanism from plasma. *Atherosclerosis* 1998; **141 Suppl 1**: S63-9.
294. Liang HJ, Zhang QY, Hu YT, Liu GQ, Qi R. Hypertriglyceridemia: A Neglected Risk Factor for Ischemic Stroke? *J Stroke* 2022; **24**(1): 21-40.
295. Chen Y, Lu T, Pettersson-Kymmer U, et al. Genomic atlas of the plasma metabolome prioritizes metabolites implicated in human diseases. *Nat Genet* 2023; **55**(1): 44-53.

296. Malik R, Traylor M, Pulit SL, et al. Low-frequency and common genetic variation in ischemic stroke: The METASTROKE collaboration. *Neurology* 2016; **86**(13): 1217-26.
297. Sanderson E, Glymour MM, Holmes MV, et al. Mendelian randomization. *Nat Rev Methods Primers* 2022; **2**.
298. Rees JMB, Wood AM, Dudbridge F, Burgess S. Robust methods in Mendelian randomization via penalization of heterogeneous causal estimates. *PLoS One* 2019; **14**(9): e0222362.
299. Brown BC, Knowles DA. Welch-weighted Egger regression reduces false positives due to correlated pleiotropy in Mendelian randomization. *Am J Hum Genet* 2021; **108**(12): 2319-35.
300. Labrecque JA, Swanson SA. Interpretation and Potential Biases of Mendelian Randomization Estimates With Time-Varying Exposures. *Am J Epidemiol* 2019; **188**(1): 231-8.
301. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol* 2016; **40**(7): 597-608.
302. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. *Int J Epidemiol* 2016; **45**(6): 1866-86.
303. Munafò MR, Tilling K, Taylor AE, Evans DM, Davey Smith G. Collider scope: when selection bias can substantially influence observed associations. *Int J Epidemiol* 2018; **47**(1): 226-35.
304. Ganna A, Ingelsson E. 5 year mortality predictors in 498,103 UK Biobank participants: a prospective population-based study. *Lancet* 2015; **386**(9993): 533-40.
305. Manolio TA, Collins R. Enhancing the feasibility of large cohort studies. *Jama* 2010; **304**(20): 2290-1.
306. Collins R. What makes UK Biobank special? *Lancet* 2012; **379**(9822): 1173-4.

Appendices

Supplementary Table 3.1 Description of 249 metabolites

ID	Abbreviation of cardiometabolic biomarkers	Cardiometabolic biomarkers	Unit	Group	Subgroup	Derived measure
1	Total_C	Total cholesterol	mmol/l	Cholesterol		No
2	non_HDL_C	Total cholesterol minus HDL-C	mmol/l	Cholesterol		No
3	Remnant_C	Remnant cholesterol (non-HDL, non-LDL -cholesterol)	mmol/l	Cholesterol		No
4	VLDL_C	VLDL cholesterol	mmol/l	Cholesterol		No
5	Clinical_LDL_C	Clinical LDL cholesterol	mmol/l	Cholesterol		No
6	LDL_C	LDL cholesterol	mmol/l	Cholesterol		No
7	HDL_C	HDL cholesterol	mmol/l	Cholesterol		No
8	Total_TG	Total triglycerides	mmol/l	Triglycerides		No
9	VLDL_TG	Triglycerides in VLDL	mmol/l	Triglycerides		No
10	LDL_TG	Triglycerides in LDL	mmol/l	Triglycerides		No
11	HDL_TG	Triglycerides in HDL	mmol/l	Triglycerides		No
12	Total_PL	Total phospholipids in lipoprotein particles	mmol/l	Phospholipids		No
13	VLDL_PL	Phospholipids in VLDL	mmol/l	Phospholipids		No
14	LDL_PL	Phospholipids in LDL	mmol/l	Phospholipids		No
15	HDL_PL	Phospholipids in HDL	mmol/l	Phospholipids		No
16	Total_CE	Total esterified cholesterol	mmol/l	Cholesteryl esters		No
17	VLDL_CE	Cholesteryl esters in VLDL	mmol/l	Cholesteryl esters		No
18	LDL_CE	Cholesteryl esters in LDL	mmol/l	Cholesteryl esters		No
19	HDL_CE	Cholesteryl esters in HDL	mmol/l	Cholesteryl esters		No
20	Total_FC	Total free cholesterol	mmol/l	Free cholesterol		No
21	VLDL_FC	Free cholesterol in VLDL	mmol/l	Free cholesterol		No
22	LDL_FC	Free cholesterol in LDL	mmol/l	Free cholesterol		No
23	HDL_FC	Free cholesterol in HDL	mmol/l	Free cholesterol		No
24	Total_L	Total lipids in lipoprotein particles	mmol/l	Total lipids		No
25	VLDL_L	Total lipids in VLDL	mmol/l	Total lipids		No
26	LDL_L	Total lipids in LDL	mmol/l	Total lipids		No
27	HDL_L	Total lipids in HDL	mmol/l	Total lipids		No
28	Total_P	Total concentration of lipoprotein particles	mmol/l	Lipoprotein particle concentrations		No
29	VLDL_P	Concentration of VLDL particles	mmol/l	Lipoprotein particle concentrations		No
30	LDL_P	Concentration of LDL particles	mmol/l	Lipoprotein particle concentrations		No
31	HDL_P	Concentration of HDL particles	mmol/l	Lipoprotein particle concentrations		No
32	VLDL_size	Average diameter for VLDL particles	nm	Lipoprotein particle sizes		No
33	LDL_size	Average diameter for LDL particles	nm	Lipoprotein particle sizes		No

ID	Abbreviation of cardiometabolic biomarkers	Cardiometabolic biomarkers	Unit	Group	Subgroup	Derived measure
34	HDL_size	Average diameter for HDL particles	nm	Lipoprotein particle sizes		No
35	Phosphoglyc	Phosphoglycerides	mmol/l	Other lipids		No
36	TG_by_PG	Ratio of triglycerides to phosphoglycerides	ratio	Other lipids		Yes
37	Cholines	Total cholines	mmol/l	Other lipids		No
38	Phosphatidylc	Phosphatidylcholines	mmol/l	Other lipids		No
39	Sphingomyelins	Sphingomyelins	mmol/l	Other lipids		No
40	ApoB	Apolipoprotein B	g/l	Apolipoproteins		No
41	ApoA1	Apolipoprotein A1	g/l	Apolipoproteins		No
42	ApoB_by_ApoA1	Ratio of apolipoprotein B to apolipoprotein A1	ratio	Apolipoproteins		Yes
43	Total_FA	Total fatty acids	mmol/l	Fatty acids		No
44	Unsaturation	Degree of unsaturation	degree	Fatty acids		No
45	Omega_3	Omega-3 fatty acids	mmol/l	Fatty acids		No
46	Omega_6	Omega-6 fatty acids	mmol/l	Fatty acids		No
47	PUFA	Polyunsaturated fatty acids	mmol/l	Fatty acids		No
48	MUFA	Monounsaturated fatty acids	mmol/l	Fatty acids		No
49	SFA	Saturated fatty acids	mmol/l	Fatty acids		No
50	LA	Linoleic acid	mmol/l	Fatty acids		No
51	DHA	Docosahexaenoic acid	mmol/l	Fatty acids		No
52	Omega_3_pct	Ratio of omega-3 fatty acids to total fatty acids	%	Fatty acids	Fatty acid ratios	Yes
53	Omega_6_pct	Ratio of omega-6 fatty acids to total fatty acids	%	Fatty acids	Fatty acid ratios	Yes
54	PUFA_pct	Ratio of polyunsaturated fatty acids to total fatty acids	%	Fatty acids	Fatty acid ratios	Yes
55	MUFA_pct	Ratio of monounsaturated fatty acids to total fatty acids	%	Fatty acids	Fatty acid ratios	Yes
56	SFA_pct	Ratio of saturated fatty acids to total fatty acids	%	Fatty acids	Fatty acid ratios	Yes
57	LA_pct	Ratio of linoleic acid to total fatty acids	%	Fatty acids	Fatty acid ratios	Yes
58	DHA_pct	Ratio of docosahexaenoic acid to total fatty acids	%	Fatty acids	Fatty acid ratios	Yes
59	PUFA_by_MUFA	Ratio of polyunsaturated fatty acids to monounsaturated fatty acids	ratio	Fatty acids	Fatty acid ratios	Yes
60	Omega_6_by_Omega_3	Ratio of omega-6 fatty acids to omega-3 fatty acids	ratio	Fatty acids	Fatty acid ratios	Yes
61	Ala	Alanine	mmol/l	Amino acids		No
62	Gln	Glutamine	mmol/l	Amino acids		No
63	Gly	Glycine	mmol/l	Amino acids		No
64	His	Histidine	mmol/l	Amino acids		No
65	Total_BCAA	Total concentration of branched-chain	mmol/l	Amino acids	Branched-chain amino acids	No

ID	Abbreviation of cardiometabolic biomarkers	Cardiometabolic biomarkers	Unit	Group	Subgroup	Derived measure
		amino acids (leucine + isoleucine + valine)				
66	Ile	Isoleucine	mmol/l	Amino acids	Branched-chain amino acids	No
67	Leu	Leucine	mmol/l	Amino acids	Branched-chain amino acids	No
68	Val	Valine	mmol/l	Amino acids	Branched-chain amino acids	No
69	Phe	Phenylalanine	mmol/l	Amino acids	Aromatic amino acids	No
70	Tyr	Tyrosine	mmol/l	Amino acids	Aromatic amino acids	No
71	Glucose	Glucose	mmol/l	Glycolysis related metabolites		No
72	Lactate	Lactate	mmol/l	Glycolysis related metabolites		No
73	Pyruvate	Pyruvate	mmol/l	Glycolysis related metabolites		No
74	Citrate	Citrate	mmol/l	Glycolysis related metabolites		No
75	bOHbutyrate	3-Hydroxybutyrate	mmol/l	Ketone bodies		No
76	Acetate	Acetate	mmol/l	Ketone bodies		No
77	Acetoacetate	Acetoacetate	mmol/l	Ketone bodies		No
78	Acetone	Acetone	mmol/l	Ketone bodies		No
79	Creatinine	Creatinine	mmol/l	Fluid balance		No
80	Albumin	Albumin	g/l	Fluid balance		No
81	GlycA	Glycoprotein acetyls	mmol/l	Inflammation		No
82	XXL_VLDL_P	Concentration of chylomicrons and extremely large VLDL particles	mmol/l	Lipoprotein subclasses	Chylomicrons and extremely large VLDL (particle diameters from 75 nm upwards)	No
83	XXL_VLDL_L	Total lipids in chylomicrons and extremely large VLDL	mmol/l	Lipoprotein subclasses	Chylomicrons and extremely large VLDL (particle diameters from 75 nm upwards)	No
84	XXL_VLDL_PL	Phospholipids in chylomicrons and extremely large VLDL	mmol/l	Lipoprotein subclasses	Chylomicrons and extremely large VLDL (particle diameters from 75 nm upwards)	No
85	XXL_VLDL_C	Cholesterol in chylomicrons and extremely large VLDL	mmol/l	Lipoprotein subclasses	Chylomicrons and extremely large VLDL (particle diameters from 75 nm upwards)	No
86	XXL_VLDL_CE	Cholesteryl esters in chylomicrons and extremely large VLDL	mmol/l	Lipoprotein subclasses	Chylomicrons and extremely large VLDL (particle diameters from 75 nm upwards)	No
87	XXL_VLDL_FC	Free cholesterol in chylomicrons and extremely large VLDL	mmol/l	Lipoprotein subclasses	Chylomicrons and extremely large VLDL (particle diameters from 75 nm upwards)	No
88	XXL_VLDL_TG	Triglycerides in chylomicrons and extremely large VLDL	mmol/l	Lipoprotein subclasses	Chylomicrons and extremely large VLDL (particle diameters from 75 nm upwards)	No
89	XL_VLDL_P	Concentration of very large VLDL particles	mmol/l	Lipoprotein subclasses	Very large VLDL (average diameter 64 nm)	No
90	XL_VLDL_L	Total lipids in very large VLDL	mmol/l	Lipoprotein subclasses	Very large VLDL (average diameter 64 nm)	No
91	XL_VLDL_PL	Phospholipids in very large VLDL	mmol/l	Lipoprotein subclasses	Very large VLDL (average diameter 64 nm)	No
92	XL_VLDL_C	Cholesterol in very large VLDL	mmol/l	Lipoprotein subclasses	Very large VLDL (average diameter 64 nm)	No
93	XL_VLDL_CE	Cholesteryl esters in very large VLDL	mmol/l	Lipoprotein subclasses	Very large VLDL (average diameter 64 nm)	No
94	XL_VLDL_FC	Free cholesterol in very large VLDL	mmol/l	Lipoprotein subclasses	Very large VLDL (average diameter 64 nm)	No
95	XL_VLDL_TG	Triglycerides in very large VLDL	mmol/l	Lipoprotein subclasses	Very large VLDL (average diameter 64 nm)	No
96	L_VLDL_P	Concentration of large VLDL particles	mmol/l	Lipoprotein subclasses	Large VLDL (average diameter 53.6 nm)	No
97	L_VLDL_L	Total lipids in large VLDL	mmol/l	Lipoprotein subclasses	Large VLDL (average diameter 53.6 nm)	No

ID	Abbreviation of cardiometabolic biomarkers	Cardiometabolic biomarkers	Unit	Group	Subgroup	Derived measure
98	L_VLDL_PL	Phospholipids in large VLDL	mmol/l	Lipoprotein subclasses	Large VLDL (average diameter 53.6 nm)	No
99	L_VLDL_C	Cholesterol in large VLDL	mmol/l	Lipoprotein subclasses	Large VLDL (average diameter 53.6 nm)	No
100	L_VLDL_CE	Cholesteryl esters in large VLDL	mmol/l	Lipoprotein subclasses	Large VLDL (average diameter 53.6 nm)	No
101	L_VLDL_FC	Free cholesterol in large VLDL	mmol/l	Lipoprotein subclasses	Large VLDL (average diameter 53.6 nm)	No
102	L_VLDL_TG	Triglycerides in large VLDL	mmol/l	Lipoprotein subclasses	Large VLDL (average diameter 53.6 nm)	No
103	M_VLDL_P	Concentration of medium VLDL particles	mmol/l	Lipoprotein subclasses	Medium VLDL (average diameter 44.5 nm)	No
104	M_VLDL_L	Total lipids in medium VLDL	mmol/l	Lipoprotein subclasses	Medium VLDL (average diameter 44.5 nm)	No
105	M_VLDL_PL	Phospholipids in medium VLDL	mmol/l	Lipoprotein subclasses	Medium VLDL (average diameter 44.5 nm)	No
106	M_VLDL_C	Cholesterol in medium VLDL	mmol/l	Lipoprotein subclasses	Medium VLDL (average diameter 44.5 nm)	No
107	M_VLDL_CE	Cholesteryl esters in medium VLDL	mmol/l	Lipoprotein subclasses	Medium VLDL (average diameter 44.5 nm)	No
108	M_VLDL_FC	Free cholesterol in medium VLDL	mmol/l	Lipoprotein subclasses	Medium VLDL (average diameter 44.5 nm)	No
109	M_VLDL_TG	Triglycerides in medium VLDL	mmol/l	Lipoprotein subclasses	Medium VLDL (average diameter 44.5 nm)	No
110	S_VLDL_P	Concentration of small VLDL particles	mmol/l	Lipoprotein subclasses	Small VLDL (average diameter 36.8 nm)	No
111	S_VLDL_L	Total lipids in small VLDL	mmol/l	Lipoprotein subclasses	Small VLDL (average diameter 36.8 nm)	No
112	S_VLDL_PL	Phospholipids in small VLDL	mmol/l	Lipoprotein subclasses	Small VLDL (average diameter 36.8 nm)	No
113	S_VLDL_C	Cholesterol in small VLDL	mmol/l	Lipoprotein subclasses	Small VLDL (average diameter 36.8 nm)	No
114	S_VLDL_CE	Cholesteryl esters in small VLDL	mmol/l	Lipoprotein subclasses	Small VLDL (average diameter 36.8 nm)	No
115	S_VLDL_FC	Free cholesterol in small VLDL	mmol/l	Lipoprotein subclasses	Small VLDL (average diameter 36.8 nm)	No
116	S_VLDL_TG	Triglycerides in small VLDL	mmol/l	Lipoprotein subclasses	Small VLDL (average diameter 36.8 nm)	No
117	XS_VLDL_P	Concentration of very small VLDL particles	mmol/l	Lipoprotein subclasses	Very small VLDL (average diameter 31.3 nm)	No
118	XS_VLDL_L	Total lipids in very small VLDL	mmol/l	Lipoprotein subclasses	Very small VLDL (average diameter 31.3 nm)	No
119	XS_VLDL_PL	Phospholipids in very small VLDL	mmol/l	Lipoprotein subclasses	Very small VLDL (average diameter 31.3 nm)	No
120	XS_VLDL_C	Cholesterol in very small VLDL	mmol/l	Lipoprotein subclasses	Very small VLDL (average diameter 31.3 nm)	No
121	XS_VLDL_CE	Cholesteryl esters in very small VLDL	mmol/l	Lipoprotein subclasses	Very small VLDL (average diameter 31.3 nm)	No
122	XS_VLDL_FC	Free cholesterol in very small VLDL	mmol/l	Lipoprotein subclasses	Very small VLDL (average diameter 31.3 nm)	No
123	XS_VLDL_TG	Triglycerides in very small VLDL	mmol/l	Lipoprotein subclasses	Very small VLDL (average diameter 31.3 nm)	No
124	IDL_P	Concentration of IDL particles	mmol/l	Lipoprotein subclasses	IDL (average diameter 28.6 nm)	No
125	IDL_L	Total lipids in IDL	mmol/l	Lipoprotein subclasses	IDL (average diameter 28.6 nm)	No
126	IDL_PL	Phospholipids in IDL	mmol/l	Lipoprotein subclasses	IDL (average diameter 28.6 nm)	No
127	IDL_C	Cholesterol in IDL	mmol/l	Lipoprotein subclasses	IDL (average diameter 28.6 nm)	No
128	IDL_CE	Cholesteryl esters in IDL	mmol/l	Lipoprotein subclasses	IDL (average diameter 28.6 nm)	No
129	IDL_FC	Free cholesterol in IDL	mmol/l	Lipoprotein subclasses	IDL (average diameter 28.6 nm)	No
130	IDL_TG	Triglycerides in IDL	mmol/l	Lipoprotein subclasses	IDL (average diameter 28.6 nm)	No
131	L_LDL_P	Concentration of large LDL particles	mmol/l	Lipoprotein subclasses	Large LDL (average diameter 25.5 nm)	No
132	L_LDL_L	Total lipids in large LDL	mmol/l	Lipoprotein subclasses	Large LDL (average diameter 25.5 nm)	No
133	L_LDL_PL	Phospholipids in large LDL	mmol/l	Lipoprotein subclasses	Large LDL (average diameter 25.5 nm)	No
134	L_LDL_C	Cholesterol in large LDL	mmol/l	Lipoprotein subclasses	Large LDL (average diameter 25.5 nm)	No
135	L_LDL_CE	Cholesteryl esters in large LDL	mmol/l	Lipoprotein subclasses	Large LDL (average diameter 25.5 nm)	No
136	L_LDL_FC	Free cholesterol in large LDL	mmol/l	Lipoprotein subclasses	Large LDL (average diameter 25.5 nm)	No
137	L_LDL_TG	Triglycerides in large LDL	mmol/l	Lipoprotein subclasses	Large LDL (average diameter 25.5 nm)	No

ID	Abbreviation of cardiometabolic biomarkers	Cardiometabolic biomarkers	Unit	Group	Subgroup	Derived measure
138	M_LDL_P	Concentration of medium LDL particles	mmol/l	Lipoprotein subclasses	Medium LDL (average diameter 23 nm)	No
139	M_LDL_L	Total lipids in medium LDL	mmol/l	Lipoprotein subclasses	Medium LDL (average diameter 23 nm)	No
140	M_LDL_PL	Phospholipids in medium LDL	mmol/l	Lipoprotein subclasses	Medium LDL (average diameter 23 nm)	No
141	M_LDL_C	Cholesterol in medium LDL	mmol/l	Lipoprotein subclasses	Medium LDL (average diameter 23 nm)	No
142	M_LDL_CE	Cholesteryl esters in medium LDL	mmol/l	Lipoprotein subclasses	Medium LDL (average diameter 23 nm)	No
143	M_LDL_FC	Free cholesterol in medium LDL	mmol/l	Lipoprotein subclasses	Medium LDL (average diameter 23 nm)	No
144	M_LDL_TG	Triglycerides in medium LDL	mmol/l	Lipoprotein subclasses	Medium LDL (average diameter 23 nm)	No
145	S_LDL_P	Concentration of small LDL particles	mmol/l	Lipoprotein subclasses	Small LDL (average diameter 18.7 nm)	No
146	S_LDL_L	Total lipids in small LDL	mmol/l	Lipoprotein subclasses	Small LDL (average diameter 18.7 nm)	No
147	S_LDL_PL	Phospholipids in small LDL	mmol/l	Lipoprotein subclasses	Small LDL (average diameter 18.7 nm)	No
148	S_LDL_C	Cholesterol in small LDL	mmol/l	Lipoprotein subclasses	Small LDL (average diameter 18.7 nm)	No
149	S_LDL_CE	Cholesteryl esters in small LDL	mmol/l	Lipoprotein subclasses	Small LDL (average diameter 18.7 nm)	No
150	S_LDL_FC	Free cholesterol in small LDL	mmol/l	Lipoprotein subclasses	Small LDL (average diameter 18.7 nm)	No
151	S_LDL_TG	Triglycerides in small LDL	mmol/l	Lipoprotein subclasses	Small LDL (average diameter 18.7 nm)	No
152	XL_HDL_P	Concentration of very large HDL particles	mmol/l	Lipoprotein subclasses	Very large HDL (average diameter 14.3 nm)	No
153	XL_HDL_L	Total lipids in very large HDL	mmol/l	Lipoprotein subclasses	Very large HDL (average diameter 14.3 nm)	No
154	XL_HDL_PL	Phospholipids in very large HDL	mmol/l	Lipoprotein subclasses	Very large HDL (average diameter 14.3 nm)	No
155	XL_HDL_C	Cholesterol in very large HDL	mmol/l	Lipoprotein subclasses	Very large HDL (average diameter 14.3 nm)	No
156	XL_HDL_CE	Cholesteryl esters in very large HDL	mmol/l	Lipoprotein subclasses	Very large HDL (average diameter 14.3 nm)	No
157	XL_HDL_FC	Free cholesterol in very large HDL	mmol/l	Lipoprotein subclasses	Very large HDL (average diameter 14.3 nm)	No
158	XL_HDL_TG	Triglycerides in very large HDL	mmol/l	Lipoprotein subclasses	Very large HDL (average diameter 14.3 nm)	No
159	L_HDL_P	Concentration of large HDL particles	mmol/l	Lipoprotein subclasses	Large HDL (average diameter 12.1 nm)	No
160	L_HDL_L	Total lipids in large HDL	mmol/l	Lipoprotein subclasses	Large HDL (average diameter 12.1 nm)	No
161	L_HDL_PL	Phospholipids in large HDL	mmol/l	Lipoprotein subclasses	Large HDL (average diameter 12.1 nm)	No
162	L_HDL_C	Cholesterol in large HDL	mmol/l	Lipoprotein subclasses	Large HDL (average diameter 12.1 nm)	No
163	L_HDL_CE	Cholesteryl esters in large HDL	mmol/l	Lipoprotein subclasses	Large HDL (average diameter 12.1 nm)	No
164	L_HDL_FC	Free cholesterol in large HDL	mmol/l	Lipoprotein subclasses	Large HDL (average diameter 12.1 nm)	No
165	L_HDL_TG	Triglycerides in large HDL	mmol/l	Lipoprotein subclasses	Large HDL (average diameter 12.1 nm)	No
166	M_HDL_P	Concentration of medium HDL particles	mmol/l	Lipoprotein subclasses	Medium HDL (average diameter 10.9 nm)	No
167	M_HDL_L	Total lipids in medium HDL	mmol/l	Lipoprotein subclasses	Medium HDL (average diameter 10.9 nm)	No
168	M_HDL_PL	Phospholipids in medium HDL	mmol/l	Lipoprotein subclasses	Medium HDL (average diameter 10.9 nm)	No
169	M_HDL_C	Cholesterol in medium HDL	mmol/l	Lipoprotein subclasses	Medium HDL (average diameter 10.9 nm)	No
170	M_HDL_CE	Cholesteryl esters in medium HDL	mmol/l	Lipoprotein subclasses	Medium HDL (average diameter 10.9 nm)	No
171	M_HDL_FC	Free cholesterol in medium HDL	mmol/l	Lipoprotein subclasses	Medium HDL (average diameter 10.9 nm)	No
172	M_HDL_TG	Triglycerides in medium HDL	mmol/l	Lipoprotein subclasses	Medium HDL (average diameter 10.9 nm)	No
173	S_HDL_P	Concentration of small HDL particles	mmol/l	Lipoprotein subclasses	Small HDL (average diameter 8.7 nm)	No
174	S_HDL_L	Total lipids in small HDL	mmol/l	Lipoprotein subclasses	Small HDL (average diameter 8.7 nm)	No
175	S_HDL_PL	Phospholipids in small HDL	mmol/l	Lipoprotein subclasses	Small HDL (average diameter 8.7 nm)	No
176	S_HDL_C	Cholesterol in small HDL	mmol/l	Lipoprotein subclasses	Small HDL (average diameter 8.7 nm)	No
177	S_HDL_CE	Cholesteryl esters in small HDL	mmol/l	Lipoprotein subclasses	Small HDL (average diameter 8.7 nm)	No
178	S_HDL_FC	Free cholesterol in small HDL	mmol/l	Lipoprotein subclasses	Small HDL (average diameter 8.7 nm)	No

ID	Abbreviation of cardiometabolic biomarkers	Cardiometabolic biomarkers	Unit	Group	Subgroup	Derived measure
179	S_HDL_TG	Triglycerides in small HDL	mmol/l	Lipoprotein subclasses	Small HDL (average diameter 8.7 nm)	No
180	XXL_VLDL_PL_pct	Phospholipids to total lipids ratio in chylomicrons and extremely large VLDL	%	Relative lipoprotein lipid concentrations	Chylomicrons and extremely large VLDL ratios	Yes
181	XXL_VLDL_C_pct	Cholesterol to total lipids ratio in chylomicrons and extremely large VLDL	%	Relative lipoprotein lipid concentrations	Chylomicrons and extremely large VLDL ratios	Yes
182	XXL_VLDL_CE_pct	Cholesteryl esters to total lipids ratio in chylomicrons and extremely large VLDL	%	Relative lipoprotein lipid concentrations	Chylomicrons and extremely large VLDL ratios	Yes
183	XXL_VLDL_FC_pct	Free cholesterol to total lipids ratio in chylomicrons and extremely large VLDL	%	Relative lipoprotein lipid concentrations	Chylomicrons and extremely large VLDL ratios	Yes
184	XXL_VLDL_TG_pct	Triglycerides to total lipids ratio in chylomicrons and extremely large VLDL	%	Relative lipoprotein lipid concentrations	Chylomicrons and extremely large VLDL ratios	Yes
185	XL_VLDL_PL_pct	Phospholipids to total lipids ratio in very large VLDL	%	Relative lipoprotein lipid concentrations	Very large VLDL ratios	Yes
186	XL_VLDL_C_pct	Cholesterol to total lipids ratio in very large VLDL	%	Relative lipoprotein lipid concentrations	Very large VLDL ratios	Yes
187	XL_VLDL_CE_pct	Cholesteryl esters to total lipids ratio in very large VLDL	%	Relative lipoprotein lipid concentrations	Very large VLDL ratios	Yes
188	XL_VLDL_FC_pct	Free cholesterol to total lipids ratio in very large VLDL	%	Relative lipoprotein lipid concentrations	Very large VLDL ratios	Yes
189	XL_VLDL_TG_pct	Triglycerides to total lipids ratio in very large VLDL	%	Relative lipoprotein lipid concentrations	Very large VLDL ratios	Yes
190	L_VLDL_PL_pct	Phospholipids to total lipids ratio in large VLDL	%	Relative lipoprotein lipid concentrations	Large VLDL ratios	Yes
191	L_VLDL_C_pct	Cholesterol to total lipids ratio in large VLDL	%	Relative lipoprotein lipid concentrations	Large VLDL ratios	Yes
192	L_VLDL_CE_pct	Cholesteryl esters to total lipids ratio in large VLDL	%	Relative lipoprotein lipid concentrations	Large VLDL ratios	Yes
193	L_VLDL_FC_pct	Free cholesterol to total lipids ratio in large VLDL	%	Relative lipoprotein lipid concentrations	Large VLDL ratios	Yes
194	L_VLDL_TG_pct	Triglycerides to total lipids ratio in large VLDL	%	Relative lipoprotein lipid concentrations	Large VLDL ratios	Yes
195	M_VLDL_PL_pct	Phospholipids to total lipids ratio in medium VLDL	%	Relative lipoprotein lipid concentrations	Medium VLDL ratios	Yes
196	M_VLDL_C_pct	Cholesterol to total lipids ratio in medium VLDL	%	Relative lipoprotein lipid concentrations	Medium VLDL ratios	Yes
197	M_VLDL_CE_pct	Cholesteryl esters to total lipids ratio in medium VLDL	%	Relative lipoprotein lipid concentrations	Medium VLDL ratios	Yes
198	M_VLDL_FC_pct	Free cholesterol to total lipids ratio in medium VLDL	%	Relative lipoprotein lipid concentrations	Medium VLDL ratios	Yes
199	M_VLDL_TG_pct	Triglycerides to total lipids ratio in medium VLDL	%	Relative lipoprotein lipid concentrations	Medium VLDL ratios	Yes
200	S_VLDL_PL_pct	Phospholipids to total lipids ratio in small VLDL	%	Relative lipoprotein lipid concentrations	Small VLDL ratios	Yes

ID	Abbreviation of cardiometabolic biomarkers	Cardiometabolic biomarkers	Unit	Group	Subgroup	Derived measure
201	S_VLDL_C_pct	Cholesterol to total lipids ratio in small VLDL	%	Relative lipoprotein lipid concentrations	Small VLDL ratios	Yes
202	S_VLDL_CE_pct	Cholesteryl esters to total lipids ratio in small VLDL	%	Relative lipoprotein lipid concentrations	Small VLDL ratios	Yes
203	S_VLDL_FC_pct	Free cholesterol to total lipids ratio in small VLDL	%	Relative lipoprotein lipid concentrations	Small VLDL ratios	Yes
204	S_VLDL_TG_pct	Triglycerides to total lipids ratio in small VLDL	%	Relative lipoprotein lipid concentrations	Small VLDL ratios	Yes
205	XS_VLDL_PL_pct	Phospholipids to total lipids ratio in very small VLDL	%	Relative lipoprotein lipid concentrations	Very small VLDL ratios	Yes
206	XS_VLDL_C_pct	Cholesterol to total lipids ratio in very small VLDL	%	Relative lipoprotein lipid concentrations	Very small VLDL ratios	Yes
207	XS_VLDL_CE_pct	Cholesteryl esters to total lipids ratio in very small VLDL	%	Relative lipoprotein lipid concentrations	Very small VLDL ratios	Yes
208	XS_VLDL_FC_pct	Free cholesterol to total lipids ratio in very small VLDL	%	Relative lipoprotein lipid concentrations	Very small VLDL ratios	Yes
209	XS_VLDL_TG_pct	Triglycerides to total lipids ratio in very small VLDL	%	Relative lipoprotein lipid concentrations	Very small VLDL ratios	Yes
210	IDL_PL_pct	Phospholipids to total lipids ratio in IDL	%	Relative lipoprotein lipid concentrations	IDL ratios	Yes
211	IDL_C_pct	Cholesterol to total lipids ratio in IDL	%	Relative lipoprotein lipid concentrations	IDL ratios	Yes
212	IDL_CE_pct	Cholesteryl esters to total lipids ratio in IDL	%	Relative lipoprotein lipid concentrations	IDL ratios	Yes
213	IDL_FC_pct	Free cholesterol to total lipids ratio in IDL	%	Relative lipoprotein lipid concentrations	IDL ratios	Yes
214	IDL_TG_pct	Triglycerides to total lipids ratio in IDL	%	Relative lipoprotein lipid concentrations	IDL ratios	Yes
215	L_LDL_PL_pct	Phospholipids to total lipids ratio in large LDL	%	Relative lipoprotein lipid concentrations	Large LDL ratios	Yes
216	L_LDL_C_pct	Cholesterol to total lipids ratio in large LDL	%	Relative lipoprotein lipid concentrations	Large LDL ratios	Yes
217	L_LDL_CE_pct	Cholesteryl esters to total lipids ratio in large LDL	%	Relative lipoprotein lipid concentrations	Large LDL ratios	Yes
218	L_LDL_FC_pct	Free cholesterol to total lipids ratio in large LDL	%	Relative lipoprotein lipid concentrations	Large LDL ratios	Yes
219	L_LDL_TG_pct	Triglycerides to total lipids ratio in large LDL	%	Relative lipoprotein lipid concentrations	Large LDL ratios	Yes
220	M_LDL_PL_pct	Phospholipids to total lipids ratio in medium LDL	%	Relative lipoprotein lipid concentrations	Medium LDL ratios	Yes
221	M_LDL_C_pct	Cholesterol to total lipids ratio in medium LDL	%	Relative lipoprotein lipid concentrations	Medium LDL ratios	Yes
222	M_LDL_CE_pct	Cholesteryl esters to total lipids ratio in	%	Relative lipoprotein lipid	Medium LDL ratios	Yes

ID	Abbreviation of cardiometabolic biomarkers	Cardiometabolic biomarkers	Unit	Group	Subgroup	Derived measure
		medium LDL		concentrations		
223	M_LDL_FC_pct	Free cholesterol to total lipids ratio in medium LDL	%	Relative lipoprotein lipid concentrations	Medium LDL ratios	Yes
224	M_LDL_TG_pct	Triglycerides to total lipids ratio in medium LDL	%	Relative lipoprotein lipid concentrations	Medium LDL ratios	Yes
225	S_LDL_PL_pct	Phospholipids to total lipids ratio in small LDL	%	Relative lipoprotein lipid concentrations	Small LDL ratios	Yes
226	S_LDL_C_pct	Cholesterol to total lipids ratio in small LDL	%	Relative lipoprotein lipid concentrations	Small LDL ratios	Yes
227	S_LDL_CE_pct	Cholesteryl esters to total lipids ratio in small LDL	%	Relative lipoprotein lipid concentrations	Small LDL ratios	Yes
228	S_LDL_FC_pct	Free cholesterol to total lipids ratio in small LDL	%	Relative lipoprotein lipid concentrations	Small LDL ratios	Yes
229	S_LDL_TG_pct	Triglycerides to total lipids ratio in small LDL	%	Relative lipoprotein lipid concentrations	Small LDL ratios	Yes
230	XL_HDL_PL_pct	Phospholipids to total lipids ratio in very large HDL	%	Relative lipoprotein lipid concentrations	Very large HDL ratios	Yes
231	XL_HDL_C_pct	Cholesterol to total lipids ratio in very large HDL	%	Relative lipoprotein lipid concentrations	Very large HDL ratios	Yes
232	XL_HDL_CE_pct	Cholesteryl esters to total lipids ratio in very large HDL	%	Relative lipoprotein lipid concentrations	Very large HDL ratios	Yes
233	XL_HDL_FC_pct	Free cholesterol to total lipids ratio in very large HDL	%	Relative lipoprotein lipid concentrations	Very large HDL ratios	Yes
234	XL_HDL_TG_pct	Triglycerides to total lipids ratio in very large HDL	%	Relative lipoprotein lipid concentrations	Very large HDL ratios	Yes
235	L_HDL_PL_pct	Phospholipids to total lipids ratio in large HDL	%	Relative lipoprotein lipid concentrations	Large HDL ratios	Yes
236	L_HDL_C_pct	Cholesterol to total lipids ratio in large HDL	%	Relative lipoprotein lipid concentrations	Large HDL ratios	Yes
237	L_HDL_CE_pct	Cholesteryl esters to total lipids ratio in large HDL	%	Relative lipoprotein lipid concentrations	Large HDL ratios	Yes
238	L_HDL_FC_pct	Free cholesterol to total lipids ratio in large HDL	%	Relative lipoprotein lipid concentrations	Large HDL ratios	Yes
239	L_HDL_TG_pct	Triglycerides to total lipids ratio in large HDL	%	Relative lipoprotein lipid concentrations	Large HDL ratios	Yes
240	M_HDL_PL_pct	Phospholipids to total lipids ratio in medium HDL	%	Relative lipoprotein lipid concentrations	Medium HDL ratios	Yes
241	M_HDL_C_pct	Cholesterol to total lipids ratio in medium HDL	%	Relative lipoprotein lipid concentrations	Medium HDL ratios	Yes
242	M_HDL_CE_pct	Cholesteryl esters to total lipids ratio in medium HDL	%	Relative lipoprotein lipid concentrations	Medium HDL ratios	Yes
243	M_HDL_FC_pct	Free cholesterol to total lipids ratio in medium HDL	%	Relative lipoprotein lipid concentrations	Medium HDL ratios	Yes

ID	Abbreviation of cardiometabolic biomarkers	Cardiometabolic biomarkers	Unit	Group	Subgroup	Derived measure
244	M_HDL_TG_pct	Triglycerides to total lipids ratio in medium HDL	%	Relative lipoprotein lipid concentrations	Medium HDL ratios	Yes
245	S_HDL_PL_pct	Phospholipids to total lipids ratio in small HDL	%	Relative lipoprotein lipid concentrations	Small HDL ratios	Yes
246	S_HDL_C_pct	Cholesterol to total lipids ratio in small HDL	%	Relative lipoprotein lipid concentrations	Small HDL ratios	Yes
247	S_HDL_CE_pct	Cholesteryl esters to total lipids ratio in small HDL	%	Relative lipoprotein lipid concentrations	Small HDL ratios	Yes
248	S_HDL_FC_pct	Free cholesterol to total lipids ratio in small HDL	%	Relative lipoprotein lipid concentrations	Small HDL ratios	Yes
249	S_HDL_TG_pct	Triglycerides to total lipids ratio in small HDL	%	Relative lipoprotein lipid concentrations	Small HDL ratios	Yes

*Li and Ji method was applied to get 67 cluster numbers.

Supplementary Table 3.2 Variables for imputation of missing data

Variables	Number of missing data in whole cohort (n = 502,364)	%	Number of missing data in participants with metabolic data (n = 274,355)	%
Age	1	< 0.01	0	0
Sex	1	< 0.01	0	0
Body mass index (BMI)	3,105	0.62	1,027	0.37
Ethnicity	2,777	0.55	1,229	0.45
Diastolic blood pressure	30,110	5.99	13,545	4.94
Systolic blood pressure	30,115	5.99	13,548	4.94
Alcohol intake frequency	1,501	0.3	556	0.2
Smoking status	2,949	0.59	1,336	0.49
Physical activity based on International Physical Activity Questionnaire	100,093	19.92	53,163	19.38
Fasting time	1,233	0.25	5	< 0.01
Waist-hip ratios	2,266	0.45	608	0.22
Blood pressure medication	8,603	1.71	2,367	0.86
Vitamin and mineral supplements	5,591	1.11	683	0.25
Fresh fruit	3,239	0.64	1,480	0.54
Whole grain foods	12,036	2.4	4,235	1.54
Refined grain foods	12,036	2.4	4,235	1.54
Caffeinated coffee	2,267	0.45	975	0.36
Decaffeinated coffee	2,267	0.45	975	0.36
Tea	2,193	0.44	934	0.34
Vegetables	10,948	2.18	5,689	2.07
Cheese	14,449	2.88	7,528	2.74
Unprocessed red meat	6,999	1.39	3,533	1.29
Processed meat	2,230	0.44	941	0.34
Oily fish	4,079	0.81	1,950	0.71
Non oily fish	3,723	0.74	1,713	0.62
Spread type	2,039	0.41	805	0.29
Education	10,130	2.02	3,193	1.16

Supplementary Table 3.3 The original dataset and the imputed dataset

	Total (n = 502364)	**Original (n = 499259)**	**Imputed (n = 3105)**
Body mass index, Median (Q1,Q3)	26.75 (24.14, 29.91)	26.74 (24.14, 29.91)	27.43 (24.81, 30.57)
Ethnicity, n (%)	Total (n = 502364)	Original (n = 499587)	Imputed (n = 2777)
Asian	9896 (1.97)	9879 (1.98)	17 (0.61)
Black	8068 (1.61)	8058 (1.61)	10 (0.36)
Chinese	1573 (0.31)	1573 (0.31)	0 (0)
Mixed	2953 (0.59)	2953 (0.59)	0 (0)
Others	4555 (0.91)	4555 (0.91)	0 (0)
White	475319 (94.62)	472569 (94.59)	2750 (99.03)
Diastolic blood pressure, mmHg, median (Q1, Q3)	Total (n = 502364)	Original (n = 472254)	Imputed (n = 30110)
	82 (75, 89)	82 (75, 89)	82 (75, 89.5)
Systolic blood pressure, mmHg, median (Q1, Q3)	Total (n = 502364)	Original (n = 472249)	Imputed (n = 30115)
	136 (124.5, 149.5)	136 (124.5, 149.5)	136 (123.5, 149.5)
Alcohol intake frequency, n (%)	Total (n = 502364)	Original (n = 500863)	Imputed (n = 1501)
Daily or almost daily	102309 (20.37)	101745 (20.31)	564 (37.57)
Three or four times a week	115864 (23.06)	115413 (23.04)	451 (30.05)
Once or twice a week	129519 (25.78)	129257 (25.81)	262 (17.46)
One to three times a month	55836 (11.11)	55836 (11.15)	0 (0)
Special occasions only	58040 (11.55)	57989 (11.58)	51 (3.4)
Non-drinker	40796 (8.12)	40623 (8.11)	173 (11.53)
Smoking status, n (%)	Total (n = 502364)	Original (n = 499415)	Imputed (n = 2949)
Never	274033 (54.55)	273449 (54.75)	584 (19.8)
Previous	174907 (34.82)	173006 (34.64)	1901 (64.46)
Current	53424 (10.63)	52960 (10.6)	464 (15.73)
Variables	Total (n = 502364)	Original (n = 402271)	Imputed (n = 100093)
Physical activity based on International Physical Activity Questionnaire, n (%)			
Low	97275 (19.36)	76190 (18.94)	21085 (21.07)

	Total (n = 502364)	**Original (n = 499259)**	**Imputed (n = 3105)**
Moderate	211118 (42.02)	163986 (40.77)	47132 (47.09)
High	193971 (38.61)	162095 (40.29)	31876 (31.85)
Education, n (%)	Total (n = 502364)	Original (n = 492234)	Imputed (n = 10130)
College or University degree	29088 (5.79)	27225 (5.53)	1863 (18.39)
NVQ or HND or HNC or equivalent	164085 (32.66)	161102 (32.73)	2983 (29.45)
Other professional qualifications	76360 (15.2)	76190 (15.48)	170 (1.68)
A levels/AS levels or equivalent	86760 (17.27)	85256 (17.32)	1504 (14.85)
O levels/GCSEs/CSEs or equivalent	84651 (16.85)	83274 (16.92)	1377 (13.59)
None of the above	61420 (12.23)	59187 (12.02)	2233 (22.04)
Fasting time, hours, median (Q1, Q3)	Total (n = 502364)	Original (n = 501131)	Imputed (n = 1233)
	3 (2, 4)	3 (2, 4)	4 (2, 4)
Blood pressure medication, n (%)	Total (n = 502364)	Original (n = 493761)	Imputed (n = 8603)
No	395262 (78.68)	389779 (78.94)	5483 (63.73)
Yes	107102 (21.32)	103982 (21.06)	3120 (36.27)
Waist hip ratio, median (Q1,Q3)	Total (n = 502364)	Original (n = 500098)	Imputed (n = 2266)
	0.87 (0.8, 0.94)	0.87 (0.8, 0.94)	0.89 (0.82, 0.95)
Dietary supplements, n (%)	Total (n = 502364)	Original (n = 496773)	Imputed (n = 5591)
No	248602 (49.49)	244735 (49.26)	3867 (69.16)
Yes	253762 (50.51)	252038 (50.74)	1724 (30.84)
Fresh fruit, n (%)	Total (n = 502364)	Original (n = 499125)	Imputed (n = 3239)
< 1 servings/day	47014 (9.36)	46597 (9.34)	417 (12.87)
≥ 1 and < 2 servings/day	131558 (26.19)	130625 (26.17)	933 (28.81)
≥ 2 and < 3 servings/day	141711 (28.21)	140865 (28.22)	846 (26.12)
≥ 3 servings/day	182081 (36.24)	181038 (36.27)	1043 (32.2)
Whole grain foods, n (%)	Total (n = 502364)	Original (n = 490328)	Imputed (n = 12036)
0 serve/day	143922 (28.65)	139916 (28.54)	4006 (33.28)
> 0 and < 1 serve/day	197624 (39.34)	192877 (39.34)	4747 (39.44)

	Total (n = 502364)	**Original (n = 499259)**	**Imputed (n = 3105)**
≥ 1 and < 3 serves/day	160383 (31.93)	157110 (32.04)	3273 (27.19)
≥ 3 serves/day	435 (0.09)	425 (0.09)	10 (0.08)
Refined grain foods, n (%)	Total (n = 502364)	Original (n = 490328)	Imputed (n = 12036)
0 time/day	9160 (1.82)	8949 (1.83)	211 (1.75)
> 0 and < 1 serve/day	114549 (22.8)	111712 (22.78)	2837 (23.57)
≥ 1 and < 3 serves/day	297191 (59.16)	290114 (59.17)	7077 (58.8)
≥ 3 serves/day	81464 (16.22)	79553 (16.22)	1911 (15.88)
Caffeinated coffee, n (%)	Total (n = 502364)	Original (n = 500097)	Imputed (n = 2267)
0 cup/day	186368 (37.1)	185515 (37.1)	853 (37.63)
> 0 and < 2 cups/day	114102 (22.71)	113610 (22.72)	492 (21.7)
≥ 2 and < 4 cups/day	126891 (25.26)	126319 (25.26)	572 (25.23)
≥ 4 cups/day	75003 (14.93)	74653 (14.93)	350 (15.44)
Decaffeinated coffee, n (%)	Total (n = 502364)	Original (n = 500097)	Imputed (n = 2267)
0 cup/day	423050 (84.21)	421130 (84.21)	1920 (84.69)
> 0 and < 3 cups/day	47983 (9.55)	47751 (9.55)	232 (10.23)
≥ 3 cups/day	31331 (6.24)	31216 (6.24)	115 (5.07)
Tea, n (%)	Total (n = 502364)	Original (n = 500171)	Imputed (n = 2193)
0 cup/day	73830 (14.7)	73500 (14.69)	330 (15.05)
> 0 and < 3 cups/day	130231 (25.92)	129689 (25.93)	542 (24.72)
≥ 3 and < 5 cups/day	147044 (29.27)	146412 (29.27)	632 (28.82)
≥ 5 cups/day	151259 (30.11)	150570 (30.1)	689 (31.42)
Vegetables servings, n (%)	Total (n = 502364)	Original (n = 491416)	Imputed (n = 10948)
< 1 serving/day	31104 (6.19)	30207 (6.15)	897 (8.19)
≥ 1 and < 2 servings/day	142861 (28.44)	139515 (28.39)	3346 (30.56)
≥ 2 and < 3 servings/day	169884 (33.82)	166290 (33.84)	3594 (32.83)
≥ 3 servings/day	158515 (31.55)	155404 (31.62)	3111 (28.42)
Cheese, n (%)	Total (n = 502364)	Original (n = 487915)	Imputed (n = 14449)

	Total (n = 502364)	**Original (n = 499259)**	**Imputed (n = 3105)**
< 1 time/week	101921 (20.29)	98295 (20.15)	3626 (25.1)
≥ 1 and < 2 times/week	108084 (21.52)	104775 (21.47)	3309 (22.9)
≥ 2 and < 5 times/week	226510 (45.09)	220535 (45.2)	5975 (41.35)
≥ 5 times/week	65849 (13.11)	64310 (13.18)	1539 (10.65)
Unprocessed red meat, n (%)	Total (n = 502364)	Original (n = 495365)	Imputed (n = 6999)
0 time/week	34085 (6.78)	33724 (6.81)	361 (5.16)
< 1 time/week	16945 (3.37)	16681 (3.37)	264 (3.77)
≥ 1 and < 2 times/week	41675 (8.3)	41003 (8.28)	672 (9.6)
≥ 2 times/week	409659 (81.55)	403957 (81.55)	5702 (81.47)
Processed meat, n (%)	Total (n = 502364)	Original (n = 500134)	Imputed (n = 2230)
0 time/week	46976 (9.35)	46779 (9.35)	197 (8.83)
< 1 time/week	152983 (30.45)	152299 (30.45)	684 (30.67)
≥ 1 and < 2 times/week	146610 (29.18)	145990 (29.19)	620 (27.8)
≥ 2 times/week	155795 (31.01)	155066 (31)	729 (32.69)
Oily fish, n (%)	Total (n = 502364)	Original (n = 498285)	Imputed (n = 4079)
0 time/week	55337 (11.02)	54815 (11)	522 (12.8)
< 1 time/week	166308 (33.11)	164839 (33.08)	1469 (36.01)
≥ 1 and < 2 times/week	189899 (37.8)	188452 (37.82)	1447 (35.47)
≥ 2 times/week	90820 (18.08)	90179 (18.1)	641 (15.71)
Non-oily fish, n (%)	Total (n = 502364)	Original (n = 498641)	Imputed (n = 3723)
0 time/week	23730 (4.72)	23482 (4.71)	248 (6.66)
< 1 time/week	145968 (29.06)	144808 (29.04)	1160 (31.16)
≥ 1 and < 2 times/week	250050 (49.77)	248316 (49.8)	1734 (46.58)
≥ 2 times/week	82616 (16.45)	82035 (16.45)	581 (15.61)
Spread type, n (%)	Total (n = 502364)	Original (n = 500325)	Imputed (n = 2039)
Never/rarely use spread	55412 (11.03)	55178 (11.03)	234 (11.48)
Butter/spreadable butter	183933 (36.61)	183177 (36.61)	756 (37.08)

	Total (n = 502364)	Original (n = 499259)	Imputed (n = 3105)
Other type of spread/margarine	263019 (52.36)	261970 (52.36)	1049 (51.45)

NVQ, national vocational qualification; HND, higher national diploma; HNC, higher national certificate; GCSE, General Certificate of Secondary Education; CSEs, certificates of secondary education.

Supplementary Table 4.1 Association of individual metabolites with incident ischaemic stroke in Cox proportional hazards regression in model 1 and 2 (n = 184,564)

		Model 1					Model 2					
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***
1	Total_C	-0.014	0.020	0.99 (0.95-1.03)	4.91E-01	5.96E-01	-0.011	0.020	0.99 (0.95-1.03)	5.77E-01	6.73E-01	NA
2	non_HDL_C	0.017	0.020	1.02 (0.98-1.06)	3.75E-01	4.82E-01	0.017	0.020	1.02 (0.98-1.06)	3.84E-01	5.28E-01	NA
3	Remnant_C	0.027	0.020	1.03 (0.99-1.07)	1.73E-01	2.73E-01	0.025	0.020	1.03 (0.99-1.07)	2.03E-01	3.44E-01	NA
4	VLDL_C	0.037	0.020	1.04 (1-1.08)	5.64E-02	1.18E-01	0.032	0.020	1.03 (0.99-1.07)	1.03E-01	2.22E-01	NA
5	Clinical_LDL_C	0.011	0.019	1.01 (0.97-1.05)	5.69E-01	6.62E-01	0.013	0.019	1.01 (0.98-1.05)	5.04E-01	6.18E-01	NA
6	LDL_C	0.008	0.019	1.01 (0.97-1.05)	6.91E-01	7.51E-01	0.009	0.019	1.01 (0.97-1.05)	6.51E-01	7.30E-01	NA
7	HDL_C	-0.097	0.021	0.91 (0.87-0.95)	2.50E-06	4.57E-05	-0.086	0.021	0.92 (0.88-0.96)	2.64E-05	5.98E-04	11%
8	Total_TG	0.039	0.019	1.04 (1-1.08)	4.49E-02	1.04E-01	0.028	0.019	1.03 (0.99-1.07)	1.53E-01	2.84E-01	NA
9	VLDL_TG	0.032	0.019	1.03 (0.99-1.07)	9.96E-02	1.78E-01	0.022	0.019	1.02 (0.98-1.06)	2.60E-01	4.12E-01	NA
10	LDL_TG	0.073	0.019	1.08 (1.04-1.12)	1.11E-04	6.72E-04	0.059	0.019	1.06 (1.02-1.1)	1.93E-03	9.62E-03	-19%
11	HDL_TG	0.029	0.019	1.03 (0.99-1.07)	1.29E-01	2.17E-01	0.019	0.019	1.02 (0.98-1.06)	3.27E-01	4.62E-01	NA
12	Total_PL	-0.020	0.020	0.98 (0.94-1.02)	3.11E-01	4.27E-01	-0.021	0.020	0.98 (0.94-1.02)	2.86E-01	4.40E-01	NA
13	VLDL_PL	0.042	0.019	1.04 (1-1.08)	3.27E-02	8.22E-02	0.033	0.019	1.03 (1-1.07)	9.03E-02	2.04E-01	NA
14	LDL_PL	0.020	0.019	1.02 (0.98-1.06)	3.00E-01	4.16E-01	0.020	0.019	1.02 (0.98-1.06)	3.07E-01	4.52E-01	NA
15	HDL_PL	-0.076	0.021	0.93 (0.89-0.97)	2.19E-04	1.27E-03	-0.071	0.020	0.93 (0.9-0.97)	5.55E-04	3.74E-03	NA
16	Total_CE	-0.018	0.020	0.98 (0.95-1.02)	3.72E-01	4.80E-01	-0.014	0.020	0.99 (0.95-1.03)	4.67E-01	5.87E-01	NA
17	VLDL_CE	0.037	0.020	1.04 (1-1.08)	5.96E-02	1.21E-01	0.033	0.020	1.03 (1-1.07)	9.34E-02	2.09E-01	NA
18	LDL_CE	0.010	0.019	1.01 (0.97-1.05)	6.16E-01	6.94E-01	0.010	0.019	1.01 (0.97-1.05)	6.13E-01	7.01E-01	NA
19	HDL_CE	-0.102	0.020	0.9 (0.87-0.94)	6.71E-07	1.52E-05	-0.090	0.020	0.91 (0.88-0.95)	9.98E-06	3.11E-04	11%
20	Total_FC	-0.003	0.020	1 (0.96-1.04)	8.67E-01	9.07E-01	-0.002	0.020	1 (0.96-1.04)	9.02E-01	9.28E-01	NA
21	VLDL_FC	0.037	0.020	1.04 (1-1.08)	5.76E-02	1.19E-01	0.030	0.020	1.03 (0.99-1.07)	1.26E-01	2.47E-01	NA
22	LDL_FC	0.002	0.019	1 (0.96-1.04)	9.24E-01	9.55E-01	0.006	0.019	1.01 (0.97-1.05)	7.72E-01	8.25E-01	NA
23	HDL_FC	-0.074	0.021	0.93 (0.89-0.97)	3.78E-04	1.88E-03	-0.067	0.021	0.94 (0.9-0.97)	1.23E-03	6.54E-03	NA
24	Total_L	-0.001	0.019	1 (0.96-1.04)	9.77E-01	9.82E-01	-0.003	0.019	1 (0.96-1.04)	8.65E-01	8.97E-01	NA
25	VLDL_L	0.037	0.019	1.04 (1-1.08)	5.84E-02	1.19E-01	0.028	0.019	1.03 (0.99-1.07)	1.52E-01	2.84E-01	NA
26	LDL_L	0.016	0.019	1.02 (0.98-1.06)	4.22E-01	5.31E-01	0.015	0.019	1.02 (0.98-1.06)	4.31E-01	5.63E-01	NA
27	HDL_L	-0.085	0.021	0.92 (0.88-0.96)	4.05E-05	3.15E-04	-0.078	0.021	0.93 (0.89-0.96)	1.61E-04	1.75E-03	8%
28	Total_P	-0.072	0.020	0.93 (0.9-0.97)	3.05E-04	1.62E-03	-0.066	0.020	0.94 (0.9-0.97)	8.84E-04	5.37E-03	NA
29	VLDL_P	0.041	0.019	1.04 (1-1.08)	3.27E-02	8.22E-02	0.034	0.019	1.04 (1-1.08)	8.02E-02	1.91E-01	NA
30	LDL_P	0.026	0.019	1.03 (0.99-1.07)	1.78E-01	2.77E-01	0.024	0.019	1.03 (0.99-1.06)	2.07E-01	3.46E-01	NA

		Model 1					Model 2					
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***
31	HDL_P	-0.080	0.020	0.92 (0.89-0.96)	5.27E-05	3.75E-04	-0.073	0.020	0.93 (0.89-0.97)	2.05E-04	1.96E-03	8%
32	VLDL_size	0.009	0.020	1.01 (0.97-1.05)	6.33E-01	7.10E-01	0.001	0.020	1 (0.96-1.04)	9.43E-01	9.61E-01	NA
33	LDL_size	-0.019	0.019	0.98 (0.95-1.02)	3.22E-01	4.31E-01	-0.017	0.019	0.98 (0.95-1.02)	3.72E-01	5.14E-01	NA
34	HDL_size	-0.078	0.021	0.93 (0.89-0.96)	2.52E-04	1.39E-03	-0.070	0.021	0.93 (0.89-0.97)	1.13E-03	6.53E-03	NA
35	Phosphoglyc	-0.029	0.020	0.97 (0.93-1.01)	1.39E-01	2.30E-01	-0.030	0.020	0.97 (0.93-1.01)	1.37E-01	2.67E-01	NA
36	TG_by_PG	0.058	0.020	1.06 (1.02-1.1)	3.73E-03	1.29E-02	0.045	0.020	1.05 (1.01-1.09)	2.37E-02	7.87E-02	NA
37	Cholines	-0.036	0.020	0.97 (0.93-1)	7.48E-02	1.43E-01	-0.035	0.020	0.97 (0.93-1)	8.26E-02	1.92E-01	NA
38	Phosphatidylc	-0.041	0.020	0.96 (0.92-1)	4.01E-02	9.52E-02	-0.040	0.020	0.96 (0.92-1)	4.63E-02	1.32E-01	NA
39	Sphingomyelins	0.008	0.020	1.01 (0.97-1.05)	7.06E-01	7.64E-01	0.007	0.020	1.01 (0.97-1.05)	7.25E-01	7.81E-01	NA
40	ApoB	0.031	0.019	1.03 (0.99-1.07)	1.11E-01	1.95E-01	0.029	0.019	1.03 (0.99-1.07)	1.40E-01	2.68E-01	NA
41	ApoA1	-0.083	0.020	0.92 (0.89-0.96)	4.71E-05	3.55E-04	-0.076	0.020	0.93 (0.89-0.96)	1.73E-04	1.80E-03	8%
42	ApoB_by_ApoA1	0.072	0.020	1.07 (1.03-1.12)	2.62E-04	1.42E-03	0.066	0.020	1.07 (1.03-1.11)	7.50E-04	4.67E-03	NA
43	Total_FA	0.021	0.019	1.02 (0.98-1.06)	2.65E-01	3.76E-01	0.014	0.019	1.01 (0.98-1.05)	4.70E-01	5.89E-01	NA
44	Unsaturation	-0.121	0.019	0.89 (0.85-0.92)	9.26E-11	6.70E-09	-0.095	0.019	0.91 (0.88-0.94)	5.08E-07	3.16E-05	22%
45	Omega_3	-0.084	0.019	0.92 (0.89-0.96)	1.22E-05	1.38E-04	-0.068	0.019	0.94 (0.9-0.97)	4.48E-04	3.34E-03	20%
46	Omega_6	-0.028	0.019	0.97 (0.94-1.01)	1.51E-01	2.44E-01	-0.026	0.019	0.97 (0.94-1.01)	1.77E-01	3.09E-01	NA
47	PUFA	-0.043	0.019	0.96 (0.92-1)	2.58E-02	7.15E-02	-0.037	0.019	0.96 (0.93-1)	5.35E-02	1.47E-01	NA
48	MUFA	0.062	0.019	1.06 (1.03-1.11)	1.05E-03	4.38E-03	0.047	0.019	1.05 (1.01-1.09)	1.28E-02	4.68E-02	NA
49	SFA	0.037	0.019	1.04 (1-1.08)	5.30E-02	1.14E-01	0.026	0.019	1.03 (0.99-1.07)	1.70E-01	2.99E-01	NA
50	LA	-0.038	0.019	0.96 (0.93-1)	4.85E-02	1.09E-01	-0.035	0.019	0.97 (0.93-1)	6.87E-02	1.74E-01	NA
51	DHA	-0.091	0.019	0.91 (0.88-0.95)	1.03E-06	2.14E-05	-0.071	0.019	0.93 (0.9-0.97)	1.62E-04	1.75E-03	22%
52	Omega_3_pct	-0.113	0.019	0.89 (0.86-0.93)	4.16E-09	1.48E-07	-0.090	0.019	0.91 (0.88-0.95)	3.52E-06	1.25E-04	21%
53	Omega_6_pct	-0.079	0.018	0.92 (0.89-0.96)	1.24E-05	1.38E-04	-0.063	0.018	0.94 (0.91-0.97)	5.26E-04	3.64E-03	20%
54	PUFA_pct	-0.110	0.018	0.9 (0.87-0.93)	6.90E-10	2.86E-08	-0.088	0.018	0.92 (0.88-0.95)	1.20E-06	4.99E-05	20%
55	MUFA_pct	0.119	0.019	1.13 (1.09-1.17)	4.71E-10	2.34E-08	0.096	0.019	1.1 (1.06-1.14)	6.40E-07	3.19E-05	-20%
56	SFA_pct	0.066	0.018	1.07 (1.03-1.11)	3.34E-04	1.70E-03	0.051	0.018	1.05 (1.02-1.09)	5.69E-03	2.36E-02	NA
57	LA_pct	-0.093	0.019	0.91 (0.88-0.95)	5.73E-07	1.52E-05	-0.077	0.019	0.93 (0.89-0.96)	3.59E-05	6.38E-04	17%
58	DHA_pct	-0.101	0.018	0.9 (0.87-0.94)	2.71E-08	8.45E-07	-0.078	0.018	0.93 (0.89-0.96)	2.46E-05	5.98E-04	23%
59	PUFA_by_MUFA	-0.120	0.019	0.89 (0.86-0.92)	1.08E-10	6.70E-09	-0.096	0.019	0.91 (0.88-0.94)	2.93E-07	2.43E-05	20%
60	Omega_6_by_Omega_3	0.083	0.019	1.09 (1.05-1.13)	1.50E-05	1.50E-04	0.066	0.019	1.07 (1.03-1.11)	6.80E-04	4.46E-03	-21%
61	Ala	0.037	0.023	1.04 (0.99-1.09)	1.11E-01	1.95E-01	0.049	0.023	1.05 (1-1.1)	3.72E-02	1.10E-01	NA
62	Gln	-0.032	0.019	0.97 (0.93-1.01)	9.92E-02	1.78E-01	-0.034	0.019	0.97 (0.93-1)	7.67E-02	1.89E-01	NA

		Model 1					Model 2					
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***
63	Gly	-0.011	0.017	0.99 (0.96-1.02)	5.20E-01	6.17E-01	-0.012	0.017	0.99 (0.96-1.02)	4.55E-01	5.81E-01	NA
64	His	-0.060	0.019	0.94 (0.91-0.98)	1.11E-03	4.55E-03	-0.056	0.019	0.95 (0.91-0.98)	2.65E-03	1.20E-02	NA
65	Total_BCAA	-0.001	0.020	1 (0.96-1.04)	9.74E-01	9.82E-01	0.005	0.020	1.01 (0.97-1.05)	8.01E-01	8.45E-01	NA
66	Ile	0.011	0.019	1.01 (0.97-1.05)	5.51E-01	6.48E-01	0.015	0.019	1.02 (0.98-1.05)	4.40E-01	5.70E-01	NA
67	Leu	-0.009	0.020	0.99 (0.95-1.03)	6.49E-01	7.24E-01	-0.005	0.020	1 (0.96-1.03)	8.04E-01	8.45E-01	NA
68	Val	0.000	0.020	1 (0.96-1.04)	9.82E-01	9.82E-01	0.007	0.020	1.01 (0.97-1.05)	7.41E-01	7.95E-01	NA
69	Phe	0.066	0.019	1.07 (1.03-1.11)	5.28E-04	2.39E-03	0.068	0.019	1.07 (1.03-1.11)	3.03E-04	2.51E-03	NA
70	Tyr	0.041	0.019	1.04 (1-1.08)	3.23E-02	8.22E-02	0.046	0.019	1.05 (1.01-1.09)	1.64E-02	5.66E-02	NA
71	Glucose	0.038	0.019	1.04 (1-1.08)	4.97E-02	1.11E-01	0.047	0.019	1.05 (1.01-1.09)	1.45E-02	5.15E-02	NA
72	Lactate	0.081	0.020	1.08 (1.04-1.13)	5.08E-05	3.72E-04	0.070	0.020	1.07 (1.03-1.12)	4.56E-04	3.34E-03	-14%
73	Pyruvate	0.044	0.020	1.05 (1.01-1.09)	2.70E-02	7.24E-02	0.039	0.020	1.04 (1-1.08)	4.67E-02	1.32E-01	NA
74	Citrate	-0.009	0.020	0.99 (0.95-1.03)	6.51E-01	7.24E-01	0.006	0.020	1.01 (0.97-1.05)	7.78E-01	8.28E-01	NA
75	bOHbutyrate	0.024	0.021	1.03 (0.98-1.07)	2.57E-01	3.73E-01	0.017	0.021	1.02 (0.98-1.06)	4.32E-01	5.63E-01	NA
76	Acetate	-0.031	0.016	0.97 (0.94-1)	4.86E-02	1.09E-01	-0.025	0.016	0.98 (0.95-1.01)	1.26E-01	2.47E-01	NA
77	Acetoacetate	0.001	0.020	1 (0.96-1.04)	9.58E-01	9.74E-01	-0.003	0.020	1 (0.96-1.04)	8.63E-01	8.97E-01	NA
78	Acetone	-0.007	0.019	0.99 (0.96-1.03)	7.28E-01	7.85E-01	-0.010	0.019	0.99 (0.95-1.03)	5.81E-01	6.73E-01	NA
79	Creatinine	0.058	0.022	1.06 (1.02-1.11)	8.21E-03	2.63E-02	0.064	0.022	1.07 (1.02-1.11)	3.75E-03	1.64E-02	NA
80	Albumin	-0.160	0.018	0.85 (0.82-0.88)	3.86E-18	9.61E-16	-0.149	0.018	0.86 (0.83-0.89)	6.39E-16	1.59E-13	7%
81	GlycA	0.146	0.020	1.16 (1.11-1.2)	8.94E-14	1.11E-11	0.126	0.020	1.13 (1.09-1.18)	1.64E-10	2.04E-08	-14%
82	XXL_VLDL_P	0.039	0.020	1.04 (1-1.08)	5.56E-02	1.18E-01	0.029	0.020	1.03 (0.99-1.07)	1.62E-01	2.90E-01	NA
83	XXL_VLDL_L	0.031	0.020	1.03 (0.99-1.07)	1.17E-01	2.02E-01	0.021	0.020	1.02 (0.98-1.06)	2.77E-01	4.37E-01	NA
84	XXL_VLDL_PL	0.036	0.021	1.04 (1-1.08)	8.82E-02	1.64E-01	0.026	0.021	1.03 (0.99-1.07)	2.20E-01	3.65E-01	NA
85	XXL_VLDL_C	0.041	0.020	1.04 (1-1.08)	4.15E-02	9.75E-02	0.031	0.020	1.03 (0.99-1.07)	1.24E-01	2.47E-01	NA
86	XXL_VLDL_CE	0.051	0.020	1.05 (1.01-1.1)	1.30E-02	4.01E-02	0.041	0.020	1.04 (1-1.08)	4.54E-02	1.31E-01	NA
87	XXL_VLDL_FC	0.027	0.020	1.03 (0.99-1.07)	1.73E-01	2.73E-01	0.017	0.020	1.02 (0.98-1.06)	3.94E-01	5.35E-01	NA
88	XXL_VLDL_TG	0.014	0.019	1.01 (0.98-1.05)	4.79E-01	5.85E-01	0.007	0.019	1.01 (0.97-1.05)	7.09E-01	7.70E-01	NA
89	XL_VLDL_P	0.046	0.020	1.05 (1.01-1.09)	2.28E-02	6.60E-02	0.035	0.020	1.04 (1-1.08)	8.34E-02	1.92E-01	NA
90	XL_VLDL_L	0.043	0.020	1.04 (1-1.09)	3.18E-02	8.22E-02	0.033	0.020	1.03 (0.99-1.07)	1.04E-01	2.22E-01	NA
91	XL_VLDL_PL	0.045	0.020	1.05 (1.01-1.09)	2.70E-02	7.24E-02	0.035	0.020	1.04 (1-1.08)	8.75E-02	2.00E-01	NA
92	XL_VLDL_C	0.044	0.020	1.05 (1.01-1.09)	2.68E-02	7.24E-02	0.036	0.020	1.04 (1-1.08)	7.34E-02	1.83E-01	NA
93	XL_VLDL_CE	0.045	0.020	1.05 (1.01-1.09)	2.45E-02	6.84E-02	0.038	0.020	1.04 (1-1.08)	5.76E-02	1.54E-01	NA
94	XL_VLDL_FC	0.042	0.020	1.04 (1-1.09)	3.52E-02	8.68E-02	0.032	0.020	1.03 (0.99-1.07)	1.05E-01	2.22E-01	NA

		Model 1					Model 2					
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***
95	XL_VLDL_TG	0.040	0.020	1.04 (1-1.08)	4.58E-02	1.05E-01	0.029	0.020	1.03 (0.99-1.07)	1.48E-01	2.80E-01	NA
96	L_VLDL_P	0.038	0.020	1.04 (1-1.08)	5.59E-02	1.18E-01	0.028	0.020	1.03 (0.99-1.07)	1.56E-01	2.87E-01	NA
97	L_VLDL_L	0.033	0.020	1.03 (0.99-1.07)	9.75E-02	1.78E-01	0.024	0.020	1.02 (0.99-1.06)	2.31E-01	3.81E-01	NA
98	L_VLDL_PL	0.043	0.020	1.04 (1-1.09)	3.36E-02	8.37E-02	0.032	0.020	1.03 (0.99-1.08)	1.09E-01	2.25E-01	NA
99	L_VLDL_C	0.036	0.020	1.04 (1-1.08)	7.21E-02	1.39E-01	0.027	0.020	1.03 (0.99-1.07)	1.68E-01	2.96E-01	NA
100	L_VLDL_CE	0.030	0.020	1.03 (0.99-1.07)	1.34E-01	2.25E-01	0.023	0.020	1.02 (0.98-1.06)	2.51E-01	4.06E-01	NA
101	L_VLDL_FC	0.041	0.020	1.04 (1-1.08)	3.89E-02	9.42E-02	0.032	0.020	1.03 (0.99-1.07)	1.13E-01	2.31E-01	NA
102	L_VLDL_TG	0.025	0.019	1.03 (0.99-1.07)	2.00E-01	3.06E-01	0.017	0.019	1.02 (0.98-1.06)	3.96E-01	5.35E-01	NA
103	M_VLDL_P	0.020	0.019	1.02 (0.98-1.06)	3.14E-01	4.27E-01	0.015	0.019	1.02 (0.98-1.06)	4.26E-01	5.62E-01	NA
104	M_VLDL_L	0.020	0.019	1.02 (0.98-1.06)	3.12E-01	4.27E-01	0.014	0.019	1.01 (0.98-1.05)	4.60E-01	5.84E-01	NA
105	M_VLDL_PL	0.022	0.019	1.02 (0.98-1.06)	2.58E-01	3.73E-01	0.018	0.019	1.02 (0.98-1.06)	3.57E-01	4.99E-01	NA
106	M_VLDL_C	0.012	0.019	1.01 (0.97-1.05)	5.51E-01	6.48E-01	0.012	0.019	1.01 (0.98-1.05)	5.18E-01	6.33E-01	NA
107	M_VLDL_CE	0.008	0.019	1.01 (0.97-1.05)	6.59E-01	7.28E-01	0.012	0.019	1.01 (0.98-1.05)	5.36E-01	6.51E-01	NA
108	M_VLDL_FC	0.019	0.019	1.02 (0.98-1.06)	3.40E-01	4.50E-01	0.016	0.019	1.02 (0.98-1.06)	4.05E-01	5.45E-01	NA
109	M_VLDL_TG	0.021	0.019	1.02 (0.98-1.06)	2.74E-01	3.85E-01	0.013	0.019	1.01 (0.98-1.05)	5.01E-01	6.18E-01	NA
110	S_VLDL_P	0.036	0.019	1.04 (1-1.08)	6.22E-02	1.24E-01	0.028	0.019	1.03 (0.99-1.07)	1.46E-01	2.77E-01	NA
111	S_VLDL_L	0.036	0.020	1.04 (1-1.08)	6.13E-02	1.23E-01	0.029	0.019	1.03 (0.99-1.07)	1.40E-01	2.68E-01	NA
112	S_VLDL_PL	0.032	0.019	1.03 (0.99-1.07)	9.59E-02	1.77E-01	0.027	0.019	1.03 (0.99-1.07)	1.60E-01	2.88E-01	NA
113	S_VLDL_C	0.041	0.020	1.04 (1-1.08)	3.79E-02	9.26E-02	0.036	0.020	1.04 (1-1.08)	6.58E-02	1.71E-01	NA
114	S_VLDL_CE	0.048	0.020	1.05 (1.01-1.09)	1.50E-02	4.55E-02	0.042	0.020	1.04 (1-1.08)	3.23E-02	9.78E-02	NA
115	S_VLDL_FC	0.027	0.019	1.03 (0.99-1.07)	1.63E-01	2.60E-01	0.025	0.019	1.03 (0.99-1.07)	2.07E-01	3.46E-01	NA
116	S_VLDL_TG	0.030	0.020	1.03 (0.99-1.07)	1.27E-01	2.17E-01	0.020	0.020	1.02 (0.98-1.06)	3.15E-01	4.59E-01	NA
117	XS_VLDL_P	0.060	0.020	1.06 (1.02-1.1)	2.21E-03	8.35E-03	0.053	0.020	1.06 (1.02-1.1)	6.33E-03	2.59E-02	NA
118	XS_VLDL_L	0.058	0.020	1.06 (1.02-1.1)	2.92E-03	1.03E-02	0.052	0.020	1.05 (1.01-1.09)	8.81E-03	3.32E-02	NA
119	XS_VLDL_PL	0.068	0.020	1.07 (1.03-1.11)	5.38E-04	2.39E-03	0.059	0.020	1.06 (1.02-1.1)	2.47E-03	1.14E-02	NA
120	XS_VLDL_C	0.045	0.020	1.05 (1.01-1.09)	2.43E-02	6.84E-02	0.043	0.020	1.04 (1-1.09)	3.20E-02	9.78E-02	NA
121	XS_VLDL_CE	0.038	0.020	1.04 (1-1.08)	5.74E-02	1.19E-01	0.037	0.020	1.04 (1-1.08)	6.15E-02	1.63E-01	NA
122	XS_VLDL_FC	0.059	0.020	1.06 (1.02-1.1)	2.89E-03	1.03E-02	0.053	0.020	1.06 (1.02-1.1)	6.98E-03	2.80E-02	NA
123	XS_VLDL_TG	0.058	0.019	1.06 (1.02-1.1)	2.43E-03	9.05E-03	0.045	0.019	1.05 (1.01-1.09)	1.89E-02	6.44E-02	NA
124	IDL_P	0.036	0.020	1.04 (1-1.08)	6.68E-02	1.31E-01	0.035	0.020	1.04 (1-1.08)	7.28E-02	1.83E-01	NA
125	IDL_L	0.020	0.020	1.02 (0.98-1.06)	3.22E-01	4.31E-01	0.021	0.020	1.02 (0.98-1.06)	2.96E-01	4.44E-01	NA
126	IDL_PL	0.025	0.020	1.03 (0.99-1.07)	2.06E-01	3.13E-01	0.025	0.020	1.03 (0.99-1.07)	2.01E-01	3.43E-01	NA

		Model 1					Model 2					
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***
127	IDL_C	0.008	0.020	1.01 (0.97-1.05)	6.70E-01	7.35E-01	0.012	0.020	1.01 (0.97-1.05)	5.54E-01	6.61E-01	NA
128	IDL_CE	0.006	0.020	1.01 (0.97-1.05)	7.58E-01	8.04E-01	0.009	0.020	1.01 (0.97-1.05)	6.46E-01	7.28E-01	NA
129	IDL_FC	0.014	0.020	1.01 (0.98-1.06)	4.67E-01	5.73E-01	0.018	0.020	1.02 (0.98-1.06)	3.50E-01	4.93E-01	NA
130	IDL_TG	0.076	0.019	1.08 (1.04-1.12)	6.40E-05	4.30E-04	0.062	0.019	1.06 (1.03-1.11)	1.16E-03	6.54E-03	-18%
131	L_LDL_P	0.022	0.019	1.02 (0.99-1.06)	2.47E-01	3.62E-01	0.021	0.019	1.02 (0.98-1.06)	2.79E-01	4.37E-01	NA
132	L_LDL_L	0.010	0.020	1.01 (0.97-1.05)	5.94E-01	6.73E-01	0.011	0.020	1.01 (0.97-1.05)	5.57E-01	6.61E-01	NA
133	L_LDL_PL	0.013	0.020	1.01 (0.98-1.05)	4.94E-01	5.98E-01	0.015	0.020	1.02 (0.98-1.06)	4.44E-01	5.70E-01	NA
134	L_LDL_C	0.002	0.019	1 (0.97-1.04)	9.12E-01	9.50E-01	0.005	0.019	1.01 (0.97-1.04)	8.16E-01	8.54E-01	NA
135	L_LDL_CE	0.003	0.019	1 (0.97-1.04)	8.62E-01	9.05E-01	0.005	0.019	1.01 (0.97-1.04)	7.95E-01	8.43E-01	NA
136	L_LDL_FC	-0.001	0.020	1 (0.96-1.04)	9.44E-01	9.71E-01	0.003	0.020	1 (0.97-1.04)	8.81E-01	9.11E-01	NA
137	L_LDL_TG	0.081	0.019	1.09 (1.05-1.13)	1.87E-05	1.79E-04	0.067	0.019	1.07 (1.03-1.11)	4.75E-04	3.38E-03	-18%
138	M_LDL_P	0.028	0.020	1.03 (0.99-1.07)	1.58E-01	2.53E-01	0.026	0.020	1.03 (0.99-1.07)	1.78E-01	3.09E-01	NA
139	M_LDL_L	0.022	0.020	1.02 (0.98-1.06)	2.64E-01	3.76E-01	0.020	0.020	1.02 (0.98-1.06)	3.15E-01	4.59E-01	NA
140	M_LDL_PL	0.025	0.020	1.03 (0.99-1.07)	2.08E-01	3.13E-01	0.023	0.020	1.02 (0.98-1.06)	2.47E-01	4.02E-01	NA
141	M_LDL_C	0.016	0.020	1.02 (0.98-1.06)	4.11E-01	5.19E-01	0.015	0.020	1.02 (0.98-1.06)	4.44E-01	5.70E-01	NA
142	M_LDL_CE	0.020	0.020	1.02 (0.98-1.06)	3.18E-01	4.31E-01	0.017	0.020	1.02 (0.98-1.06)	3.86E-01	5.28E-01	NA
143	M_LDL_FC	0.006	0.019	1.01 (0.97-1.05)	7.73E-01	8.15E-01	0.008	0.019	1.01 (0.97-1.05)	6.62E-01	7.39E-01	NA
144	M_LDL_TG	0.065	0.019	1.07 (1.03-1.11)	6.70E-04	2.93E-03	0.051	0.019	1.05 (1.01-1.09)	7.53E-03	2.98E-02	NA
145	S_LDL_P	0.034	0.019	1.03 (1-1.07)	8.54E-02	1.61E-01	0.031	0.019	1.03 (0.99-1.07)	1.14E-01	2.31E-01	NA
146	S_LDL_L	0.028	0.019	1.03 (0.99-1.07)	1.47E-01	2.41E-01	0.025	0.019	1.03 (0.99-1.07)	1.97E-01	3.39E-01	NA
147	S_LDL_PL	0.036	0.019	1.04 (1-1.08)	6.43E-02	1.27E-01	0.032	0.019	1.03 (0.99-1.07)	1.01E-01	2.19E-01	NA
148	S_LDL_C	0.021	0.019	1.02 (0.98-1.06)	2.91E-01	4.04E-01	0.019	0.019	1.02 (0.98-1.06)	3.23E-01	4.60E-01	NA
149	S_LDL_CE	0.023	0.019	1.02 (0.99-1.06)	2.31E-01	3.42E-01	0.020	0.019	1.02 (0.98-1.06)	2.92E-01	4.44E-01	NA
150	S_LDL_FC	0.011	0.019	1.01 (0.97-1.05)	5.77E-01	6.63E-01	0.014	0.019	1.01 (0.98-1.05)	4.80E-01	5.97E-01	NA
151	S_LDL_TG	0.045	0.019	1.05 (1.01-1.09)	1.65E-02	4.88E-02	0.033	0.019	1.03 (1-1.07)	8.03E-02	1.91E-01	NA
152	XL_HDL_P	-0.044	0.021	0.96 (0.92-1)	3.22E-02	8.22E-02	-0.040	0.021	0.96 (0.92-1)	5.38E-02	1.47E-01	NA
153	XL_HDL_L	-0.040	0.021	0.96 (0.92-1)	5.11E-02	1.12E-01	-0.036	0.021	0.97 (0.93-1.01)	8.28E-02	1.92E-01	NA
154	XL_HDL_PL	-0.037	0.020	0.96 (0.93-1)	7.08E-02	1.38E-01	-0.033	0.020	0.97 (0.93-1.01)	1.00E-01	2.19E-01	NA
155	XL_HDL_C	-0.042	0.020	0.96 (0.92-1)	3.95E-02	9.45E-02	-0.036	0.020	0.97 (0.93-1)	7.85E-02	1.91E-01	NA
156	XL_HDL_CE	-0.053	0.020	0.95 (0.91-0.99)	9.80E-03	3.09E-02	-0.046	0.020	0.96 (0.92-0.99)	2.51E-02	8.11E-02	NA
157	XL_HDL_FC	-0.001	0.020	1 (0.96-1.04)	9.55E-01	9.74E-01	0.001	0.020	1 (0.96-1.04)	9.45E-01	9.61E-01	NA
158	XL_HDL_TG	0.017	0.019	1.02 (0.98-1.06)	3.81E-01	4.87E-01	0.008	0.019	1.01 (0.97-1.05)	6.65E-01	7.39E-01	NA

		Model 1					Model 2					
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***
159	L_HDL_P	-0.091	0.021	0.91 (0.88-0.95)	9.95E-06	1.24E-04	-0.082	0.021	0.92 (0.89-0.96)	7.33E-05	1.07E-03	10%
160	L_HDL_L	-0.087	0.021	0.92 (0.88-0.96)	3.09E-05	2.66E-04	-0.078	0.021	0.93 (0.89-0.96)	1.86E-04	1.86E-03	10%
161	L_HDL_PL	-0.082	0.021	0.92 (0.89-0.96)	8.46E-05	5.40E-04	-0.074	0.021	0.93 (0.89-0.97)	3.91E-04	3.14E-03	10%
162	L_HDL_C	-0.089	0.020	0.92 (0.88-0.95)	1.28E-05	1.38E-04	-0.079	0.020	0.92 (0.89-0.96)	1.10E-04	1.38E-03	11%
163	L_HDL_CE	-0.093	0.020	0.91 (0.88-0.95)	4.59E-06	6.35E-05	-0.082	0.020	0.92 (0.89-0.96)	5.20E-05	8.63E-04	12%
164	L_HDL_FC	-0.072	0.021	0.93 (0.89-0.97)	4.93E-04	2.32E-03	-0.064	0.021	0.94 (0.9-0.98)	1.77E-03	9.00E-03	NA
165	L_HDL_TG	-0.011	0.020	0.99 (0.95-1.03)	5.74E-01	6.62E-01	-0.016	0.020	0.98 (0.95-1.02)	4.20E-01	5.56E-01	NA
166	M_HDL_P	-0.087	0.020	0.92 (0.88-0.95)	1.41E-05	1.47E-04	-0.079	0.020	0.92 (0.89-0.96)	6.63E-05	1.03E-03	8%
167	M_HDL_L	-0.078	0.020	0.93 (0.89-0.96)	8.79E-05	5.47E-04	-0.073	0.020	0.93 (0.89-0.97)	2.58E-04	2.25E-03	7%
168	M_HDL_PL	-0.069	0.020	0.93 (0.9-0.97)	5.26E-04	2.39E-03	-0.065	0.020	0.94 (0.9-0.97)	9.84E-04	5.83E-03	NA
169	M_HDL_C	-0.092	0.020	0.91 (0.88-0.95)	4.32E-06	6.35E-05	-0.082	0.020	0.92 (0.89-0.96)	3.30E-05	6.33E-04	10%
170	M_HDL_CE	-0.093	0.020	0.91 (0.88-0.95)	2.57E-06	4.57E-05	-0.084	0.020	0.92 (0.89-0.96)	2.27E-05	5.98E-04	10%
171	M_HDL_FC	-0.081	0.020	0.92 (0.89-0.96)	5.51E-05	3.81E-04	-0.074	0.020	0.93 (0.89-0.97)	2.22E-04	2.05E-03	9%
172	M_HDL_TG	0.026	0.019	1.03 (0.99-1.07)	1.82E-01	2.81E-01	0.016	0.019	1.02 (0.98-1.06)	4.12E-01	5.49E-01	NA
173	S_HDL_P	-0.039	0.019	0.96 (0.93-1)	4.28E-02	9.95E-02	-0.038	0.019	0.96 (0.93-1)	5.11E-02	1.43E-01	NA
174	S_HDL_L	-0.029	0.019	0.97 (0.94-1.01)	1.28E-01	2.17E-01	-0.031	0.019	0.97 (0.93-1.01)	1.07E-01	2.23E-01	NA
175	S_HDL_PL	-0.029	0.019	0.97 (0.94-1.01)	1.29E-01	2.17E-01	-0.032	0.019	0.97 (0.93-1.01)	9.87E-02	2.19E-01	NA
176	S_HDL_C	-0.046	0.019	0.96 (0.92-0.99)	1.63E-02	4.88E-02	-0.043	0.019	0.96 (0.92-1)	2.48E-02	8.11E-02	NA
177	S_HDL_CE	-0.051	0.019	0.95 (0.92-0.99)	7.94E-03	2.60E-02	-0.047	0.019	0.95 (0.92-0.99)	1.47E-02	5.15E-02	NA
178	S_HDL_FC	-0.026	0.019	0.97 (0.94-1.01)	1.78E-01	2.77E-01	-0.027	0.020	0.97 (0.94-1.01)	1.67E-01	2.96E-01	NA
179	S_HDL_TG	0.055	0.020	1.06 (1.02-1.1)	5.08E-03	1.71E-02	0.042	0.020	1.04 (1-1.08)	3.26E-02	9.78E-02	NA
180	XXL_VLDL_PL_pct	0.029	0.023	1.03 (0.99-1.08)	2.00E-01	3.06E-01	0.022	0.023	1.02 (0.98-1.07)	3.23E-01	4.60E-01	NA
181	XXL_VLDL_C_pct	0.006	0.020	1.01 (0.97-1.05)	7.49E-01	8.00E-01	0.011	0.020	1.01 (0.97-1.05)	5.79E-01	6.73E-01	NA
182	XXL_VLDL_CE_pct	0.029	0.020	1.03 (0.99-1.07)	1.49E-01	2.43E-01	0.033	0.020	1.03 (0.99-1.07)	1.01E-01	2.19E-01	NA
183	XXL_VLDL_FC_pct	-0.032	0.020	0.97 (0.93-1.01)	1.05E-01	1.87E-01	-0.028	0.020	0.97 (0.94-1.01)	1.59E-01	2.88E-01	NA
184	XXL_VLDL_TG_pct	-0.010	0.018	0.99 (0.96-1.03)	5.74E-01	6.62E-01	-0.011	0.018	0.99 (0.96-1.02)	5.47E-01	6.61E-01	NA
185	XL_VLDL_PL_pct	0.034	0.022	1.04 (0.99-1.08)	1.17E-01	2.02E-01	0.031	0.022	1.03 (0.99-1.08)	1.57E-01	2.88E-01	NA
186	XL_VLDL_C_pct	-0.023	0.020	0.98 (0.94-1.02)	2.47E-01	3.62E-01	-0.011	0.020	0.99 (0.95-1.03)	5.64E-01	6.65E-01	NA
187	XL_VLDL_CE_pct	-0.021	0.020	0.98 (0.94-1.02)	2.76E-01	3.86E-01	-0.010	0.020	0.99 (0.95-1.03)	6.23E-01	7.08E-01	NA
188	XL_VLDL_FC_pct	-0.023	0.020	0.98 (0.94-1.02)	2.59E-01	3.73E-01	-0.014	0.020	0.99 (0.95-1.03)	4.82E-01	5.97E-01	NA
189	XL_VLDL_TG_pct	0.011	0.020	1.01 (0.97-1.05)	5.83E-01	6.66E-01	0.001	0.020	1 (0.96-1.04)	9.54E-01	9.63E-01	NA
190	L_VLDL_PL_pct	0.072	0.022	1.08 (1.03-1.12)	1.24E-03	4.97E-03	0.059	0.022	1.06 (1.02-1.11)	8.26E-03	3.17E-02	NA

		Model 1					Model 2					
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***
191	L_VLDL_C_pct	0.002	0.019	1 (0.97-1.04)	9.20E-01	9.54E-01	0.007	0.019	1.01 (0.97-1.05)	7.06E-01	7.70E-01	NA
192	L_VLDL_CE_pct	-0.018	0.019	0.98 (0.95-1.02)	3.44E-01	4.50E-01	-0.010	0.019	0.99 (0.95-1.03)	5.99E-01	6.87E-01	NA
193	L_VLDL_FC_pct	0.059	0.020	1.06 (1.02-1.1)	2.83E-03	1.02E-02	0.055	0.020	1.06 (1.02-1.1)	5.48E-03	2.33E-02	NA
194	L_VLDL_TG_pct	-0.037	0.019	0.96 (0.93-1)	5.28E-02	1.14E-01	-0.035	0.019	0.97 (0.93-1)	6.75E-02	1.73E-01	NA
195	M_VLDL_PL_pct	0.015	0.019	1.02 (0.98-1.06)	4.27E-01	5.34E-01	0.020	0.019	1.02 (0.98-1.06)	2.96E-01	4.44E-01	NA
196	M_VLDL_C_pct	-0.010	0.019	0.99 (0.95-1.03)	5.86E-01	6.66E-01	-0.001	0.019	1 (0.96-1.04)	9.66E-01	9.67E-01	NA
197	M_VLDL_CE_pct	-0.008	0.019	0.99 (0.96-1.03)	6.61E-01	7.28E-01	0.001	0.019	1 (0.97-1.04)	9.55E-01	9.63E-01	NA
198	M_VLDL_FC_pct	-0.001	0.019	1 (0.96-1.04)	9.78E-01	9.82E-01	0.008	0.019	1.01 (0.97-1.05)	6.91E-01	7.65E-01	NA
199	M_VLDL_TG_pct	0.011	0.020	1.01 (0.97-1.05)	5.74E-01	6.62E-01	0.002	0.020	1 (0.96-1.04)	9.29E-01	9.51E-01	NA
200	S_VLDL_PL_pct	-0.018	0.020	0.98 (0.95-1.02)	3.56E-01	4.62E-01	-0.008	0.020	0.99 (0.96-1.03)	7.02E-01	7.70E-01	NA
201	S_VLDL_C_pct	0.014	0.019	1.02 (0.98-1.05)	4.53E-01	5.61E-01	0.022	0.019	1.02 (0.98-1.06)	2.58E-01	4.12E-01	NA
202	S_VLDL_CE_pct	0.037	0.019	1.04 (1-1.08)	5.12E-02	1.12E-01	0.041	0.019	1.04 (1-1.08)	3.06E-02	9.63E-02	NA
203	S_VLDL_FC_pct	-0.018	0.019	0.98 (0.95-1.02)	3.44E-01	4.50E-01	-0.008	0.019	0.99 (0.96-1.03)	6.94E-01	7.65E-01	NA
204	S_VLDL_TG_pct	-0.001	0.020	1 (0.96-1.04)	9.58E-01	9.74E-01	-0.010	0.020	0.99 (0.95-1.03)	6.27E-01	7.10E-01	NA
205	XS_VLDL_PL_pct	0.090	0.020	1.09 (1.05-1.14)	4.39E-06	6.35E-05	0.076	0.020	1.08 (1.04-1.12)	1.10E-04	1.38E-03	-16%
206	XS_VLDL_C_pct	-0.028	0.019	0.97 (0.94-1.01)	1.37E-01	2.28E-01	-0.016	0.019	0.98 (0.95-1.02)	4.11E-01	5.49E-01	NA
207	XS_VLDL_CE_pct	-0.033	0.019	0.97 (0.93-1)	8.19E-02	1.56E-01	-0.021	0.019	0.98 (0.94-1.02)	2.85E-01	4.40E-01	NA
208	XS_VLDL_FC_pct	0.015	0.019	1.02 (0.98-1.05)	4.31E-01	5.37E-01	0.022	0.019	1.02 (0.99-1.06)	2.35E-01	3.85E-01	NA
209	XS_VLDL_TG_pct	0.020	0.020	1.02 (0.98-1.06)	3.25E-01	4.32E-01	0.007	0.020	1.01 (0.97-1.05)	7.15E-01	7.74E-01	NA
210	IDL_PL_pct	0.023	0.019	1.02 (0.99-1.06)	2.19E-01	3.28E-01	0.019	0.019	1.02 (0.98-1.06)	3.17E-01	4.59E-01	NA
211	IDL_C_pct	-0.052	0.018	0.95 (0.92-0.98)	4.35E-03	1.48E-02	-0.040	0.018	0.96 (0.93-1)	3.16E-02	9.78E-02	NA
212	IDL_CE_pct	-0.057	0.018	0.94 (0.91-0.98)	1.87E-03	7.26E-03	-0.047	0.018	0.95 (0.92-0.99)	1.03E-02	3.82E-02	NA
213	IDL_FC_pct	-0.012	0.018	0.99 (0.95-1.02)	5.09E-01	6.09E-01	-0.001	0.018	1 (0.96-1.04)	9.67E-01	9.67E-01	NA
214	IDL_TG_pct	0.060	0.019	1.06 (1.02-1.1)	1.84E-03	7.26E-03	0.045	0.019	1.05 (1.01-1.09)	2.10E-02	7.08E-02	NA
215	L_LDL_PL_pct	0.016	0.019	1.02 (0.98-1.05)	3.93E-01	4.99E-01	0.019	0.019	1.02 (0.98-1.06)	3.06E-01	4.52E-01	NA
216	L_LDL_C_pct	-0.086	0.019	0.92 (0.89-0.95)	3.87E-06	6.35E-05	-0.071	0.019	0.93 (0.9-0.97)	1.41E-04	1.67E-03	17%
217	L_LDL_CE_pct	-0.066	0.018	0.94 (0.9-0.97)	3.29E-04	1.70E-03	-0.059	0.018	0.94 (0.91-0.98)	1.26E-03	6.54E-03	NA
218	L_LDL_FC_pct	-0.041	0.019	0.96 (0.93-1)	2.75E-02	7.28E-02	-0.029	0.019	0.97 (0.94-1.01)	1.26E-01	2.47E-01	NA
219	L_LDL_TG_pct	0.079	0.019	1.08 (1.04-1.12)	3.43E-05	2.75E-04	0.062	0.019	1.06 (1.02-1.1)	1.26E-03	6.54E-03	-22%
220	M_LDL_PL_pct	0.018	0.019	1.02 (0.98-1.06)	3.45E-01	4.50E-01	0.020	0.019	1.02 (0.98-1.06)	2.99E-01	4.45E-01	NA
221	M_LDL_C_pct	-0.060	0.018	0.94 (0.91-0.98)	9.12E-04	3.85E-03	-0.049	0.018	0.95 (0.92-0.99)	7.75E-03	3.01E-02	NA
222	M_LDL_CE_pct	-0.006	0.019	0.99 (0.96-1.03)	7.44E-01	7.98E-01	-0.010	0.019	0.99 (0.95-1.03)	5.95E-01	6.86E-01	NA

		Model 1					Model 2					
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***
223	M_LDL_FC_pct	-0.043	0.019	0.96 (0.92-0.99)	2.03E-02	5.94E-02	-0.030	0.019	0.97 (0.94-1.01)	1.14E-01	2.31E-01	NA
224	M_LDL_TG_pct	0.057	0.019	1.06 (1.02-1.1)	2.81E-03	1.02E-02	0.042	0.019	1.04 (1.01-1.08)	2.79E-02	8.91E-02	NA
225	S_LDL_PL_pct	0.024	0.020	1.02 (0.99-1.06)	2.24E-01	3.34E-01	0.021	0.020	1.02 (0.98-1.06)	2.84E-01	4.40E-01	NA
226	S_LDL_C_pct	-0.051	0.019	0.95 (0.92-0.99)	8.23E-03	2.63E-02	-0.037	0.019	0.96 (0.93-1)	5.49E-02	1.48E-01	NA
227	S_LDL_CE_pct	-0.014	0.019	0.99 (0.95-1.02)	4.61E-01	5.68E-01	-0.014	0.019	0.99 (0.95-1.02)	4.67E-01	5.87E-01	NA
228	S_LDL_FC_pct	-0.030	0.018	0.97 (0.94-1.01)	9.87E-02	1.78E-01	-0.018	0.018	0.98 (0.95-1.02)	3.23E-01	4.60E-01	NA
229	S_LDL_TG_pct	0.033	0.019	1.03 (1-1.07)	8.63E-02	1.62E-01	0.020	0.019	1.02 (0.98-1.06)	2.89E-01	4.41E-01	NA
230	XL_HDL_PL_pct	-0.021	0.019	0.98 (0.94-1.02)	2.65E-01	3.76E-01	-0.022	0.019	0.98 (0.94-1.02)	2.58E-01	4.12E-01	NA
231	XL_HDL_C_pct	0.013	0.019	1.01 (0.98-1.05)	5.11E-01	6.09E-01	0.017	0.019	1.02 (0.98-1.06)	3.72E-01	5.14E-01	NA
232	XL_HDL_CE_pct	-0.056	0.018	0.95 (0.91-0.98)	2.13E-03	8.15E-03	-0.045	0.018	0.96 (0.92-0.99)	1.40E-02	5.07E-02	NA
233	XL_HDL_FC_pct	0.085	0.020	1.09 (1.05-1.13)	2.36E-05	2.18E-04	0.079	0.020	1.08 (1.04-1.13)	8.78E-05	1.22E-03	-7%
234	XL_HDL_TG_pct	0.044	0.019	1.05 (1.01-1.09)	2.44E-02	6.84E-02	0.034	0.019	1.04 (1-1.08)	8.05E-02	1.91E-01	NA
235	L_HDL_PL_pct	0.065	0.018	1.07 (1.03-1.11)	4.34E-04	2.08E-03	0.055	0.019	1.06 (1.02-1.1)	2.90E-03	1.29E-02	NA
236	L_HDL_C_pct	-0.065	0.018	0.94 (0.91-0.97)	2.39E-04	1.35E-03	-0.055	0.018	0.95 (0.91-0.98)	2.10E-03	1.02E-02	NA
237	L_HDL_CE_pct	-0.073	0.017	0.93 (0.9-0.96)	2.55E-05	2.27E-04	-0.061	0.017	0.94 (0.91-0.97)	4.19E-04	3.26E-03	16%
238	L_HDL_FC_pct	0.013	0.019	1.01 (0.98-1.05)	5.08E-01	6.09E-01	0.012	0.019	1.01 (0.97-1.05)	5.49E-01	6.61E-01	NA
239	L_HDL_TG_pct	0.065	0.019	1.07 (1.03-1.11)	7.73E-04	3.32E-03	0.054	0.019	1.06 (1.02-1.1)	5.51E-03	2.33E-02	NA
240	M_HDL_PL_pct	0.094	0.019	1.1 (1.06-1.14)	6.47E-07	1.52E-05	0.079	0.019	1.08 (1.04-1.12)	3.21E-05	6.33E-04	-16%
241	M_HDL_C_pct	-0.083	0.019	0.92 (0.89-0.96)	8.90E-06	1.17E-04	-0.068	0.019	0.93 (0.9-0.97)	2.62E-04	2.25E-03	18%
242	M_HDL_CE_pct	-0.076	0.018	0.93 (0.89-0.96)	3.25E-05	2.70E-04	-0.062	0.018	0.94 (0.91-0.97)	7.34E-04	4.67E-03	19%
243	M_HDL_FC_pct	-0.071	0.020	0.93 (0.9-0.97)	4.07E-04	1.99E-03	-0.062	0.020	0.94 (0.9-0.98)	2.16E-03	1.02E-02	NA
244	M_HDL_TG_pct	0.074	0.020	1.08 (1.04-1.12)	1.88E-04	1.11E-03	0.061	0.020	1.06 (1.02-1.11)	2.18E-03	1.02E-02	NA
245	S_HDL_PL_pct	-0.006	0.020	0.99 (0.96-1.03)	7.59E-01	8.04E-01	-0.012	0.020	0.99 (0.95-1.03)	5.56E-01	6.61E-01	NA
246	S_HDL_C_pct	-0.048	0.019	0.95 (0.92-0.99)	1.10E-02	3.44E-02	-0.035	0.019	0.97 (0.93-1)	6.58E-02	1.71E-01	NA
247	S_HDL_CE_pct	-0.050	0.019	0.95 (0.92-0.99)	7.17E-03	2.38E-02	-0.038	0.019	0.96 (0.93-1)	4.33E-02	1.27E-01	NA
248	S_HDL_FC_pct	0.008	0.019	1.01 (0.97-1.05)	6.91E-01	7.51E-01	0.011	0.020	1.01 (0.97-1.05)	5.80E-01	6.73E-01	NA
249	S_HDL_TG_pct	0.079	0.020	1.08 (1.04-1.13)	7.21E-05	4.72E-04	0.065	0.020	1.07 (1.03-1.11)	1.19E-03	6.54E-03	-18%

* Percentage change in beta from previous model.

SE, standard error. HR, hazard ratio. CI, confidence intervals. FDR, false discovery rate. NA, not applicable.

Abbreviations are in line with **Supplementary Table 3.1**.

Supplementary Table 4.2 Association of individual metabolites with incident ischaemic stroke in Cox proportional hazards regression in model 3 and 4 (n = 184,564)

		Model 3						Model 4					
ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR	% change*	Beta	SE	HR (95% CI)	p value	FDR	% change*
1	Total_C	0.002	0.020	1 (0.96-1.04)	9.17E-01	9.47E-01	NA	-0.002	0.020	1 (0.96-1.04)	9.02E-01	9.38E-01	NA
2	non_HDL_C	0.022	0.020	1.02 (0.98-1.06)	2.60E-01	4.41E-01	NA	0.015	0.020	1.02 (0.98-1.06)	4.37E-01	7.21E-01	NA
3	Remnant_C	0.028	0.020	1.03 (0.99-1.07)	1.50E-01	3.44E-01	NA	0.021	0.020	1.02 (0.98-1.06)	2.79E-01	5.87E-01	NA
4	VLDL_C	0.028	0.020	1.03 (0.99-1.07)	1.51E-01	3.44E-01	NA	0.019	0.020	1.02 (0.98-1.06)	3.38E-01	6.52E-01	NA
5	Clinical_LDL_C	0.019	0.019	1.02 (0.98-1.06)	3.25E-01	4.94E-01	NA	0.014	0.020	1.01 (0.98-1.05)	4.89E-01	7.56E-01	NA
6	LDL_C	0.015	0.019	1.02 (0.98-1.06)	4.34E-01	5.78E-01	NA	0.009	0.020	1.01 (0.97-1.05)	6.49E-01	8.25E-01	NA
7	HDL_C	-0.065	0.021	0.94 (0.9-0.98)	2.57E-03	2.57E-02	25%	-0.058	0.022	0.94 (0.9-0.98)	6.84E-03	6.81E-02	NA
8	Total_TG	0.023	0.019	1.02 (0.99-1.06)	2.28E-01	4.28E-01	NA	0.009	0.019	1.01 (0.97-1.05)	6.32E-01	8.25E-01	NA
9	VLDL_TG	0.016	0.019	1.02 (0.98-1.06)	4.07E-01	5.66E-01	NA	0.002	0.020	1 (0.96-1.04)	9.07E-01	9.38E-01	NA
10	LDL_TG	0.060	0.019	1.06 (1.02-1.1)	1.74E-03	2.10E-02	NA	0.047	0.019	1.05 (1.01-1.09)	1.45E-02	1.02E-01	NA
11	HDL_TG	0.022	0.019	1.02 (0.98-1.06)	2.65E-01	4.45E-01	NA	0.009	0.019	1.01 (0.97-1.05)	6.49E-01	8.25E-01	NA
12	Total_PL	-0.003	0.020	1 (0.96-1.04)	8.66E-01	9.09E-01	NA	-0.011	0.020	0.99 (0.95-1.03)	5.85E-01	8.18E-01	NA
13	VLDL_PL	0.028	0.020	1.03 (0.99-1.07)	1.56E-01	3.53E-01	NA	0.016	0.020	1.02 (0.98-1.06)	4.22E-01	7.07E-01	NA
14	LDL_PL	0.025	0.019	1.03 (0.99-1.07)	1.98E-01	4.09E-01	NA	0.018	0.020	1.02 (0.98-1.06)	3.67E-01	6.80E-01	NA
15	HDL_PL	-0.048	0.022	0.95 (0.91-0.99)	2.56E-02	9.96E-02	NA	-0.048	0.022	0.95 (0.91-1)	2.75E-02	1.40E-01	NA
16	Total_CE	0.000	0.020	1 (0.96-1.04)	9.97E-01	9.97E-01	NA	-0.004	0.020	1 (0.96-1.04)	8.25E-01	9.09E-01	NA
17	VLDL_CE	0.030	0.020	1.03 (0.99-1.07)	1.27E-01	3.16E-01	NA	0.022	0.020	1.02 (0.98-1.06)	2.60E-01	5.69E-01	NA
18	LDL_CE	0.016	0.019	1.02 (0.98-1.06)	4.19E-01	5.72E-01	NA	0.008	0.020	1.01 (0.97-1.05)	6.64E-01	8.35E-01	NA
19	HDL_CE	-0.069	0.021	0.93 (0.9-0.97)	1.18E-03	1.63E-02	23%	-0.063	0.021	0.94 (0.9-0.98)	3.43E-03	6.35E-02	NA
20	Total_FC	0.007	0.020	1.01 (0.97-1.05)	7.13E-01	7.89E-01	NA	0.002	0.020	1 (0.96-1.04)	9.04E-01	9.38E-01	NA
21	VLDL_FC	0.025	0.020	1.03 (0.99-1.07)	1.96E-01	4.09E-01	NA	0.014	0.020	1.01 (0.98-1.05)	4.65E-01	7.47E-01	NA
22	LDL_FC	0.013	0.019	1.01 (0.98-1.05)	4.97E-01	6.25E-01	NA	0.010	0.019	1.01 (0.97-1.05)	6.24E-01	8.25E-01	NA
23	HDL_FC	-0.044	0.022	0.96 (0.92-1)	4.14E-02	1.43E-01	NA	-0.039	0.022	0.96 (0.92-1)	7.76E-02	2.61E-01	NA
24	Total_L	0.007	0.020	1.01 (0.97-1.05)	7.06E-01	7.88E-01	NA	-0.002	0.020	1 (0.96-1.04)	9.17E-01	9.44E-01	NA
25	VLDL_L	0.022	0.019	1.02 (0.98-1.06)	2.49E-01	4.33E-01	NA	0.010	0.020	1.01 (0.97-1.05)	6.23E-01	8.25E-01	NA
26	LDL_L	0.021	0.019	1.02 (0.98-1.06)	2.75E-01	4.51E-01	NA	0.014	0.020	1.01 (0.98-1.05)	4.76E-01	7.56E-01	NA
27	HDL_L	-0.055	0.022	0.95 (0.91-0.99)	1.06E-02	5.26E-02	NA	-0.053	0.022	0.95 (0.91-0.99)	1.47E-02	1.02E-01	NA
28	Total_P	-0.044	0.021	0.96 (0.92-1)	3.54E-02	1.28E-01	NA	-0.051	0.021	0.95 (0.91-0.99)	1.49E-02	1.02E-01	NA
29	VLDL_P	0.030	0.019	1.03 (0.99-1.07)	1.19E-01	3.07E-01	NA	0.021	0.020	1.02 (0.98-1.06)	2.91E-01	5.93E-01	NA
30	LDL_P	0.025	0.019	1.03 (0.99-1.07)	2.04E-01	4.17E-01	NA	0.019	0.019	1.02 (0.98-1.06)	3.23E-01	6.38E-01	NA
31	HDL_P	-0.051	0.021	0.95 (0.91-0.99)	1.37E-02	5.99E-02	NA	-0.057	0.021	0.94 (0.91-0.98)	5.95E-03	6.47E-02	NA
32	VLDL_size	-0.007	0.020	0.99 (0.96-1.03)	7.13E-01	7.89E-01	NA	-0.023	0.020	0.98 (0.94-1.02)	2.49E-01	5.53E-01	NA
33	LDL_size	-0.012	0.019	0.99 (0.95-1.03)	5.16E-01	6.40E-01	NA	-0.005	0.019	1 (0.96-1.03)	8.07E-01	9.05E-01	NA
34	HDL_size	-0.054	0.022	0.95 (0.91-0.99)	1.21E-02	5.60E-02	NA	-0.037	0.022	0.96 (0.92-1.01)	9.15E-02	2.96E-01	NA
35	Phosphoglyc	-0.010	0.020	0.99 (0.95-1.03)	6.38E-01	7.28E-01	NA	-0.017	0.021	0.98 (0.95-1.02)	4.16E-01	7.04E-01	NA
36	TG_by_PG	0.032	0.020	1.03 (0.99-1.07)	1.17E-01	3.06E-01	NA	0.018	0.020	1.02 (0.98-1.06)	3.67E-01	6.80E-01	NA
37	Cholines	-0.015	0.021	0.99 (0.95-1.03)	4.63E-01	6.01E-01	NA	-0.021	0.021	0.98 (0.94-1.02)	3.10E-01	6.27E-01	NA
38	Phosphatidylc	-0.019	0.021	0.98 (0.94-1.02)	3.48E-01	5.17E-01	NA	-0.024	0.021	0.98 (0.94-1.02)	2.44E-01	5.46E-01	NA
39	Sphingomyelins	0.026	0.021	1.03 (0.99-1.07)	2.12E-01	4.24E-01	NA	0.022	0.021	1.02 (0.98-1.07)	2.88E-01	5.93E-01	NA
40	ApoB	0.029	0.019	1.03 (0.99-1.07)	1.36E-01	3.29E-01	NA	0.023	0.019	1.02 (0.99-1.06)	2.44E-01	5.46E-01	NA

		Model 3						Model 4					
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***
41	ApoA1	-0.054	0.021	0.95 (0.91-0.99)	1.20E-02	5.60E-02	NA	-0.056	0.021	0.95 (0.91-0.99)	9.30E-03	8.58E-02	NA
42	ApoB_by_ApoA1	0.053	0.020	1.05 (1.01-1.1)	8.03E-03	4.45E-02	NA	0.048	0.020	1.05 (1.01-1.09)	1.55E-02	1.02E-01	NA
43	Total_FA	0.020	0.019	1.02 (0.98-1.06)	2.86E-01	4.63E-01	NA	0.008	0.019	1.01 (0.97-1.05)	6.97E-01	8.55E-01	NA
44	Unsaturation	-0.087	0.019	0.92 (0.88-0.95)	4.74E-06	2.36E-04	9%	-0.063	0.020	0.94 (0.9-0.98)	1.66E-03	5.64E-02	27%
45	Omega_3	-0.054	0.019	0.95 (0.91-0.98)	5.54E-03	3.60E-02	NA	-0.039	0.021	0.96 (0.92-1)	5.93E-02	2.25E-01	NA
46	Omega_6	-0.020	0.019	0.98 (0.94-1.02)	2.98E-01	4.72E-01	NA	-0.026	0.020	0.97 (0.94-1.01)	1.82E-01	4.28E-01	NA
47	PUFA	-0.029	0.019	0.97 (0.94-1.01)	1.39E-01	3.29E-01	NA	-0.029	0.020	0.97 (0.94-1.01)	1.43E-01	3.72E-01	NA
48	MUFA	0.048	0.019	1.05 (1.01-1.09)	1.19E-02	5.60E-02	NA	0.029	0.019	1.03 (0.99-1.07)	1.36E-01	3.70E-01	NA
49	SFA	0.035	0.019	1.04 (1-1.08)	6.47E-02	1.89E-01	NA	0.019	0.019	1.02 (0.98-1.06)	3.20E-01	6.38E-01	NA
50	LA	-0.033	0.019	0.97 (0.93-1.01)	8.83E-02	2.53E-01	NA	-0.036	0.019	0.96 (0.93-1)	6.16E-02	2.28E-01	NA
51	DHA	-0.054	0.019	0.95 (0.91-0.98)	4.79E-03	3.41E-02	NA	-0.032	0.021	0.97 (0.93-1.01)	1.26E-01	3.54E-01	NA
52	Omega_3_pct	-0.077	0.020	0.93 (0.89-0.96)	8.04E-05	2.50E-03	14%	-0.057	0.022	0.95 (0.91-0.99)	9.09E-03	8.58E-02	NA
53	Omega_6_pct	-0.067	0.018	0.94 (0.9-0.97)	2.39E-04	6.60E-03	NA	-0.051	0.018	0.95 (0.92-0.99)	5.29E-03	6.47E-02	NA
54	PUFA_pct	-0.087	0.018	0.92 (0.89-0.95)	1.65E-06	1.03E-04	1%	-0.064	0.019	0.94 (0.91-0.97)	6.88E-04	3.62E-02	27%
55	MUFA_pct	0.086	0.019	1.09 (1.05-1.13)	8.57E-06	3.05E-04	-10%	0.062	0.020	1.06 (1.02-1.11)	1.81E-03	5.64E-02	-28%
56	SFA_pct	0.064	0.019	1.07 (1.03-1.11)	6.08E-04	1.16E-02	NA	0.048	0.019	1.05 (1.01-1.09)	1.27E-02	9.90E-02	NA
57	LA_pct	-0.084	0.019	0.92 (0.89-0.95)	6.39E-06	2.65E-04	-9%	-0.071	0.019	0.93 (0.9-0.97)	2.00E-04	1.66E-02	16%
58	DHA_pct	-0.064	0.019	0.94 (0.9-0.97)	6.05E-04	1.16E-02	18%	-0.037	0.021	0.96 (0.93-1)	7.35E-02	2.51E-01	NA
59	PUFA_by_MUFA	-0.090	0.019	0.91 (0.88-0.95)	1.64E-06	1.03E-04	6%	-0.065	0.019	0.94 (0.9-0.97)	7.28E-04	3.62E-02	27%
60	Omega_6_by_Omega_3	0.052	0.020	1.05 (1.01-1.1)	7.17E-03	4.19E-02	NA	0.034	0.021	1.03 (0.99-1.08)	1.15E-01	3.44E-01	NA
61	Ala	0.044	0.023	1.05 (1-1.09)	5.99E-02	1.78E-01	NA	0.045	0.023	1.05 (1-1.1)	5.41E-02	2.14E-01	NA
62	Gln	-0.039	0.019	0.96 (0.93-1)	4.24E-02	1.45E-01	NA	-0.035	0.019	0.97 (0.93-1)	6.76E-02	2.40E-01	NA
63	Gly	-0.014	0.016	0.99 (0.96-1.02)	3.92E-01	5.61E-01	NA	-0.014	0.017	0.99 (0.96-1.02)	3.99E-01	7.04E-01	NA
64	His	-0.054	0.019	0.95 (0.91-0.98)	3.83E-03	2.89E-02	NA	-0.052	0.019	0.95 (0.92-0.99)	5.34E-03	6.47E-02	NA
65	Total_BCAA	0.003	0.020	1 (0.96-1.04)	8.98E-01	9.36E-01	NA	0.000	0.020	1 (0.96-1.04)	9.97E-01	9.97E-01	NA
66	Ile	0.011	0.019	1.01 (0.97-1.05)	5.80E-01	6.75E-01	NA	0.007	0.019	1.01 (0.97-1.05)	7.24E-01	8.73E-01	NA
67	Leu	-0.006	0.020	0.99 (0.96-1.03)	7.73E-01	8.29E-01	NA	-0.011	0.020	0.99 (0.95-1.03)	5.80E-01	8.18E-01	NA
68	Val	0.004	0.020	1 (0.97-1.04)	8.37E-01	8.87E-01	NA	0.004	0.020	1 (0.97-1.04)	8.34E-01	9.09E-01	NA
69	Phe	0.065	0.019	1.07 (1.03-1.11)	5.58E-04	1.16E-02	NA	0.063	0.019	1.07 (1.03-1.11)	8.72E-04	3.62E-02	NA
70	Tyr	0.050	0.019	1.05 (1.01-1.09)	9.14E-03	4.74E-02	NA	0.053	0.019	1.05 (1.02-1.1)	5.97E-03	6.47E-02	NA
71	Glucose	0.047	0.019	1.05 (1.01-1.09)	1.35E-02	5.99E-02	NA	0.045	0.019	1.05 (1.01-1.09)	1.72E-02	1.10E-01	NA
72	Lactate	0.067	0.020	1.07 (1.03-1.11)	7.18E-04	1.19E-02	NA	0.059	0.020	1.06 (1.02-1.1)	2.87E-03	6.35E-02	NA
73	Pyruvate	0.038	0.020	1.04 (1-1.08)	5.35E-02	1.67E-01	NA	0.032	0.020	1.03 (0.99-1.07)	1.04E-01	3.21E-01	NA
74	Citrate	-0.006	0.020	0.99 (0.96-1.03)	7.52E-01	8.17E-01	NA	-0.007	0.020	0.99 (0.95-1.03)	7.16E-01	8.73E-01	NA
75	bOHbutyrate	0.019	0.021	1.02 (0.98-1.06)	3.63E-01	5.28E-01	NA	0.012	0.021	1.01 (0.97-1.06)	5.64E-01	8.18E-01	NA
76	Acetate	-0.023	0.016	0.98 (0.95-1.01)	1.45E-01	3.38E-01	NA	-0.018	0.017	0.98 (0.95-1.01)	2.73E-01	5.87E-01	NA
77	Acetoacetate	0.000	0.020	1 (0.96-1.04)	9.97E-01	9.97E-01	NA	-0.004	0.020	1 (0.96-1.04)	8.43E-01	9.09E-01	NA
78	Acetone	-0.003	0.019	1 (0.96-1.03)	8.61E-01	9.09E-01	NA	-0.006	0.019	0.99 (0.96-1.03)	7.60E-01	8.77E-01	NA
79	Creatinine	0.059	0.022	1.06 (1.02-1.11)	7.71E-03	4.36E-02	NA	0.045	0.022	1.05 (1-1.09)	4.20E-02	1.83E-01	NA
80	Albumin	-0.142	0.019	0.87 (0.84-0.9)	2.10E-14	5.23E-12	5%	-0.135	0.019	0.87 (0.84-0.91)	4.29E-13	1.07E-10	5%
81	GlycA	0.116	0.020	1.12 (1.08-1.17)	4.26E-09	5.31E-07	-8%	0.102	0.020	1.11 (1.06-1.15)	3.81E-07	4.74E-05	-13%
82	XXL_VLDL_P	0.021	0.020	1.02 (0.98-1.06)	3.03E-01	4.72E-01	NA	0.007	0.021	1.01 (0.97-1.05)	7.40E-01	8.76E-01	NA
83	XXL_VLDL_L	0.014	0.020	1.01 (0.98-1.05)	4.84E-01	6.19E-01	NA	-0.001	0.020	1 (0.96-1.04)	9.80E-01	9.84E-01	NA

		Model 3						Model 4					
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***
84	XXL_VLDL_PL	0.018	0.021	1.02 (0.98-1.06)	3.91E-01	5.61E-01	NA	0.005	0.021	1.01 (0.96-1.05)	8.22E-01	9.09E-01	NA
85	XXL_VLDL_C	0.024	0.020	1.02 (0.98-1.07)	2.43E-01	4.33E-01	NA	0.008	0.020	1.01 (0.97-1.05)	6.86E-01	8.50E-01	NA
86	XXL_VLDL_CE	0.033	0.020	1.03 (0.99-1.08)	1.05E-01	2.84E-01	NA	0.017	0.021	1.02 (0.98-1.06)	4.01E-01	7.04E-01	NA
87	XXL_VLDL_FC	0.010	0.020	1.01 (0.97-1.05)	6.15E-01	7.09E-01	NA	-0.004	0.020	1 (0.96-1.04)	8.36E-01	9.09E-01	NA
88	XXL_VLDL_TG	0.000	0.019	1 (0.96-1.04)	9.89E-01	9.97E-01	NA	-0.010	0.019	0.99 (0.95-1.03)	6.15E-01	8.25E-01	NA
89	XL_VLDL_P	0.027	0.020	1.03 (0.99-1.07)	1.78E-01	3.83E-01	NA	0.012	0.020	1.01 (0.97-1.05)	5.68E-01	8.18E-01	NA
90	XL_VLDL_L	0.025	0.020	1.03 (0.99-1.07)	2.14E-01	4.24E-01	NA	0.010	0.020	1.01 (0.97-1.05)	6.32E-01	8.25E-01	NA
91	XL_VLDL_PL	0.026	0.020	1.03 (0.99-1.07)	1.99E-01	4.09E-01	NA	0.012	0.021	1.01 (0.97-1.05)	5.72E-01	8.18E-01	NA
92	XL_VLDL_C	0.028	0.020	1.03 (0.99-1.07)	1.71E-01	3.76E-01	NA	0.014	0.020	1.01 (0.97-1.06)	5.01E-01	7.61E-01	NA
93	XL_VLDL_CE	0.030	0.020	1.03 (0.99-1.07)	1.36E-01	3.29E-01	NA	0.017	0.020	1.02 (0.98-1.06)	4.14E-01	7.04E-01	NA
94	XL_VLDL_FC	0.024	0.020	1.02 (0.99-1.07)	2.30E-01	4.28E-01	NA	0.010	0.020	1.01 (0.97-1.05)	6.15E-01	8.25E-01	NA
95	XL_VLDL_TG	0.022	0.020	1.02 (0.98-1.06)	2.71E-01	4.47E-01	NA	0.007	0.020	1.01 (0.97-1.05)	7.42E-01	8.76E-01	NA
96	L_VLDL_P	0.021	0.020	1.02 (0.98-1.06)	2.94E-01	4.70E-01	NA	0.007	0.020	1.01 (0.97-1.05)	7.35E-01	8.76E-01	NA
97	L_VLDL_L	0.017	0.020	1.02 (0.98-1.06)	4.01E-01	5.66E-01	NA	0.003	0.020	1 (0.97-1.04)	8.82E-01	9.34E-01	NA
98	L_VLDL_PL	0.024	0.020	1.02 (0.98-1.07)	2.34E-01	4.31E-01	NA	0.009	0.020	1.01 (0.97-1.05)	6.46E-01	8.25E-01	NA
99	L_VLDL_C	0.019	0.020	1.02 (0.98-1.06)	3.31E-01	4.99E-01	NA	0.007	0.020	1.01 (0.97-1.05)	7.36E-01	8.76E-01	NA
100	L_VLDL_CE	0.015	0.020	1.02 (0.98-1.06)	4.55E-01	5.97E-01	NA	0.004	0.020	1 (0.97-1.04)	8.33E-01	9.09E-01	NA
101	L_VLDL_FC	0.024	0.020	1.02 (0.98-1.07)	2.37E-01	4.31E-01	NA	0.009	0.020	1.01 (0.97-1.05)	6.41E-01	8.25E-01	NA
102	L_VLDL_TG	0.011	0.019	1.01 (0.97-1.05)	5.79E-01	6.75E-01	NA	-0.002	0.020	1 (0.96-1.04)	9.05E-01	9.38E-01	NA
103	M_VLDL_P	0.012	0.019	1.01 (0.98-1.05)	5.25E-01	6.42E-01	NA	0.004	0.020	1 (0.97-1.04)	8.57E-01	9.16E-01	NA
104	M_VLDL_L	0.011	0.019	1.01 (0.97-1.05)	5.75E-01	6.75E-01	NA	0.001	0.019	1 (0.96-1.04)	9.54E-01	9.66E-01	NA
105	M_VLDL_PL	0.015	0.019	1.02 (0.98-1.06)	4.32E-01	5.78E-01	NA	0.007	0.020	1.01 (0.97-1.05)	7.26E-01	8.73E-01	NA
106	M_VLDL_C	0.012	0.019	1.01 (0.98-1.05)	5.26E-01	6.42E-01	NA	0.009	0.019	1.01 (0.97-1.05)	6.26E-01	8.25E-01	NA
107	M_VLDL_CE	0.013	0.019	1.01 (0.98-1.05)	4.91E-01	6.21E-01	NA	0.013	0.019	1.01 (0.98-1.05)	4.86E-01	7.56E-01	NA
108	M_VLDL_FC	0.014	0.019	1.01 (0.98-1.05)	4.63E-01	6.01E-01	NA	0.008	0.020	1.01 (0.97-1.05)	6.84E-01	8.50E-01	NA
109	M_VLDL_TG	0.008	0.019	1.01 (0.97-1.05)	6.66E-01	7.54E-01	NA	-0.004	0.020	1 (0.96-1.04)	8.42E-01	9.09E-01	NA
110	S_VLDL_P	0.023	0.019	1.02 (0.99-1.06)	2.44E-01	4.33E-01	NA	0.013	0.020	1.01 (0.98-1.05)	5.01E-01	7.61E-01	NA
111	S_VLDL_L	0.024	0.020	1.03 (0.99-1.07)	2.15E-01	4.24E-01	NA	0.015	0.020	1.02 (0.98-1.05)	4.59E-01	7.42E-01	NA
112	S_VLDL_PL	0.024	0.020	1.03 (0.99-1.07)	2.12E-01	4.24E-01	NA	0.016	0.020	1.02 (0.98-1.06)	4.05E-01	7.04E-01	NA
113	S_VLDL_C	0.032	0.020	1.03 (0.99-1.07)	1.00E-01	2.74E-01	NA	0.025	0.020	1.03 (0.99-1.07)	2.12E-01	4.93E-01	NA
114	S_VLDL_CE	0.037	0.020	1.04 (1-1.08)	5.97E-02	1.78E-01	NA	0.029	0.020	1.03 (0.99-1.07)	1.47E-01	3.72E-01	NA
115	S_VLDL_FC	0.023	0.019	1.02 (0.99-1.06)	2.36E-01	4.31E-01	NA	0.017	0.019	1.02 (0.98-1.06)	3.87E-01	7.01E-01	NA
116	S_VLDL_TG	0.015	0.020	1.02 (0.98-1.05)	4.56E-01	5.97E-01	NA	0.004	0.020	1 (0.97-1.04)	8.24E-01	9.09E-01	NA
117	XS_VLDL_P	0.054	0.020	1.06 (1.02-1.1)	5.72E-03	3.60E-02	NA	0.049	0.020	1.05 (1.01-1.09)	1.23E-02	9.90E-02	NA
118	XS_VLDL_L	0.053	0.020	1.05 (1.01-1.1)	7.24E-03	4.19E-02	NA	0.048	0.020	1.05 (1.01-1.09)	1.45E-02	1.02E-01	NA
119	XS_VLDL_PL	0.058	0.020	1.06 (1.02-1.1)	2.89E-03	2.57E-02	NA	0.054	0.020	1.06 (1.02-1.1)	6.41E-03	6.65E-02	NA
120	XS_VLDL_C	0.047	0.020	1.05 (1.01-1.09)	1.89E-02	7.72E-02	NA	0.046	0.020	1.05 (1.01-1.09)	2.28E-02	1.23E-01	NA
121	XS_VLDL_CE	0.043	0.020	1.04 (1-1.09)	3.28E-02	1.20E-01	NA	0.042	0.020	1.04 (1-1.09)	3.45E-02	1.62E-01	NA
122	XS_VLDL_FC	0.054	0.020	1.06 (1.02-1.1)	5.93E-03	3.60E-02	NA	0.051	0.020	1.05 (1.01-1.09)	1.01E-02	8.66E-02	NA
123	XS_VLDL_TG	0.043	0.019	1.04 (1.01-1.09)	2.54E-02	9.96E-02	NA	0.034	0.019	1.03 (1-1.07)	8.25E-02	2.74E-01	NA
124	IDL_P	0.038	0.020	1.04 (1-1.08)	4.99E-02	1.64E-01	NA	0.032	0.020	1.03 (0.99-1.07)	1.01E-01	3.16E-01	NA
125	IDL_L	0.031	0.020	1.03 (0.99-1.07)	1.22E-01	3.09E-01	NA	0.027	0.020	1.03 (0.99-1.07)	1.71E-01	4.04E-01	NA
126	IDL_PL	0.033	0.020	1.03 (0.99-1.08)	9.40E-02	2.66E-01	NA	0.032	0.020	1.03 (0.99-1.07)	1.13E-01	3.44E-01	NA

		Model 3						Model 4					
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***
127	IDL_C	0.023	0.020	1.02 (0.98-1.06)	2.54E-01	4.36E-01	NA	0.020	0.020	1.02 (0.98-1.06)	3.15E-01	6.32E-01	NA
128	IDL_CE	0.020	0.020	1.02 (0.98-1.06)	3.04E-01	4.72E-01	NA	0.017	0.020	1.02 (0.98-1.06)	3.91E-01	7.01E-01	NA
129	IDL_FC	0.028	0.020	1.03 (0.99-1.07)	1.59E-01	3.57E-01	NA	0.027	0.020	1.03 (0.99-1.07)	1.69E-01	4.04E-01	NA
130	IDL_TG	0.063	0.019	1.07 (1.03-1.11)	9.71E-04	1.42E-02	NA	0.053	0.019	1.06 (1.02-1.1)	5.72E-03	6.47E-02	NA
131	L_LDL_P	0.021	0.019	1.02 (0.98-1.06)	2.69E-01	4.47E-01	NA	0.018	0.019	1.02 (0.98-1.06)	3.63E-01	6.80E-01	NA
132	L_LDL_L	0.019	0.020	1.02 (0.98-1.06)	3.21E-01	4.91E-01	NA	0.014	0.020	1.01 (0.98-1.05)	4.88E-01	7.56E-01	NA
133	L_LDL_PL	0.023	0.020	1.02 (0.98-1.06)	2.50E-01	4.33E-01	NA	0.016	0.020	1.02 (0.98-1.06)	4.14E-01	7.04E-01	NA
134	L_LDL_C	0.013	0.020	1.01 (0.98-1.05)	5.13E-01	6.38E-01	NA	0.008	0.020	1.01 (0.97-1.05)	6.78E-01	8.48E-01	NA
135	L_LDL_CE	0.013	0.019	1.01 (0.98-1.05)	5.09E-01	6.37E-01	NA	0.008	0.020	1.01 (0.97-1.05)	6.92E-01	8.53E-01	NA
136	L_LDL_FC	0.012	0.020	1.01 (0.97-1.05)	5.39E-01	6.44E-01	NA	0.009	0.020	1.01 (0.97-1.05)	6.49E-01	8.25E-01	NA
137	L_LDL_TG	0.069	0.019	1.07 (1.03-1.11)	3.23E-04	8.04E-03	NA	0.056	0.019	1.06 (1.02-1.1)	3.34E-03	6.35E-02	NA
138	M_LDL_P	0.027	0.020	1.03 (0.99-1.07)	1.69E-01	3.75E-01	NA	0.019	0.020	1.02 (0.98-1.06)	3.28E-01	6.43E-01	NA
139	M_LDL_L	0.023	0.020	1.02 (0.99-1.06)	2.42E-01	4.33E-01	NA	0.013	0.020	1.01 (0.98-1.05)	5.16E-01	7.72E-01	NA
140	M_LDL_PL	0.025	0.020	1.03 (0.99-1.07)	1.94E-01	4.08E-01	NA	0.015	0.020	1.02 (0.98-1.06)	4.37E-01	7.21E-01	NA
141	M_LDL_C	0.019	0.020	1.02 (0.98-1.06)	3.42E-01	5.13E-01	NA	0.009	0.020	1.01 (0.97-1.05)	6.41E-01	8.25E-01	NA
142	M_LDL_CE	0.020	0.020	1.02 (0.98-1.06)	3.11E-01	4.81E-01	NA	0.009	0.020	1.01 (0.97-1.05)	6.41E-01	8.25E-01	NA
143	M_LDL_FC	0.014	0.019	1.01 (0.98-1.05)	4.71E-01	6.06E-01	NA	0.009	0.019	1.01 (0.97-1.05)	6.57E-01	8.30E-01	NA
144	M_LDL_TG	0.050	0.019	1.05 (1.01-1.09)	8.23E-03	4.46E-02	NA	0.037	0.019	1.04 (1-1.08)	5.33E-02	2.14E-01	NA
145	S_LDL_P	0.029	0.019	1.03 (0.99-1.07)	1.43E-01	3.35E-01	NA	0.021	0.020	1.02 (0.98-1.06)	2.74E-01	5.87E-01	NA
146	S_LDL_L	0.026	0.019	1.03 (0.99-1.07)	1.84E-01	3.92E-01	NA	0.018	0.019	1.02 (0.98-1.06)	3.61E-01	6.80E-01	NA
147	S_LDL_PL	0.032	0.019	1.03 (0.99-1.07)	9.92E-02	2.74E-01	NA	0.027	0.019	1.03 (0.99-1.07)	1.63E-01	3.95E-01	NA
148	S_LDL_C	0.021	0.019	1.02 (0.98-1.06)	2.85E-01	4.63E-01	NA	0.013	0.019	1.01 (0.98-1.05)	5.16E-01	7.72E-01	NA
149	S_LDL_CE	0.021	0.019	1.02 (0.98-1.06)	2.70E-01	4.47E-01	NA	0.011	0.020	1.01 (0.97-1.05)	5.58E-01	8.18E-01	NA
150	S_LDL_FC	0.016	0.019	1.02 (0.98-1.06)	3.91E-01	5.61E-01	NA	0.014	0.019	1.01 (0.98-1.05)	4.56E-01	7.42E-01	NA
151	S_LDL_TG	0.030	0.019	1.03 (0.99-1.07)	1.20E-01	3.07E-01	NA	0.017	0.019	1.02 (0.98-1.06)	3.86E-01	7.01E-01	NA
152	XL_HDL_P	-0.028	0.021	0.97 (0.93-1.01)	1.74E-01	3.79E-01	NA	-0.012	0.021	0.99 (0.95-1.03)	5.73E-01	8.18E-01	NA
153	XL_HDL_L	-0.025	0.021	0.98 (0.94-1.02)	2.28E-01	4.28E-01	NA	-0.006	0.021	0.99 (0.95-1.04)	7.60E-01	8.77E-01	NA
154	XL_HDL_PL	-0.024	0.021	0.98 (0.94-1.02)	2.48E-01	4.33E-01	NA	-0.005	0.021	1 (0.96-1.04)	8.04E-01	9.05E-01	NA
155	XL_HDL_C	-0.025	0.021	0.98 (0.94-1.02)	2.29E-01	4.28E-01	NA	-0.006	0.021	0.99 (0.95-1.04)	7.61E-01	8.77E-01	NA
156	XL_HDL_CE	-0.033	0.021	0.97 (0.93-1.01)	1.08E-01	2.89E-01	NA	-0.015	0.021	0.99 (0.95-1.03)	4.78E-01	7.56E-01	NA
157	XL_HDL_FC	0.006	0.020	1.01 (0.97-1.05)	7.62E-01	8.21E-01	NA	0.024	0.021	1.02 (0.98-1.07)	2.52E-01	5.54E-01	NA
158	XL_HDL_TG	0.009	0.019	1.01 (0.97-1.05)	6.23E-01	7.15E-01	NA	0.005	0.019	1.01 (0.97-1.04)	7.77E-01	8.81E-01	NA
159	L_HDL_P	-0.064	0.021	0.94 (0.9-0.98)	2.59E-03	2.57E-02	NA	-0.049	0.021	0.95 (0.91-0.99)	2.08E-02	1.17E-01	NA
160	L_HDL_L	-0.059	0.021	0.94 (0.9-0.98)	5.71E-03	3.60E-02	NA	-0.045	0.022	0.96 (0.92-1)	3.77E-02	1.68E-01	NA
161	L_HDL_PL	-0.054	0.021	0.95 (0.91-0.99)	1.12E-02	5.48E-02	NA	-0.041	0.022	0.96 (0.92-1)	5.56E-02	2.16E-01	NA
162	L_HDL_C	-0.061	0.021	0.94 (0.9-0.98)	3.27E-03	2.81E-02	NA	-0.046	0.021	0.96 (0.92-1)	2.96E-02	1.47E-01	NA
163	L_HDL_CE	-0.065	0.021	0.94 (0.9-0.98)	1.77E-03	2.10E-02	21%	-0.049	0.021	0.95 (0.91-0.99)	1.87E-02	1.14E-01	NA
164	L_HDL_FC	-0.046	0.021	0.96 (0.92-1)	2.76E-02	1.06E-01	NA	-0.032	0.021	0.97 (0.93-1.01)	1.38E-01	3.70E-01	NA
165	L_HDL_TG	-0.008	0.020	0.99 (0.95-1.03)	6.97E-01	7.82E-01	NA	-0.011	0.020	0.99 (0.95-1.03)	5.92E-01	8.19E-01	NA
166	M_HDL_P	-0.058	0.021	0.94 (0.91-0.98)	5.88E-03	3.60E-02	NA	-0.060	0.021	0.94 (0.9-0.98)	4.40E-03	6.35E-02	NA
167	M_HDL_L	-0.051	0.021	0.95 (0.91-0.99)	1.58E-02	6.69E-02	NA	-0.054	0.021	0.95 (0.91-0.99)	9.80E-03	8.66E-02	NA
168	M_HDL_PL	-0.043	0.021	0.96 (0.92-1)	3.85E-02	1.37E-01	NA	-0.048	0.021	0.95 (0.92-0.99)	2.04E-02	1.17E-01	NA
169	M_HDL_C	-0.061	0.021	0.94 (0.9-0.98)	3.66E-03	2.85E-02	26%	-0.061	0.021	0.94 (0.9-0.98)	3.80E-03	6.35E-02	NA

		Model 3						Model 4					
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***
170	M_HDL_CE	-0.063	0.021	0.94 (0.9-0.98)	2.69E-03	2.57E-02	25%	-0.062	0.021	0.94 (0.9-0.98)	2.80E-03	6.35E-02	NA
171	M_HDL_FC	-0.051	0.021	0.95 (0.91-0.99)	1.50E-02	6.42E-02	NA	-0.051	0.021	0.95 (0.91-0.99)	1.53E-02	1.02E-01	NA
172	M_HDL_TG	0.020	0.019	1.02 (0.98-1.06)	3.01E-01	4.72E-01	NA	0.006	0.020	1.01 (0.97-1.05)	7.55E-01	8.77E-01	NA
173	S_HDL_P	-0.020	0.020	0.98 (0.94-1.02)	3.13E-01	4.82E-01	NA	-0.037	0.020	0.96 (0.93-1)	6.24E-02	2.28E-01	NA
174	S_HDL_L	-0.012	0.020	0.99 (0.95-1.03)	5.30E-01	6.43E-01	NA	-0.031	0.020	0.97 (0.93-1.01)	1.18E-01	3.50E-01	NA
175	S_HDL_PL	-0.012	0.020	0.99 (0.95-1.03)	5.32E-01	6.43E-01	NA	-0.031	0.020	0.97 (0.93-1.01)	1.23E-01	3.54E-01	NA
176	S_HDL_C	-0.024	0.020	0.98 (0.94-1.02)	2.20E-01	4.28E-01	NA	-0.040	0.020	0.96 (0.92-1)	4.63E-02	1.91E-01	NA
177	S_HDL_CE	-0.029	0.020	0.97 (0.93-1.01)	1.38E-01	3.29E-01	NA	-0.045	0.020	0.96 (0.92-0.99)	2.45E-02	1.27E-01	NA
178	S_HDL_FC	-0.006	0.020	0.99 (0.96-1.03)	7.55E-01	8.17E-01	NA	-0.020	0.020	0.98 (0.94-1.02)	3.34E-01	6.50E-01	NA
179	S_HDL_TG	0.038	0.020	1.04 (1-1.08)	5.20E-02	1.66E-01	NA	0.021	0.020	1.02 (0.98-1.06)	2.83E-01	5.87E-01	NA
180	XXL_VLDL_PL_pct	0.018	0.023	1.02 (0.97-1.06)	4.32E-01	5.78E-01	NA	0.012	0.023	1.01 (0.97-1.06)	5.83E-01	8.18E-01	NA
181	XXL_VLDL_C_pct	0.016	0.020	1.02 (0.98-1.06)	4.02E-01	5.66E-01	NA	0.023	0.020	1.02 (0.99-1.06)	2.42E-01	5.46E-01	NA
182	XXL_VLDL_CE_pct	0.037	0.020	1.04 (1-1.08)	6.82E-02	1.98E-01	NA	0.040	0.020	1.04 (1-1.08)	4.69E-02	1.91E-01	NA
183	XXL_VLDL_FC_pct	-0.021	0.020	0.98 (0.94-1.02)	2.95E-01	4.70E-01	NA	-0.011	0.020	0.99 (0.95-1.03)	5.83E-01	8.18E-01	NA
184	XXL_VLDL_TG_pct	-0.014	0.018	0.99 (0.95-1.02)	4.17E-01	5.72E-01	NA	-0.016	0.018	0.98 (0.95-1.02)	3.69E-01	6.80E-01	NA
185	XL_VLDL_PL_pct	0.021	0.022	1.02 (0.98-1.07)	3.49E-01	5.17E-01	NA	0.018	0.022	1.02 (0.98-1.06)	4.23E-01	7.07E-01	NA
186	XL_VLDL_C_pct	-0.009	0.020	0.99 (0.95-1.03)	6.60E-01	7.50E-01	NA	0.004	0.020	1 (0.97-1.04)	8.43E-01	9.09E-01	NA
187	XL_VLDL_CE_pct	-0.006	0.020	0.99 (0.96-1.03)	7.51E-01	8.17E-01	NA	0.006	0.020	1.01 (0.97-1.05)	7.78E-01	8.81E-01	NA
188	XL_VLDL_FC_pct	-0.015	0.020	0.99 (0.95-1.03)	4.52E-01	5.97E-01	NA	-0.001	0.020	1 (0.96-1.04)	9.43E-01	9.63E-01	NA
189	XL_VLDL_TG_pct	0.001	0.020	1 (0.96-1.04)	9.78E-01	9.90E-01	NA	-0.010	0.020	0.99 (0.95-1.03)	6.08E-01	8.25E-01	NA
190	L_VLDL_PL_pct	0.048	0.022	1.05 (1-1.1)	3.20E-02	1.19E-01	NA	0.033	0.022	1.03 (0.99-1.08)	1.39E-01	3.70E-01	NA
191	L_VLDL_C_pct	0.005	0.019	1.01 (0.97-1.04)	7.82E-01	8.35E-01	NA	0.013	0.019	1.01 (0.98-1.05)	4.98E-01	7.61E-01	NA
192	L_VLDL_CE_pct	-0.010	0.019	0.99 (0.95-1.03)	6.11E-01	7.08E-01	NA	0.002	0.019	1 (0.96-1.04)	9.23E-01	9.46E-01	NA
193	L_VLDL_FC_pct	0.048	0.020	1.05 (1.01-1.09)	1.61E-02	6.69E-02	NA	0.043	0.020	1.04 (1-1.09)	3.44E-02	1.62E-01	NA
194	L_VLDL_TG_pct	-0.029	0.019	0.97 (0.94-1.01)	1.33E-01	3.27E-01	NA	-0.028	0.019	0.97 (0.94-1.01)	1.40E-01	3.70E-01	NA
195	M_VLDL_PL_pct	0.024	0.019	1.02 (0.99-1.06)	2.24E-01	4.28E-01	NA	0.028	0.020	1.03 (0.99-1.07)	1.48E-01	3.72E-01	NA
196	M_VLDL_C_pct	0.004	0.019	1 (0.97-1.04)	8.30E-01	8.83E-01	NA	0.014	0.019	1.01 (0.98-1.05)	4.59E-01	7.42E-01	NA
197	M_VLDL_CE_pct	0.006	0.019	1.01 (0.97-1.04)	7.45E-01	8.17E-01	NA	0.016	0.019	1.02 (0.98-1.05)	4.05E-01	7.04E-01	NA
198	M_VLDL_FC_pct	0.012	0.019	1.01 (0.97-1.05)	5.41E-01	6.44E-01	NA	0.021	0.019	1.02 (0.98-1.06)	2.81E-01	5.87E-01	NA
199	M_VLDL_TG_pct	-0.003	0.020	1 (0.96-1.04)	8.99E-01	9.36E-01	NA	-0.012	0.020	0.99 (0.95-1.03)	5.32E-01	7.88E-01	NA
200	S_VLDL_PL_pct	-0.001	0.020	1 (0.96-1.04)	9.43E-01	9.67E-01	NA	0.006	0.020	1.01 (0.97-1.05)	7.75E-01	8.81E-01	NA
201	S_VLDL_C_pct	0.023	0.019	1.02 (0.99-1.06)	2.26E-01	4.28E-01	NA	0.028	0.019	1.03 (0.99-1.07)	1.54E-01	3.75E-01	NA
202	S_VLDL_CE_pct	0.039	0.019	1.04 (1-1.08)	4.10E-02	1.43E-01	NA	0.041	0.019	1.04 (1-1.08)	3.35E-02	1.62E-01	NA
203	S_VLDL_FC_pct	-0.001	0.020	1 (0.96-1.04)	9.52E-01	9.72E-01	NA	0.006	0.020	1.01 (0.97-1.05)	7.59E-01	8.77E-01	NA
204	S_VLDL_TG_pct	-0.012	0.020	0.99 (0.95-1.03)	5.34E-01	6.43E-01	NA	-0.017	0.020	0.98 (0.95-1.02)	3.89E-01	7.01E-01	NA
205	XS_VLDL_PL_pct	0.060	0.020	1.06 (1.02-1.1)	2.88E-03	2.57E-02	NA	0.057	0.020	1.06 (1.02-1.1)	4.35E-03	6.35E-02	NA
206	XS_VLDL_C_pct	-0.008	0.019	0.99 (0.96-1.03)	6.91E-01	7.78E-01	NA	0.001	0.019	1 (0.96-1.04)	9.53E-01	9.66E-01	NA
207	XS_VLDL_CE_pct	-0.011	0.019	0.99 (0.95-1.03)	5.54E-01	6.54E-01	NA	-0.003	0.019	1 (0.96-1.04)	8.91E-01	9.37E-01	NA
208	XS_VLDL_FC_pct	0.021	0.019	1.02 (0.99-1.06)	2.59E-01	4.41E-01	NA	0.028	0.019	1.03 (0.99-1.07)	1.43E-01	3.72E-01	NA
209	XS_VLDL_TG_pct	0.002	0.020	1 (0.96-1.04)	9.11E-01	9.45E-01	NA	-0.007	0.020	0.99 (0.96-1.03)	7.22E-01	8.73E-01	NA
210	IDL_PL_pct	0.006	0.019	1.01 (0.97-1.04)	7.50E-01	8.17E-01	NA	0.016	0.019	1.02 (0.98-1.05)	4.09E-01	7.04E-01	NA
211	IDL_C_pct	-0.030	0.019	0.97 (0.94-1.01)	1.11E-01	2.94E-01	NA	-0.027	0.019	0.97 (0.94-1.01)	1.50E-01	3.72E-01	NA
212	IDL_CE_pct	-0.036	0.019	0.96 (0.93-1)	5.09E-02	1.64E-01	NA	-0.037	0.019	0.96 (0.93-1)	4.48E-02	1.91E-01	NA

		Model 3						Model 4					
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***
213	IDL_FC_pct	0.001	0.018	1 (0.97-1.04)	9.67E-01	9.83E-01	NA	0.010	0.018	1.01 (0.97-1.05)	5.89E-01	8.19E-01	NA
214	IDL_TG_pct	0.037	0.019	1.04 (1-1.08)	5.63E-02	1.71E-01	NA	0.030	0.019	1.03 (0.99-1.07)	1.26E-01	3.54E-01	NA
215	L_LDL_PL_pct	0.015	0.019	1.02 (0.98-1.05)	4.20E-01	5.72E-01	NA	0.012	0.019	1.01 (0.98-1.05)	5.18E-01	7.72E-01	NA
216	L_LDL_C_pct	-0.063	0.019	0.94 (0.91-0.97)	7.74E-04	1.20E-02	NA	-0.053	0.019	0.95 (0.91-0.98)	4.59E-03	6.35E-02	NA
217	L_LDL_CE_pct	-0.056	0.018	0.95 (0.91-0.98)	2.27E-03	2.45E-02	NA	-0.052	0.018	0.95 (0.92-0.98)	4.56E-03	6.35E-02	NA
218	L_LDL_FC_pct	-0.022	0.019	0.98 (0.94-1.02)	2.50E-01	4.33E-01	NA	-0.013	0.019	0.99 (0.95-1.02)	4.83E-01	7.56E-01	NA
219	L_LDL_TG_pct	0.056	0.019	1.06 (1.02-1.1)	3.50E-03	2.81E-02	NA	0.048	0.019	1.05 (1.01-1.09)	1.27E-02	9.90E-02	NA
220	M_LDL_PL_pct	0.016	0.019	1.02 (0.98-1.05)	4.09E-01	5.66E-01	NA	0.017	0.019	1.02 (0.98-1.06)	3.62E-01	6.80E-01	NA
221	M_LDL_C_pct	-0.043	0.018	0.96 (0.92-0.99)	2.01E-02	8.07E-02	NA	-0.038	0.018	0.96 (0.93-1)	3.74E-02	1.68E-01	NA
222	M_LDL_CE_pct	-0.011	0.019	0.99 (0.95-1.03)	5.52E-01	6.54E-01	NA	-0.021	0.019	0.98 (0.94-1.02)	2.81E-01	5.87E-01	NA
223	M_LDL_FC_pct	-0.023	0.019	0.98 (0.94-1.01)	2.18E-01	4.28E-01	NA	-0.011	0.019	0.99 (0.95-1.03)	5.79E-01	8.18E-01	NA
224	M_LDL_TG_pct	0.038	0.019	1.04 (1-1.08)	4.82E-02	1.60E-01	NA	0.032	0.019	1.03 (1-1.07)	9.47E-02	3.02E-01	NA
225	S_LDL_PL_pct	0.018	0.020	1.02 (0.98-1.06)	3.54E-01	5.21E-01	NA	0.034	0.020	1.04 (1-1.08)	8.69E-02	2.85E-01	NA
226	S_LDL_C_pct	-0.030	0.019	0.97 (0.93-1.01)	1.23E-01	3.09E-01	NA	-0.035	0.019	0.97 (0.93-1)	7.12E-02	2.48E-01	NA
227	S_LDL_CE_pct	-0.012	0.019	0.99 (0.95-1.03)	5.20E-01	6.41E-01	NA	-0.028	0.019	0.97 (0.94-1.01)	1.46E-01	3.72E-01	NA
228	S_LDL_FC_pct	-0.013	0.018	0.99 (0.95-1.02)	4.72E-01	6.06E-01	NA	-0.002	0.018	1 (0.96-1.03)	8.92E-01	9.37E-01	NA
229	S_LDL_TG_pct	0.015	0.019	1.02 (0.98-1.05)	4.26E-01	5.77E-01	NA	0.006	0.019	1.01 (0.97-1.04)	7.64E-01	8.77E-01	NA
230	XL_HDL_PL_pct	-0.016	0.019	0.98 (0.95-1.02)	4.01E-01	5.66E-01	NA	-0.001	0.020	1 (0.96-1.04)	9.67E-01	9.75E-01	NA
231	XL_HDL_C_pct	0.014	0.020	1.01 (0.98-1.05)	4.88E-01	6.20E-01	NA	0.004	0.020	1 (0.97-1.04)	8.57E-01	9.16E-01	NA
232	XL_HDL_CE_pct	-0.035	0.018	0.97 (0.93-1)	5.32E-02	1.67E-01	NA	-0.034	0.018	0.97 (0.93-1)	5.97E-02	2.25E-01	NA
233	XL_HDL_FC_pct	0.064	0.021	1.07 (1.02-1.11)	1.77E-03	2.10E-02	NA	0.049	0.021	1.05 (1.01-1.09)	1.87E-02	1.14E-01	NA
234	XL_HDL_TG_pct	0.026	0.020	1.03 (0.99-1.07)	1.77E-01	3.83E-01	NA	0.009	0.020	1.01 (0.97-1.05)	6.45E-01	8.25E-01	NA
235	L_HDL_PL_pct	0.053	0.019	1.06 (1.02-1.09)	4.04E-03	2.96E-02	NA	0.039	0.019	1.04 (1-1.08)	3.77E-02	1.68E-01	NA
236	L_HDL_C_pct	-0.047	0.018	0.95 (0.92-0.99)	8.81E-03	4.67E-02	NA	-0.033	0.018	0.97 (0.93-1)	7.18E-02	2.48E-01	NA
237	L_HDL_CE_pct	-0.054	0.017	0.95 (0.92-0.98)	2.06E-03	2.33E-02	NA	-0.041	0.018	0.96 (0.93-0.99)	2.14E-02	1.18E-01	NA
238	L_HDL_FC_pct	0.018	0.019	1.02 (0.98-1.06)	3.59E-01	5.27E-01	NA	0.030	0.020	1.03 (0.99-1.07)	1.26E-01	3.54E-01	NA
239	L_HDL_TG_pct	0.042	0.020	1.04 (1-1.08)	3.04E-02	1.15E-01	NA	0.028	0.020	1.03 (0.99-1.07)	1.53E-01	3.75E-01	NA
240	M_HDL_PL_pct	0.065	0.019	1.07 (1.03-1.11)	6.81E-04	1.19E-02	-17%	0.056	0.019	1.06 (1.02-1.1)	4.04E-03	6.35E-02	NA
241	M_HDL_C_pct	-0.055	0.019	0.95 (0.91-0.98)	3.49E-03	2.81E-02	NA	-0.044	0.019	0.96 (0.92-0.99)	1.99E-02	1.17E-01	NA
242	M_HDL_CE_pct	-0.052	0.018	0.95 (0.92-0.98)	4.93E-03	3.41E-02	NA	-0.042	0.019	0.96 (0.93-0.99)	2.37E-02	1.26E-01	NA
243	M_HDL_FC_pct	-0.041	0.021	0.96 (0.92-1)	4.83E-02	1.60E-01	NA	-0.032	0.021	0.97 (0.93-1.01)	1.35E-01	3.70E-01	NA
244	M_HDL_TG_pct	0.049	0.020	1.05 (1.01-1.09)	1.37E-02	5.99E-02	NA	0.037	0.020	1.04 (1-1.08)	6.58E-02	2.37E-01	NA
245	S_HDL_PL_pct	-0.002	0.020	1 (0.96-1.04)	9.21E-01	9.47E-01	NA	-0.003	0.020	1 (0.96-1.04)	8.66E-01	9.22E-01	NA
246	S_HDL_C_pct	-0.031	0.019	0.97 (0.93-1.01)	9.56E-02	2.68E-01	NA	-0.023	0.019	0.98 (0.94-1.01)	2.16E-01	4.99E-01	NA
247	S_HDL_CE_pct	-0.036	0.019	0.97 (0.93-1)	5.51E-02	1.69E-01	NA	-0.031	0.019	0.97 (0.93-1.01)	9.57E-02	3.02E-01	NA
248	S_HDL_FC_pct	0.016	0.020	1.02 (0.98-1.06)	4.06E-01	5.66E-01	NA	0.030	0.020	1.03 (0.99-1.07)	1.23E-01	3.54E-01	NA
249	S_HDL_TG_pct	0.052	0.020	1.05 (1.01-1.1)	1.04E-02	5.26E-02	NA	0.040	0.020	1.04 (1-1.08)	4.68E-02	1.91E-01	NA

* Percentage change in beta from previous model.

SE, standard error. HR, hazard ratio. CI, confidence intervals. FDR, false discovery rate. NA, not applicable.

Abbreviations are in line with **Supplementary Table 3.1**.

Supplementary Table 4.3 Association of individual metabolites with incident ischaemic stroke in Cox proportional hazards regression in model 5 and 6 (n = 184,564)

		Model 5						Model 6					
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***
1	Total_C	-0.002	0.020	1 (0.96-1.04)	9.03E-01	9.37E-01	NA	0.007	0.020	1.01 (0.97-1.05)	7.23E-01	8.58E-01	NA
2	non_HDL_C	0.015	0.020	1.02 (0.98-1.06)	4.39E-01	7.27E-01	NA	0.012	0.020	1.01 (0.97-1.05)	5.29E-01	7.50E-01	NA
3	Remnant_C	0.021	0.020	1.02 (0.98-1.06)	2.80E-01	5.86E-01	NA	0.015	0.020	1.02 (0.98-1.06)	4.32E-01	6.82E-01	NA
4	VLDL_C	0.019	0.020	1.02 (0.98-1.06)	3.40E-01	6.56E-01	NA	0.000	0.020	1 (0.96-1.04)	9.94E-01	9.98E-01	NA
5	Clinical_LDL_C	0.013	0.020	1.01 (0.98-1.05)	4.90E-01	7.56E-01	NA	0.015	0.020	1.02 (0.98-1.06)	4.37E-01	6.82E-01	NA
6	LDL_C	0.009	0.020	1.01 (0.97-1.05)	6.51E-01	8.31E-01	NA	0.009	0.020	1.01 (0.97-1.05)	6.48E-01	8.31E-01	NA
7	HDL_C	-0.059	0.022	0.94 (0.9-0.98)	6.88E-03	6.85E-02	NA	-0.018	0.022	0.98 (0.94-1.03)	4.18E-01	6.73E-01	NA
8	Total_TG	0.009	0.019	1.01 (0.97-1.05)	6.41E-01	8.31E-01	NA	-0.028	0.020	0.97 (0.94-1.01)	1.70E-01	4.02E-01	NA
9	VLDL_TG	0.002	0.020	1 (0.96-1.04)	9.17E-01	9.40E-01	NA	-0.037	0.020	0.96 (0.93-1)	7.07E-02	2.97E-01	NA
10	LDL_TG	0.047	0.019	1.05 (1.01-1.09)	1.51E-02	1.03E-01	NA	0.019	0.019	1.02 (0.98-1.06)	3.18E-01	5.69E-01	NA
11	HDL_TG	0.009	0.019	1.01 (0.97-1.05)	6.54E-01	8.31E-01	NA	-0.014	0.020	0.99 (0.95-1.02)	4.62E-01	6.98E-01	NA
12	Total_PL	-0.011	0.020	0.99 (0.95-1.03)	5.84E-01	8.17E-01	NA	-0.007	0.020	0.99 (0.95-1.03)	7.27E-01	8.58E-01	NA
13	VLDL_PL	0.016	0.020	1.02 (0.98-1.06)	4.27E-01	7.14E-01	NA	-0.013	0.020	0.99 (0.95-1.03)	5.15E-01	7.38E-01	NA
14	LDL_PL	0.018	0.020	1.02 (0.98-1.06)	3.70E-01	6.87E-01	NA	0.014	0.020	1.01 (0.98-1.05)	4.73E-01	7.00E-01	NA
15	HDL_PL	-0.048	0.022	0.95 (0.91-1)	2.76E-02	1.40E-01	NA	-0.023	0.022	0.98 (0.94-1.02)	2.96E-01	5.53E-01	NA
16	Total_CE	-0.004	0.020	1 (0.96-1.04)	8.26E-01	9.10E-01	NA	0.007	0.020	1.01 (0.97-1.05)	7.26E-01	8.58E-01	NA
17	VLDL_CE	0.022	0.020	1.02 (0.98-1.06)	2.62E-01	5.72E-01	NA	0.009	0.020	1.01 (0.97-1.05)	6.57E-01	8.31E-01	NA
18	LDL_CE	0.008	0.020	1.01 (0.97-1.05)	6.66E-01	8.38E-01	NA	0.006	0.020	1.01 (0.97-1.05)	7.74E-01	9.01E-01	NA
19	HDL_CE	-0.063	0.021	0.94 (0.9-0.98)	3.44E-03	6.54E-02	NA	-0.022	0.022	0.98 (0.94-1.02)	3.16E-01	5.69E-01	NA
20	Total_FC	0.002	0.020	1 (0.96-1.04)	9.03E-01	9.37E-01	NA	0.007	0.020	1.01 (0.97-1.05)	7.20E-01	8.58E-01	NA
21	VLDL_FC	0.014	0.020	1.01 (0.98-1.05)	4.70E-01	7.55E-01	NA	-0.011	0.020	0.99 (0.95-1.03)	5.81E-01	7.86E-01	NA
22	LDL_FC	0.010	0.020	1.01 (0.97-1.05)	6.23E-01	8.31E-01	NA	0.018	0.020	1.02 (0.98-1.06)	3.68E-01	6.23E-01	NA
23	HDL_FC	-0.039	0.022	0.96 (0.92-1)	7.84E-02	2.64E-01	NA	-0.002	0.023	1 (0.96-1.04)	9.46E-01	9.85E-01	NA
24	Total_L	-0.002	0.020	1 (0.96-1.04)	9.14E-01	9.40E-01	NA	-0.008	0.020	0.99 (0.96-1.03)	6.93E-01	8.45E-01	NA
25	VLDL_L	0.009	0.020	1.01 (0.97-1.05)	6.30E-01	8.31E-01	NA	-0.023	0.020	0.98 (0.94-1.02)	2.55E-01	5.15E-01	NA
26	LDL_L	0.014	0.020	1.01 (0.98-1.05)	4.79E-01	7.56E-01	NA	0.011	0.020	1.01 (0.97-1.05)	5.69E-01	7.82E-01	NA
27	HDL_L	-0.053	0.022	0.95 (0.91-0.99)	1.47E-02	1.03E-01	NA	-0.022	0.022	0.98 (0.94-1.02)	3.15E-01	5.69E-01	NA
28	Total_P	-0.051	0.021	0.95 (0.91-0.99)	1.48E-02	1.03E-01	NA	-0.035	0.021	0.97 (0.93-1.01)	9.57E-02	3.24E-01	NA
29	VLDL_P	0.020	0.020	1.02 (0.98-1.06)	2.94E-01	6.00E-01	NA	-0.001	0.020	1 (0.96-1.04)	9.47E-01	9.85E-01	NA
30	LDL_P	0.019	0.019	1.02 (0.98-1.06)	3.24E-01	6.43E-01	NA	0.014	0.019	1.01 (0.98-1.05)	4.70E-01	7.00E-01	NA
31	HDL_P	-0.058	0.021	0.94 (0.91-0.98)	5.90E-03	6.68E-02	NA	-0.040	0.021	0.96 (0.92-1)	5.89E-02	2.90E-01	NA
32	VLDL_size	-0.023	0.020	0.98 (0.94-1.02)	2.43E-01	5.45E-01	NA	-0.067	0.021	0.94 (0.9-0.97)	1.23E-03	7.56E-02	NA
33	LDL_size	-0.004	0.019	1 (0.96-1.03)	8.14E-01	9.10E-01	NA	0.019	0.019	1.02 (0.98-1.06)	3.34E-01	5.78E-01	NA
34	HDL_size	-0.037	0.022	0.96 (0.92-1.01)	9.39E-02	3.04E-01	NA	0.016	0.023	1.02 (0.97-1.06)	4.90E-01	7.18E-01	NA
35	Phosphoglyc	-0.017	0.021	0.98 (0.95-1.02)	4.15E-01	7.07E-01	NA	-0.012	0.021	0.99 (0.95-1.03)	5.47E-01	7.65E-01	NA
36	TG_by_PG	0.018	0.020	1.02 (0.98-1.06)	3.73E-01	6.88E-01	NA	-0.027	0.021	0.97 (0.93-1.01)	1.96E-01	4.35E-01	NA
37	Cholines	-0.021	0.021	0.98 (0.94-1.02)	3.10E-01	6.28E-01	NA	-0.011	0.021	0.99 (0.95-1.03)	5.81E-01	7.86E-01	NA
38	Phosphatidylc	-0.024	0.021	0.98 (0.94-1.02)	2.44E-01	5.45E-01	NA	-0.014	0.021	0.99 (0.95-1.03)	5.06E-01	7.32E-01	NA
39	Sphingomyelins	0.022	0.021	1.02 (0.98-1.07)	2.86E-01	5.94E-01	NA	0.038	0.021	1.04 (1-1.08)	7.20E-02	2.97E-01	NA
40	ApoB	0.023	0.020	1.02 (0.99-1.06)	2.45E-01	5.45E-01	NA	0.016	0.020	1.02 (0.98-1.06)	4.19E-01	6.73E-01	NA

		Model 5						Model 6					
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***
41	ApoA1	-0.056	0.021	0.95 (0.91-0.99)	9.28E-03	8.70E-02	NA	-0.031	0.022	0.97 (0.93-1.01)	1.52E-01	3.86E-01	NA
42	ApoB_by_ApoA1	0.048	0.020	1.05 (1.01-1.09)	1.57E-02	1.03E-01	NA	0.030	0.020	1.03 (0.99-1.07)	1.39E-01	3.72E-01	NA
43	Total_FA	0.007	0.019	1.01 (0.97-1.05)	7.03E-01	8.62E-01	NA	-0.013	0.019	0.99 (0.95-1.03)	5.15E-01	7.38E-01	NA
44	Unsaturation	-0.063	0.020	0.94 (0.9-0.98)	1.78E-03	5.86E-02	NA	-0.027	0.021	0.97 (0.93-1.01)	1.92E-01	4.34E-01	NA
45	Omega_3	-0.039	0.021	0.96 (0.92-1)	6.06E-02	2.26E-01	NA	-0.038	0.021	0.96 (0.93-1)	7.01E-02	2.97E-01	NA
46	Omega_6	-0.026	0.020	0.97 (0.94-1.01)	1.82E-01	4.27E-01	NA	-0.026	0.020	0.97 (0.94-1.01)	1.85E-01	4.24E-01	NA
47	PUFA	-0.029	0.020	0.97 (0.94-1.01)	1.44E-01	3.72E-01	NA	-0.028	0.020	0.97 (0.94-1.01)	1.55E-01	3.86E-01	NA
48	MUFA	0.029	0.019	1.03 (0.99-1.07)	1.39E-01	3.71E-01	NA	-0.004	0.020	1 (0.96-1.04)	8.55E-01	9.46E-01	NA
49	SFA	0.019	0.019	1.02 (0.98-1.06)	3.26E-01	6.43E-01	NA	-0.005	0.020	1 (0.96-1.03)	8.01E-01	9.10E-01	NA
50	LA	-0.036	0.019	0.96 (0.93-1)	6.16E-02	2.26E-01	NA	-0.032	0.019	0.97 (0.93-1.01)	1.02E-01	3.29E-01	NA
51	DHA	-0.032	0.021	0.97 (0.93-1.01)	1.30E-01	3.63E-01	NA	-0.010	0.021	0.99 (0.95-1.03)	6.50E-01	8.31E-01	NA
52	Omega_3_pct	-0.056	0.022	0.95 (0.91-0.99)	9.56E-03	8.70E-02	NA	-0.041	0.022	0.96 (0.92-1)	6.12E-02	2.93E-01	NA
53	Omega_6_pct	-0.051	0.018	0.95 (0.92-0.99)	5.43E-03	6.68E-02	NA	-0.014	0.019	0.99 (0.95-1.02)	4.52E-01	6.90E-01	NA
54	PUFA_pct	-0.063	0.019	0.94 (0.91-0.97)	7.29E-04	3.71E-02	NA	-0.022	0.020	0.98 (0.94-1.02)	2.58E-01	5.18E-01	NA
55	MUFA_pct	0.062	0.020	1.06 (1.02-1.11)	1.88E-03	5.86E-02	NA	0.016	0.021	1.02 (0.98-1.06)	4.39E-01	6.82E-01	NA
56	SFA_pct	0.047	0.019	1.05 (1.01-1.09)	1.32E-02	1.03E-01	NA	0.024	0.019	1.02 (0.99-1.06)	2.24E-01	4.73E-01	NA
57	LA_pct	-0.071	0.019	0.93 (0.9-0.97)	2.06E-04	1.71E-02	NA	-0.034	0.020	0.97 (0.93-1.01)	8.96E-02	3.16E-01	NA
58	DHA_pct	-0.036	0.021	0.96 (0.93-1)	7.70E-02	2.63E-01	NA	-0.001	0.021	1 (0.96-1.04)	9.50E-01	9.85E-01	NA
59	PUFA_by_MUFA	-0.065	0.019	0.94 (0.9-0.97)	7.64E-04	3.71E-02	NA	-0.020	0.020	0.98 (0.94-1.02)	3.26E-01	5.75E-01	NA
60	Omega_6_by_Omega_3	0.033	0.021	1.03 (0.99-1.08)	1.17E-01	3.48E-01	NA	0.032	0.021	1.03 (0.99-1.08)	1.34E-01	3.71E-01	NA
61	Ala	0.045	0.023	1.05 (1-1.1)	5.66E-02	2.21E-01	NA	0.026	0.024	1.03 (0.98-1.08)	2.64E-01	5.22E-01	NA
62	Gln	-0.035	0.019	0.97 (0.93-1)	6.84E-02	2.43E-01	NA	-0.015	0.019	0.99 (0.95-1.02)	4.45E-01	6.88E-01	NA
63	Gly	-0.014	0.017	0.99 (0.96-1.02)	4.01E-01	7.06E-01	NA	-0.007	0.017	0.99 (0.96-1.03)	6.79E-01	8.43E-01	NA
64	His	-0.051	0.019	0.95 (0.92-0.99)	5.48E-03	6.68E-02	NA	-0.050	0.019	0.95 (0.92-0.99)	7.57E-03	1.80E-01	NA
65	Total_BCAA	0.000	0.020	1 (0.96-1.04)	9.83E-01	9.83E-01	NA	-0.034	0.020	0.97 (0.93-1.01)	9.77E-02	3.24E-01	NA
66	Ile	0.007	0.019	1.01 (0.97-1.05)	7.34E-01	8.77E-01	NA	-0.016	0.020	0.98 (0.95-1.02)	4.01E-01	6.57E-01	NA
67	Leu	-0.011	0.020	0.99 (0.95-1.03)	5.69E-01	8.17E-01	NA	-0.037	0.020	0.96 (0.93-1)	6.50E-02	2.97E-01	NA
68	Val	0.004	0.020	1 (0.97-1.04)	8.48E-01	9.14E-01	NA	-0.036	0.021	0.97 (0.93-1.01)	8.61E-02	3.15E-01	NA
69	Phe	0.063	0.019	1.07 (1.03-1.11)	8.93E-04	3.71E-02	NA	0.044	0.019	1.05 (1.01-1.09)	2.29E-02	1.97E-01	NA
70	Tyr	0.052	0.019	1.05 (1.02-1.09)	6.23E-03	6.73E-02	NA	0.024	0.020	1.03 (0.99-1.07)	2.17E-01	4.61E-01	NA
71	Glucose	0.045	0.019	1.05 (1.01-1.09)	1.75E-02	1.12E-01	NA	0.030	0.019	1.03 (0.99-1.07)	1.12E-01	3.36E-01	NA
72	Lactate	0.059	0.020	1.06 (1.02-1.1)	2.90E-03	6.54E-02	NA	0.051	0.020	1.05 (1.01-1.1)	1.02E-02	1.80E-01	NA
73	Pyruvate	0.032	0.020	1.03 (0.99-1.07)	1.05E-01	3.23E-01	NA	0.022	0.020	1.02 (0.98-1.06)	2.77E-01	5.37E-01	NA
74	Citrate	-0.008	0.020	0.99 (0.95-1.03)	7.13E-01	8.67E-01	NA	-0.001	0.020	1 (0.96-1.04)	9.58E-01	9.85E-01	NA
75	bOHbutyrate	0.012	0.021	1.01 (0.97-1.06)	5.59E-01	8.17E-01	NA	0.013	0.022	1.01 (0.97-1.06)	5.45E-01	7.65E-01	NA
76	Acetate	-0.018	0.017	0.98 (0.95-1.02)	2.78E-01	5.86E-01	NA	-0.013	0.017	0.99 (0.96-1.02)	4.36E-01	6.82E-01	NA
77	Acetoacetate	-0.004	0.020	1 (0.96-1.04)	8.44E-01	9.14E-01	NA	-0.011	0.019	0.99 (0.95-1.03)	5.71E-01	7.82E-01	NA
78	Acetone	-0.006	0.019	0.99 (0.96-1.03)	7.70E-01	8.77E-01	NA	0.001	0.019	1 (0.97-1.04)	9.57E-01	9.85E-01	NA
79	Creatinine	0.045	0.022	1.05 (1-1.09)	4.32E-02	1.89E-01	NA	0.024	0.022	1.03 (0.98-1.07)	2.79E-01	5.37E-01	NA
80	Albumin	-0.135	0.019	0.87 (0.84-0.91)	4.61E-13	1.15E-10	0%	-0.119	0.019	0.89 (0.86-0.92)	2.65E-10	6.61E-08	12%
81	GlycA	0.102	0.020	1.11 (1.07-1.15)	3.86E-07	4.81E-05	0%	0.066	0.021	1.07 (1.03-1.11)	1.48E-03	7.56E-02	-35%
82	XXL_VLDL_P	0.007	0.021	1.01 (0.97-1.05)	7.48E-01	8.77E-01	NA	-0.034	0.021	0.97 (0.93-1.01)	1.02E-01	3.29E-01	NA
83	XXL_VLDL_L	-0.001	0.020	1 (0.96-1.04)	9.71E-01	9.79E-01	NA	-0.041	0.020	0.96 (0.92-1)	4.28E-02	2.57E-01	NA

		Model 5						Model 6					
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***
84	XXL_VLDL_PL	0.005	0.021	1.01 (0.96-1.05)	8.29E-01	9.10E-01	NA	-0.034	0.021	0.97 (0.93-1.01)	1.10E-01	3.36E-01	NA
85	XXL_VLDL_C	0.008	0.020	1.01 (0.97-1.05)	6.93E-01	8.54E-01	NA	-0.032	0.021	0.97 (0.93-1.01)	1.22E-01	3.55E-01	NA
86	XXL_VLDL_CE	0.017	0.021	1.02 (0.98-1.06)	4.05E-01	7.06E-01	NA	-0.022	0.021	0.98 (0.94-1.02)	2.87E-01	5.45E-01	NA
87	XXL_VLDL_FC	-0.004	0.020	1 (0.96-1.04)	8.27E-01	9.10E-01	NA	-0.044	0.021	0.96 (0.92-1)	3.10E-02	2.20E-01	NA
88	XXL_VLDL_TG	-0.010	0.019	0.99 (0.95-1.03)	6.07E-01	8.31E-01	NA	-0.039	0.019	0.96 (0.93-1)	3.87E-02	2.51E-01	NA
89	XL_VLDL_P	0.011	0.020	1.01 (0.97-1.05)	5.76E-01	8.17E-01	NA	-0.032	0.021	0.97 (0.93-1.01)	1.36E-01	3.72E-01	NA
90	XL_VLDL_L	0.009	0.020	1.01 (0.97-1.05)	6.40E-01	8.31E-01	NA	-0.033	0.021	0.97 (0.93-1.01)	1.14E-01	3.38E-01	NA
91	XL_VLDL_PL	0.011	0.021	1.01 (0.97-1.05)	5.79E-01	8.17E-01	NA	-0.030	0.021	0.97 (0.93-1.01)	1.59E-01	3.87E-01	NA
92	XL_VLDL_C	0.013	0.020	1.01 (0.97-1.06)	5.06E-01	7.69E-01	NA	-0.023	0.021	0.98 (0.94-1.02)	2.77E-01	5.37E-01	NA
93	XL_VLDL_CE	0.016	0.020	1.02 (0.98-1.06)	4.18E-01	7.07E-01	NA	-0.016	0.021	0.98 (0.95-1.03)	4.48E-01	6.89E-01	NA
94	XL_VLDL_FC	0.010	0.020	1.01 (0.97-1.05)	6.22E-01	8.31E-01	NA	-0.029	0.021	0.97 (0.93-1.01)	1.63E-01	3.91E-01	NA
95	XL_VLDL_TG	0.006	0.020	1.01 (0.97-1.05)	7.52E-01	8.77E-01	NA	-0.038	0.021	0.96 (0.92-1)	7.02E-02	2.97E-01	NA
96	L_VLDL_P	0.007	0.020	1.01 (0.97-1.05)	7.43E-01	8.77E-01	NA	-0.032	0.021	0.97 (0.93-1.01)	1.23E-01	3.55E-01	NA
97	L_VLDL_L	0.003	0.020	1 (0.96-1.04)	8.90E-01	9.37E-01	NA	-0.035	0.021	0.97 (0.93-1.01)	8.96E-02	3.16E-01	NA
98	L_VLDL_PL	0.009	0.020	1.01 (0.97-1.05)	6.54E-01	8.31E-01	NA	-0.031	0.021	0.97 (0.93-1.01)	1.41E-01	3.72E-01	NA
99	L_VLDL_C	0.007	0.020	1.01 (0.97-1.05)	7.42E-01	8.77E-01	NA	-0.027	0.021	0.97 (0.94-1.01)	1.94E-01	4.35E-01	NA
100	L_VLDL_CE	0.004	0.020	1 (0.97-1.04)	8.38E-01	9.11E-01	NA	-0.023	0.020	0.98 (0.94-1.02)	2.61E-01	5.20E-01	NA
101	L_VLDL_FC	0.009	0.020	1.01 (0.97-1.05)	6.49E-01	8.31E-01	NA	-0.030	0.021	0.97 (0.93-1.01)	1.53E-01	3.86E-01	NA
102	L_VLDL_TG	-0.003	0.020	1 (0.96-1.04)	8.95E-01	9.37E-01	NA	-0.040	0.020	0.96 (0.92-1)	4.89E-02	2.59E-01	NA
103	M_VLDL_P	0.003	0.020	1 (0.97-1.04)	8.61E-01	9.21E-01	NA	-0.013	0.020	0.99 (0.95-1.03)	4.99E-01	7.27E-01	NA
104	M_VLDL_L	0.001	0.019	1 (0.96-1.04)	9.60E-01	9.71E-01	NA	-0.020	0.020	0.98 (0.94-1.02)	3.13E-01	5.69E-01	NA
105	M_VLDL_PL	0.007	0.020	1.01 (0.97-1.05)	7.30E-01	8.77E-01	NA	-0.008	0.020	0.99 (0.95-1.03)	6.71E-01	8.43E-01	NA
106	M_VLDL_C	0.009	0.019	1.01 (0.97-1.05)	6.25E-01	8.31E-01	NA	0.010	0.019	1.01 (0.97-1.05)	6.02E-01	7.94E-01	NA
107	M_VLDL_CE	0.013	0.019	1.01 (0.98-1.05)	4.84E-01	7.56E-01	NA	0.022	0.019	1.02 (0.99-1.06)	2.42E-01	4.98E-01	NA
108	M_VLDL_FC	0.008	0.020	1.01 (0.97-1.05)	6.86E-01	8.54E-01	NA	-0.002	0.020	1 (0.96-1.04)	9.05E-01	9.76E-01	NA
109	M_VLDL_TG	-0.004	0.020	1 (0.96-1.04)	8.33E-01	9.10E-01	NA	-0.036	0.020	0.96 (0.93-1)	6.96E-02	2.97E-01	NA
110	S_VLDL_P	0.013	0.020	1.01 (0.98-1.05)	5.06E-01	7.69E-01	NA	-0.011	0.020	0.99 (0.95-1.03)	5.72E-01	7.82E-01	NA
111	S_VLDL_L	0.014	0.020	1.02 (0.98-1.05)	4.63E-01	7.49E-01	NA	-0.009	0.020	0.99 (0.95-1.03)	6.37E-01	8.27E-01	NA
112	S_VLDL_PL	0.016	0.020	1.02 (0.98-1.06)	4.08E-01	7.06E-01	NA	0.000	0.020	1 (0.96-1.04)	9.92E-01	9.98E-01	NA
113	S_VLDL_C	0.024	0.020	1.03 (0.99-1.07)	2.14E-01	4.97E-01	NA	0.010	0.020	1.01 (0.97-1.05)	6.28E-01	8.19E-01	NA
114	S_VLDL_CE	0.028	0.020	1.03 (0.99-1.07)	1.49E-01	3.74E-01	NA	0.011	0.020	1.01 (0.97-1.05)	5.92E-01	7.94E-01	NA
115	S_VLDL_FC	0.017	0.019	1.02 (0.98-1.06)	3.90E-01	7.03E-01	NA	0.007	0.020	1.01 (0.97-1.05)	7.05E-01	8.54E-01	NA
116	S_VLDL_TG	0.004	0.020	1 (0.97-1.04)	8.32E-01	9.10E-01	NA	-0.027	0.020	0.97 (0.94-1.01)	1.84E-01	4.23E-01	NA
117	XS_VLDL_P	0.049	0.020	1.05 (1.01-1.09)	1.25E-02	1.03E-01	NA	0.040	0.020	1.04 (1-1.08)	4.34E-02	2.57E-01	NA
118	XS_VLDL_L	0.048	0.020	1.05 (1.01-1.09)	1.47E-02	1.03E-01	NA	0.039	0.020	1.04 (1-1.08)	4.74E-02	2.59E-01	NA
119	XS_VLDL_PL	0.053	0.020	1.06 (1.02-1.1)	6.49E-03	6.73E-02	NA	0.042	0.020	1.04 (1-1.08)	3.45E-02	2.32E-01	NA
120	XS_VLDL_C	0.046	0.020	1.05 (1.01-1.09)	2.27E-02	1.23E-01	NA	0.048	0.020	1.05 (1.01-1.09)	1.63E-02	1.80E-01	NA
121	XS_VLDL_CE	0.043	0.020	1.04 (1-1.09)	3.42E-02	1.62E-01	NA	0.049	0.020	1.05 (1.01-1.09)	1.47E-02	1.80E-01	NA
122	XS_VLDL_FC	0.051	0.020	1.05 (1.01-1.09)	1.02E-02	8.73E-02	NA	0.044	0.020	1.05 (1.01-1.09)	2.57E-02	2.07E-01	NA
123	XS_VLDL_TG	0.034	0.019	1.03 (1-1.07)	8.43E-02	2.80E-01	NA	0.007	0.020	1.01 (0.97-1.05)	7.34E-01	8.62E-01	NA
124	IDL_P	0.032	0.020	1.03 (0.99-1.07)	1.02E-01	3.16E-01	NA	0.029	0.020	1.03 (0.99-1.07)	1.42E-01	3.72E-01	NA
125	IDL_L	0.028	0.020	1.03 (0.99-1.07)	1.70E-01	4.03E-01	NA	0.035	0.020	1.04 (1-1.08)	8.31E-02	3.09E-01	NA
126	IDL_PL	0.032	0.020	1.03 (0.99-1.07)	1.13E-01	3.43E-01	NA	0.040	0.020	1.04 (1-1.08)	4.49E-02	2.59E-01	NA

		Model 5						Model 6					
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***
127	IDL_C	0.020	0.020	1.02 (0.98-1.06)	3.13E-01	6.29E-01	NA	0.031	0.020	1.03 (0.99-1.07)	1.25E-01	3.58E-01	NA
128	IDL_CE	0.017	0.020	1.02 (0.98-1.06)	3.89E-01	7.03E-01	NA	0.027	0.020	1.03 (0.99-1.07)	1.71E-01	4.02E-01	NA
129	IDL_FC	0.027	0.020	1.03 (0.99-1.07)	1.68E-01	4.01E-01	NA	0.039	0.020	1.04 (1-1.08)	5.28E-02	2.74E-01	NA
130	IDL_TG	0.053	0.019	1.05 (1.02-1.1)	5.89E-03	6.68E-02	NA	0.030	0.020	1.03 (0.99-1.07)	1.28E-01	3.58E-01	NA
131	L_LDL_P	0.017	0.019	1.02 (0.98-1.06)	3.64E-01	6.83E-01	NA	0.016	0.019	1.02 (0.98-1.06)	3.94E-01	6.52E-01	NA
132	L_LDL_L	0.014	0.020	1.01 (0.98-1.05)	4.89E-01	7.56E-01	NA	0.016	0.020	1.02 (0.98-1.06)	4.13E-01	6.72E-01	NA
133	L_LDL_PL	0.016	0.020	1.02 (0.98-1.06)	4.15E-01	7.07E-01	NA	0.017	0.020	1.02 (0.98-1.06)	3.84E-01	6.41E-01	NA
134	L_LDL_C	0.008	0.020	1.01 (0.97-1.05)	6.78E-01	8.48E-01	NA	0.014	0.020	1.01 (0.98-1.05)	4.87E-01	7.18E-01	NA
135	L_LDL_CE	0.008	0.020	1.01 (0.97-1.05)	6.93E-01	8.54E-01	NA	0.011	0.020	1.01 (0.97-1.05)	5.72E-01	7.82E-01	NA
136	L_LDL_FC	0.009	0.020	1.01 (0.97-1.05)	6.47E-01	8.31E-01	NA	0.020	0.020	1.02 (0.98-1.06)	3.02E-01	5.60E-01	NA
137	L_LDL_TG	0.056	0.019	1.06 (1.02-1.1)	3.48E-03	6.54E-02	NA	0.031	0.019	1.03 (0.99-1.07)	1.12E-01	3.36E-01	NA
138	M_LDL_P	0.019	0.020	1.02 (0.98-1.06)	3.31E-01	6.44E-01	NA	0.008	0.020	1.01 (0.97-1.05)	6.87E-01	8.43E-01	NA
139	M_LDL_L	0.013	0.020	1.01 (0.98-1.05)	5.20E-01	7.76E-01	NA	0.001	0.020	1 (0.96-1.04)	9.55E-01	9.85E-01	NA
140	M_LDL_PL	0.015	0.020	1.02 (0.98-1.06)	4.41E-01	7.27E-01	NA	0.004	0.020	1 (0.97-1.04)	8.29E-01	9.21E-01	NA
141	M_LDL_C	0.009	0.020	1.01 (0.97-1.05)	6.46E-01	8.31E-01	NA	-0.001	0.020	1 (0.96-1.04)	9.72E-01	9.89E-01	NA
142	M_LDL_CE	0.009	0.020	1.01 (0.97-1.05)	6.47E-01	8.31E-01	NA	-0.005	0.020	1 (0.96-1.04)	8.12E-01	9.10E-01	NA
143	M_LDL_FC	0.009	0.019	1.01 (0.97-1.05)	6.59E-01	8.33E-01	NA	0.010	0.019	1.01 (0.97-1.05)	5.97E-01	7.94E-01	NA
144	M_LDL_TG	0.037	0.019	1.04 (1-1.08)	5.50E-02	2.21E-01	NA	0.008	0.020	1.01 (0.97-1.05)	6.82E-01	8.43E-01	NA
145	S_LDL_P	0.021	0.020	1.02 (0.98-1.06)	2.76E-01	5.86E-01	NA	0.010	0.020	1.01 (0.97-1.05)	5.97E-01	7.94E-01	NA
146	S_LDL_L	0.018	0.019	1.02 (0.98-1.06)	3.65E-01	6.83E-01	NA	0.008	0.019	1.01 (0.97-1.05)	6.86E-01	8.43E-01	NA
147	S_LDL_PL	0.027	0.019	1.03 (0.99-1.07)	1.66E-01	4.00E-01	NA	0.021	0.019	1.02 (0.98-1.06)	2.80E-01	5.37E-01	NA
148	S_LDL_C	0.013	0.019	1.01 (0.98-1.05)	5.19E-01	7.76E-01	NA	0.005	0.020	1.01 (0.97-1.04)	8.17E-01	9.12E-01	NA
149	S_LDL_CE	0.011	0.020	1.01 (0.97-1.05)	5.62E-01	8.17E-01	NA	-0.001	0.020	1 (0.96-1.04)	9.61E-01	9.85E-01	NA
150	S_LDL_FC	0.014	0.019	1.01 (0.98-1.05)	4.58E-01	7.45E-01	NA	0.018	0.019	1.02 (0.98-1.06)	3.44E-01	5.87E-01	NA
151	S_LDL_TG	0.016	0.019	1.02 (0.98-1.06)	3.93E-01	7.04E-01	NA	-0.015	0.020	0.99 (0.95-1.02)	4.35E-01	6.82E-01	NA
152	XL_HDL_P	-0.012	0.021	0.99 (0.95-1.03)	5.82E-01	8.17E-01	NA	0.040	0.022	1.04 (1-1.09)	7.40E-02	2.97E-01	NA
153	XL_HDL_L	-0.006	0.021	0.99 (0.95-1.04)	7.69E-01	8.77E-01	NA	0.046	0.022	1.05 (1-1.09)	3.94E-02	2.51E-01	NA
154	XL_HDL_PL	-0.005	0.021	1 (0.96-1.04)	8.14E-01	9.10E-01	NA	0.043	0.022	1.04 (1-1.09)	4.88E-02	2.59E-01	NA
155	XL_HDL_C	-0.006	0.021	0.99 (0.95-1.04)	7.71E-01	8.77E-01	NA	0.048	0.022	1.05 (1.01-1.1)	2.84E-02	2.08E-01	NA
156	XL_HDL_CE	-0.015	0.021	0.99 (0.95-1.03)	4.86E-01	7.56E-01	NA	0.041	0.022	1.04 (1-1.09)	6.56E-02	2.97E-01	NA
157	XL_HDL_FC	0.024	0.021	1.02 (0.98-1.07)	2.48E-01	5.46E-01	NA	0.069	0.021	1.07 (1.03-1.12)	1.33E-03	7.56E-02	NA
158	XL_HDL_TG	0.005	0.019	1.01 (0.97-1.04)	7.79E-01	8.82E-01	NA	0.000	0.019	1 (0.96-1.04)	1.00E+00	1.00E+00	NA
159	L_HDL_P	-0.049	0.021	0.95 (0.91-0.99)	2.14E-02	1.21E-01	NA	0.001	0.023	1 (0.96-1.05)	9.73E-01	9.89E-01	NA
160	L_HDL_L	-0.045	0.022	0.96 (0.92-1)	3.86E-02	1.75E-01	NA	0.004	0.023	1 (0.96-1.05)	8.77E-01	9.53E-01	NA
161	L_HDL_PL	-0.041	0.022	0.96 (0.92-1)	5.67E-02	2.21E-01	NA	0.002	0.023	1 (0.96-1.05)	9.17E-01	9.84E-01	NA
162	L_HDL_C	-0.046	0.021	0.96 (0.92-1)	3.05E-02	1.52E-01	NA	0.006	0.022	1.01 (0.96-1.05)	7.98E-01	9.10E-01	NA
163	L_HDL_CE	-0.049	0.021	0.95 (0.91-0.99)	1.93E-02	1.18E-01	NA	0.003	0.022	1 (0.96-1.05)	9.01E-01	9.75E-01	NA
164	L_HDL_FC	-0.031	0.021	0.97 (0.93-1.01)	1.41E-01	3.71E-01	NA	0.016	0.022	1.02 (0.97-1.06)	4.68E-01	7.00E-01	NA
165	L_HDL_TG	-0.011	0.020	0.99 (0.95-1.03)	5.94E-01	8.22E-01	NA	-0.007	0.020	0.99 (0.96-1.03)	7.45E-01	8.71E-01	NA
166	M_HDL_P	-0.060	0.021	0.94 (0.9-0.98)	4.39E-03	6.54E-02	NA	-0.038	0.021	0.96 (0.92-1)	7.27E-02	2.97E-01	NA
167	M_HDL_L	-0.054	0.021	0.95 (0.91-0.99)	9.78E-03	8.70E-02	NA	-0.037	0.021	0.96 (0.92-1)	8.02E-02	3.08E-01	NA
168	M_HDL_PL	-0.049	0.021	0.95 (0.91-0.99)	2.04E-02	1.18E-01	NA	-0.037	0.021	0.96 (0.93-1)	7.90E-02	3.08E-01	NA
169	M_HDL_C	-0.061	0.021	0.94 (0.9-0.98)	3.79E-03	6.54E-02	NA	-0.035	0.021	0.97 (0.93-1.01)	1.06E-01	3.29E-01	NA

		Model 5						Model 6					
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***
170	M_HDL_CE	-0.063	0.021	0.94 (0.9-0.98)	2.79E-03	6.54E-02	NA	-0.036	0.021	0.97 (0.93-1.01)	9.01E-02	3.16E-01	NA
171	M_HDL_FC	-0.052	0.021	0.95 (0.91-0.99)	1.53E-02	1.03E-01	NA	-0.027	0.022	0.97 (0.93-1.02)	2.07E-01	4.45E-01	NA
172	M_HDL_TG	0.006	0.020	1.01 (0.97-1.05)	7.61E-01	8.77E-01	NA	-0.019	0.020	0.98 (0.94-1.02)	3.31E-01	5.76E-01	NA
173	S_HDL_P	-0.037	0.020	0.96 (0.93-1)	6.15E-02	2.26E-01	NA	-0.049	0.020	0.95 (0.92-0.99)	1.51E-02	1.80E-01	NA
174	S_HDL_L	-0.031	0.020	0.97 (0.93-1.01)	1.16E-01	3.48E-01	NA	-0.048	0.020	0.95 (0.92-0.99)	1.66E-02	1.80E-01	NA
175	S_HDL_PL	-0.031	0.020	0.97 (0.93-1.01)	1.21E-01	3.53E-01	NA	-0.048	0.020	0.95 (0.92-0.99)	1.81E-02	1.88E-01	NA
176	S_HDL_C	-0.040	0.020	0.96 (0.92-1)	4.57E-02	1.90E-01	NA	-0.047	0.020	0.95 (0.92-0.99)	2.03E-02	1.95E-01	NA
177	S_HDL_CE	-0.045	0.020	0.96 (0.92-0.99)	2.42E-02	1.26E-01	NA	-0.052	0.020	0.95 (0.91-0.99)	9.49E-03	1.80E-01	NA
178	S_HDL_FC	-0.020	0.020	0.98 (0.94-1.02)	3.31E-01	6.44E-01	NA	-0.024	0.020	0.98 (0.94-1.02)	2.29E-01	4.80E-01	NA
179	S_HDL_TG	0.021	0.020	1.02 (0.98-1.06)	2.88E-01	5.94E-01	NA	-0.020	0.021	0.98 (0.94-1.02)	3.42E-01	5.87E-01	NA
180	XXL_VLDL_PL_pct	0.012	0.023	1.01 (0.97-1.06)	5.84E-01	8.17E-01	NA	-0.006	0.023	0.99 (0.95-1.04)	8.06E-01	9.10E-01	NA
181	XXL_VLDL_C_pct	0.023	0.020	1.02 (0.99-1.06)	2.38E-01	5.44E-01	NA	0.050	0.020	1.05 (1.01-1.09)	1.27E-02	1.80E-01	NA
182	XXL_VLDL_CE_pct	0.040	0.020	1.04 (1-1.08)	4.57E-02	1.90E-01	NA	0.062	0.020	1.06 (1.02-1.11)	2.48E-03	1.03E-01	NA
183	XXL_VLDL_FC_pct	-0.011	0.020	0.99 (0.95-1.03)	5.87E-01	8.17E-01	NA	0.017	0.020	1.02 (0.98-1.06)	3.96E-01	6.52E-01	NA
184	XXL_VLDL_TG_pct	-0.016	0.018	0.98 (0.95-1.02)	3.64E-01	6.83E-01	NA	-0.025	0.017	0.98 (0.94-1.01)	1.55E-01	3.86E-01	NA
185	XL_VLDL_PL_pct	0.018	0.022	1.02 (0.98-1.06)	4.23E-01	7.12E-01	NA	0.005	0.022	1.01 (0.96-1.05)	8.11E-01	9.10E-01	NA
186	XL_VLDL_C_pct	0.004	0.020	1 (0.97-1.05)	8.30E-01	9.10E-01	NA	0.046	0.021	1.05 (1.01-1.09)	2.84E-02	2.08E-01	NA
187	XL_VLDL_CE_pct	0.006	0.020	1.01 (0.97-1.05)	7.66E-01	8.77E-01	NA	0.046	0.021	1.05 (1.01-1.09)	2.45E-02	2.04E-01	NA
188	XL_VLDL_FC_pct	-0.001	0.020	1 (0.96-1.04)	9.55E-01	9.70E-01	NA	0.034	0.021	1.04 (0.99-1.08)	1.05E-01	3.29E-01	NA
189	XL_VLDL_TG_pct	-0.011	0.020	0.99 (0.95-1.03)	5.98E-01	8.23E-01	NA	-0.045	0.021	0.96 (0.92-1)	2.77E-02	2.08E-01	NA
190	L_VLDL_PL_pct	0.033	0.022	1.03 (0.99-1.08)	1.41E-01	3.71E-01	NA	-0.004	0.022	1 (0.95-1.04)	8.70E-01	9.50E-01	NA
191	L_VLDL_C_pct	0.013	0.019	1.01 (0.98-1.05)	4.90E-01	7.56E-01	NA	0.038	0.020	1.04 (1-1.08)	5.42E-02	2.75E-01	NA
192	L_VLDL_CE_pct	0.002	0.019	1 (0.97-1.04)	9.11E-01	9.40E-01	NA	0.038	0.020	1.04 (1-1.08)	5.93E-02	2.90E-01	NA
193	L_VLDL_FC_pct	0.043	0.020	1.04 (1-1.09)	3.44E-02	1.62E-01	NA	0.031	0.020	1.03 (0.99-1.07)	1.28E-01	3.58E-01	NA
194	L_VLDL_TG_pct	-0.028	0.019	0.97 (0.94-1.01)	1.39E-01	3.71E-01	NA	-0.033	0.019	0.97 (0.93-1.01)	9.34E-02	3.23E-01	NA
195	M_VLDL_PL_pct	0.029	0.020	1.03 (0.99-1.07)	1.46E-01	3.72E-01	NA	0.052	0.020	1.05 (1.01-1.1)	8.48E-03	1.80E-01	NA
196	M_VLDL_C_pct	0.015	0.019	1.02 (0.98-1.05)	4.51E-01	7.39E-01	NA	0.049	0.020	1.05 (1.01-1.09)	1.38E-02	1.80E-01	NA
197	M_VLDL_CE_pct	0.016	0.019	1.02 (0.98-1.06)	3.98E-01	7.06E-01	NA	0.050	0.020	1.05 (1.01-1.09)	1.18E-02	1.80E-01	NA
198	M_VLDL_FC_pct	0.021	0.020	1.02 (0.98-1.06)	2.76E-01	5.86E-01	NA	0.051	0.020	1.05 (1.01-1.1)	1.02E-02	1.80E-01	NA
199	M_VLDL_TG_pct	-0.013	0.020	0.99 (0.95-1.03)	5.24E-01	7.76E-01	NA	-0.048	0.020	0.95 (0.92-0.99)	1.93E-02	1.92E-01	NA
200	S_VLDL_PL_pct	0.006	0.020	1.01 (0.97-1.05)	7.68E-01	8.77E-01	NA	0.035	0.020	1.04 (1-1.08)	8.05E-02	3.08E-01	NA
201	S_VLDL_C_pct	0.028	0.019	1.03 (0.99-1.07)	1.51E-01	3.74E-01	NA	0.048	0.020	1.05 (1.01-1.09)	1.38E-02	1.80E-01	NA
202	S_VLDL_CE_pct	0.041	0.019	1.04 (1-1.08)	3.29E-02	1.61E-01	NA	0.052	0.019	1.05 (1.01-1.09)	6.92E-03	1.80E-01	NA
203	S_VLDL_FC_pct	0.006	0.020	1.01 (0.97-1.05)	7.52E-01	8.77E-01	NA	0.037	0.020	1.04 (1-1.08)	6.81E-02	2.97E-01	NA
204	S_VLDL_TG_pct	-0.017	0.020	0.98 (0.95-1.02)	3.83E-01	7.01E-01	NA	-0.041	0.020	0.96 (0.92-1)	4.07E-02	2.53E-01	NA
205	XS_VLDL_PL_pct	0.058	0.020	1.06 (1.02-1.1)	4.35E-03	6.54E-02	NA	0.030	0.021	1.03 (0.99-1.07)	1.39E-01	3.72E-01	NA
206	XS_VLDL_C_pct	0.001	0.019	1 (0.96-1.04)	9.43E-01	9.62E-01	NA	0.035	0.020	1.04 (1-1.08)	8.16E-02	3.08E-01	NA
207	XS_VLDL_CE_pct	-0.002	0.020	1 (0.96-1.04)	9.02E-01	9.37E-01	NA	0.033	0.020	1.03 (0.99-1.08)	1.04E-01	3.29E-01	NA
208	XS_VLDL_FC_pct	0.028	0.019	1.03 (0.99-1.07)	1.42E-01	3.71E-01	NA	0.040	0.019	1.04 (1-1.08)	3.39E-02	2.32E-01	NA
209	XS_VLDL_TG_pct	-0.007	0.020	0.99 (0.95-1.03)	7.12E-01	8.67E-01	NA	-0.040	0.020	0.96 (0.92-1)	4.81E-02	2.59E-01	NA
210	IDL_PL_pct	0.016	0.019	1.02 (0.98-1.05)	4.08E-01	7.06E-01	NA	0.020	0.019	1.02 (0.98-1.06)	2.91E-01	5.49E-01	NA
211	IDL_C_pct	-0.027	0.019	0.97 (0.94-1.01)	1.52E-01	3.74E-01	NA	-0.005	0.019	1 (0.96-1.03)	7.92E-01	9.10E-01	NA
212	IDL_CE_pct	-0.037	0.019	0.96 (0.93-1)	4.55E-02	1.90E-01	NA	-0.019	0.019	0.98 (0.95-1.02)	3.04E-01	5.60E-01	NA

		Model 5						Model 6					
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***
213	IDL_FC_pct	0.010	0.018	1.01 (0.97-1.05)	5.85E-01	8.17E-01	NA	0.027	0.019	1.03 (0.99-1.07)	1.47E-01	3.81E-01	NA
214	IDL_TG_pct	0.029	0.019	1.03 (0.99-1.07)	1.29E-01	3.63E-01	NA	-0.001	0.020	1 (0.96-1.04)	9.50E-01	9.85E-01	NA
215	L_LDL_PL_pct	0.012	0.019	1.01 (0.98-1.05)	5.19E-01	7.76E-01	NA	0.003	0.019	1 (0.97-1.04)	8.67E-01	9.50E-01	NA
216	L_LDL_C_pct	-0.053	0.019	0.95 (0.91-0.98)	4.73E-03	6.54E-02	NA	-0.019	0.019	0.98 (0.94-1.02)	3.26E-01	5.75E-01	NA
217	L_LDL_CE_pct	-0.052	0.018	0.95 (0.92-0.98)	4.62E-03	6.54E-02	NA	-0.042	0.018	0.96 (0.92-0.99)	2.18E-02	1.97E-01	NA
218	L_LDL_FC_pct	-0.013	0.019	0.99 (0.95-1.02)	4.92E-01	7.56E-01	NA	0.023	0.020	1.02 (0.98-1.06)	2.49E-01	5.08E-01	NA
219	L_LDL_TG_pct	0.048	0.019	1.05 (1.01-1.09)	1.31E-02	1.03E-01	NA	0.018	0.020	1.02 (0.98-1.06)	3.71E-01	6.24E-01	NA
220	M_LDL_PL_pct	0.017	0.019	1.02 (0.98-1.06)	3.59E-01	6.83E-01	NA	0.024	0.019	1.03 (0.99-1.06)	2.02E-01	4.40E-01	NA
221	M_LDL_C_pct	-0.038	0.018	0.96 (0.93-1)	3.77E-02	1.74E-01	NA	-0.022	0.019	0.98 (0.94-1.02)	2.41E-01	4.98E-01	NA
222	M_LDL_CE_pct	-0.021	0.019	0.98 (0.94-1.02)	2.77E-01	5.86E-01	NA	-0.045	0.020	0.96 (0.92-0.99)	2.29E-02	1.97E-01	NA
223	M_LDL_FC_pct	-0.010	0.019	0.99 (0.95-1.03)	5.87E-01	8.17E-01	NA	0.027	0.020	1.03 (0.99-1.07)	1.62E-01	3.91E-01	NA
224	M_LDL_TG_pct	0.032	0.019	1.03 (0.99-1.07)	9.63E-02	3.06E-01	NA	0.009	0.019	1.01 (0.97-1.05)	6.57E-01	8.31E-01	NA
225	S_LDL_PL_pct	0.034	0.020	1.04 (1-1.08)	8.69E-02	2.85E-01	NA	0.054	0.020	1.06 (1.02-1.1)	7.36E-03	1.80E-01	NA
226	S_LDL_C_pct	-0.035	0.020	0.97 (0.93-1)	7.21E-02	2.53E-01	NA	-0.025	0.020	0.98 (0.94-1.01)	1.98E-01	4.36E-01	NA
227	S_LDL_CE_pct	-0.028	0.019	0.97 (0.94-1.01)	1.46E-01	3.72E-01	NA	-0.048	0.020	0.95 (0.92-0.99)	1.39E-02	1.80E-01	NA
228	S_LDL_FC_pct	-0.002	0.018	1 (0.96-1.03)	8.97E-01	9.37E-01	NA	0.026	0.019	1.03 (0.99-1.07)	1.75E-01	4.06E-01	NA
229	S_LDL_TG_pct	0.006	0.019	1.01 (0.97-1.04)	7.71E-01	8.77E-01	NA	-0.025	0.020	0.98 (0.94-1.01)	2.03E-01	4.40E-01	NA
230	XL_HDL_PL_pct	-0.001	0.020	1 (0.96-1.04)	9.75E-01	9.79E-01	NA	0.029	0.020	1.03 (0.99-1.07)	1.56E-01	3.86E-01	NA
231	XL_HDL_C_pct	0.003	0.020	1 (0.97-1.04)	8.62E-01	9.21E-01	NA	-0.012	0.020	0.99 (0.95-1.03)	5.30E-01	7.50E-01	NA
232	XL_HDL_CE_pct	-0.034	0.018	0.97 (0.93-1)	6.04E-02	2.26E-01	NA	-0.018	0.018	0.98 (0.95-1.02)	3.28E-01	5.75E-01	NA
233	XL_HDL_FC_pct	0.049	0.021	1.05 (1.01-1.09)	1.92E-02	1.18E-01	NA	0.004	0.022	1 (0.96-1.05)	8.62E-01	9.50E-01	NA
234	XL_HDL_TG_pct	0.009	0.020	1.01 (0.97-1.05)	6.54E-01	8.31E-01	NA	-0.034	0.021	0.97 (0.93-1.01)	9.67E-02	3.24E-01	NA
235	L_HDL_PL_pct	0.039	0.019	1.04 (1-1.08)	3.97E-02	1.76E-01	NA	-0.009	0.020	0.99 (0.95-1.03)	6.55E-01	8.31E-01	NA
236	L_HDL_C_pct	-0.032	0.018	0.97 (0.94-1)	7.48E-02	2.59E-01	NA	0.010	0.019	1.01 (0.97-1.05)	6.03E-01	7.94E-01	NA
237	L_HDL_CE_pct	-0.040	0.018	0.96 (0.93-0.99)	2.25E-02	1.23E-01	NA	0.000	0.019	1 (0.96-1.04)	9.86E-01	9.98E-01	NA
238	L_HDL_FC_pct	0.030	0.020	1.03 (0.99-1.07)	1.24E-01	3.56E-01	NA	0.056	0.020	1.06 (1.02-1.1)	5.17E-03	1.80E-01	NA
239	L_HDL_TG_pct	0.028	0.020	1.03 (0.99-1.07)	1.56E-01	3.80E-01	NA	-0.008	0.020	0.99 (0.95-1.03)	6.82E-01	8.43E-01	NA
240	M_HDL_PL_pct	0.056	0.019	1.06 (1.02-1.1)	4.14E-03	6.54E-02	NA	0.011	0.021	1.01 (0.97-1.05)	6.08E-01	7.97E-01	NA
241	M_HDL_C_pct	-0.044	0.019	0.96 (0.92-0.99)	2.04E-02	1.18E-01	NA	-0.005	0.020	1 (0.96-1.04)	8.06E-01	9.10E-01	NA
242	M_HDL_CE_pct	-0.042	0.019	0.96 (0.93-1)	2.43E-02	1.26E-01	NA	-0.007	0.019	0.99 (0.96-1.03)	7.08E-01	8.54E-01	NA
243	M_HDL_FC_pct	-0.031	0.021	0.97 (0.93-1.01)	1.38E-01	3.71E-01	NA	0.008	0.022	1.01 (0.97-1.05)	7.10E-01	8.54E-01	NA
244	M_HDL_TG_pct	0.037	0.020	1.04 (1-1.08)	6.68E-02	2.41E-01	NA	0.001	0.021	1 (0.96-1.04)	9.62E-01	9.85E-01	NA
245	S_HDL_PL_pct	-0.003	0.020	1 (0.96-1.04)	8.66E-01	9.21E-01	NA	-0.005	0.020	1 (0.96-1.04)	7.95E-01	9.10E-01	NA
246	S_HDL_C_pct	-0.023	0.019	0.98 (0.94-1.01)	2.19E-01	5.06E-01	NA	0.002	0.019	1 (0.97-1.04)	9.31E-01	9.85E-01	NA
247	S_HDL_CE_pct	-0.031	0.019	0.97 (0.93-1.01)	9.72E-02	3.06E-01	NA	-0.014	0.019	0.99 (0.95-1.02)	4.55E-01	6.91E-01	NA
248	S_HDL_FC_pct	0.031	0.020	1.03 (0.99-1.07)	1.22E-01	3.53E-01	NA	0.064	0.020	1.07 (1.03-1.11)	1.52E-03	7.56E-02	NA
249	S_HDL_TG_pct	0.040	0.020	1.04 (1-1.08)	4.76E-02	1.94E-01	NA	0.002	0.021	1 (0.96-1.04)	9.38E-01	9.85E-01	NA

* Percentage change in beta from previous model.

SE, standard error. HR, hazard ratio. CI, confidence intervals. FDR, false discovery rate. NA, not applicable.

Abbreviations are in line with **Supplementary Table 3.1**.

Supplementary Table 4.4 Association of individual metabolites with incident ischaemic stroke in Cox proportional hazards regression in model 7 (n = 184,564)

		Model 7					
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***
1	Total_C	0.008	0.020	1.01 (0.97-1.05)	6.88E-01	8.21E-01	NA
2	non_HDL_C	0.016	0.020	1.02 (0.98-1.06)	4.21E-01	6.08E-01	NA
3	Remnant_C	0.019	0.020	1.02 (0.98-1.06)	3.44E-01	5.38E-01	NA
4	VLDL_C	-0.001	0.020	1 (0.96-1.04)	9.50E-01	9.96E-01	NA
5	Clinical_LDL_C	0.022	0.020	1.02 (0.98-1.06)	2.78E-01	4.84E-01	NA
6	LDL_C	0.013	0.020	1.01 (0.97-1.05)	5.21E-01	7.02E-01	NA
7	HDL_C	-0.026	0.022	0.97 (0.93-1.02)	2.45E-01	4.36E-01	NA
8	Total_TG	-0.040	0.020	0.96 (0.92-1)	4.53E-02	1.42E-01	NA
9	VLDL_TG	-0.048	0.020	0.95 (0.92-0.99)	1.80E-02	8.13E-02	NA
10	LDL_TG	0.006	0.020	1.01 (0.97-1.05)	7.68E-01	8.72E-01	NA
11	HDL_TG	-0.031	0.020	0.97 (0.93-1.01)	1.19E-01	2.53E-01	NA
12	Total_PL	-0.018	0.021	0.98 (0.94-1.02)	3.81E-01	5.75E-01	NA
13	VLDL_PL	-0.019	0.020	0.98 (0.94-1.02)	3.33E-01	5.32E-01	NA
14	LDL_PL	0.018	0.020	1.02 (0.98-1.06)	3.74E-01	5.70E-01	NA
15	HDL_PL	-0.040	0.022	0.96 (0.92-1)	7.12E-02	1.75E-01	NA
16	Total_CE	0.008	0.020	1.01 (0.97-1.05)	6.96E-01	8.21E-01	NA
17	VLDL_CE	0.010	0.020	1.01 (0.97-1.05)	6.03E-01	7.70E-01	NA
18	LDL_CE	0.008	0.020	1.01 (0.97-1.05)	6.94E-01	8.21E-01	NA
19	HDL_CE	-0.030	0.022	0.97 (0.93-1.01)	1.80E-01	3.39E-01	NA
20	Total_FC	0.008	0.020	1.01 (0.97-1.05)	6.73E-01	8.21E-01	NA
21	VLDL_FC	-0.016	0.020	0.98 (0.95-1.02)	4.25E-01	6.08E-01	NA
22	LDL_FC	0.026	0.020	1.03 (0.99-1.07)	1.94E-01	3.62E-01	NA
23	HDL_FC	-0.010	0.023	0.99 (0.95-1.04)	6.60E-01	8.19E-01	NA
24	Total_L	-0.015	0.020	0.99 (0.95-1.02)	4.47E-01	6.33E-01	NA
25	VLDL_L	-0.031	0.020	0.97 (0.93-1.01)	1.21E-01	2.55E-01	NA
26	LDL_L	0.014	0.020	1.01 (0.98-1.05)	4.89E-01	6.64E-01	NA
27	HDL_L	-0.037	0.022	0.96 (0.92-1.01)	1.01E-01	2.26E-01	NA
28	Total_P	-0.052	0.021	0.95 (0.91-0.99)	1.43E-02	7.28E-02	NA
29	VLDL_P	-0.005	0.020	1 (0.96-1.03)	7.90E-01	8.82E-01	NA
30	LDL_P	0.019	0.020	1.02 (0.98-1.06)	3.33E-01	5.32E-01	NA
31	HDL_P	-0.058	0.021	0.94 (0.91-0.98)	6.21E-03	4.59E-02	NA
32	VLDL_size	-0.078	0.021	0.93 (0.89-0.96)	1.50E-04	9.36E-03	NA
33	LDL_size	0.030	0.019	1.03 (0.99-1.07)	1.24E-01	2.57E-01	NA
34	HDL_size	0.021	0.023	1.02 (0.98-1.07)	3.56E-01	5.55E-01	NA
35	Phosphoglyc	-0.029	0.021	0.97 (0.93-1.01)	1.68E-01	3.19E-01	NA
36	TG_by_PG	-0.035	0.021	0.97 (0.93-1.01)	9.77E-02	2.23E-01	NA
37	Cholines	-0.025	0.021	0.98 (0.94-1.02)	2.29E-01	4.13E-01	NA
38	Phosphatidylc	-0.026	0.021	0.97 (0.94-1.02)	2.07E-01	3.82E-01	NA
39	Sphingomyelins	0.033	0.021	1.03 (0.99-1.08)	1.22E-01	2.55E-01	NA
40	ApoB	0.020	0.020	1.02 (0.98-1.06)	3.00E-01	5.08E-01	NA
41	ApoA1	-0.049	0.022	0.95 (0.91-0.99)	2.54E-02	9.65E-02	NA

		Model 7					
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***
42	ApoB_by_ApoA1	0.042	0.020	1.04 (1-1.09)	3.54E-02	1.17E-01	NA
43	Total_FA	-0.027	0.020	0.97 (0.94-1.01)	1.61E-01	3.09E-01	NA
44	Unsaturation	-0.011	0.021	0.99 (0.95-1.03)	6.02E-01	7.70E-01	NA
45	Omega_3	-0.041	0.021	0.96 (0.92-1)	4.85E-02	1.44E-01	NA
46	Omega_6	-0.031	0.020	0.97 (0.93-1.01)	1.13E-01	2.45E-01	NA
47	PUFA	-0.033	0.020	0.97 (0.93-1.01)	9.22E-02	2.13E-01	NA
48	MUFA	-0.020	0.020	0.98 (0.94-1.02)	3.09E-01	5.19E-01	NA
49	SFA	-0.024	0.020	0.98 (0.94-1.01)	2.19E-01	4.02E-01	NA
50	LA	-0.033	0.020	0.97 (0.93-1.01)	9.22E-02	2.13E-01	NA
51	DHA	-0.009	0.021	0.99 (0.95-1.03)	6.69E-01	8.21E-01	NA
52	Omega_3_pct	-0.035	0.022	0.97 (0.93-1.01)	1.08E-01	2.40E-01	NA
53	Omega_6_pct	0.007	0.019	1.01 (0.97-1.05)	7.14E-01	8.34E-01	NA
54	PUFA_pct	0.002	0.020	1 (0.96-1.04)	9.38E-01	9.96E-01	NA
55	MUFA_pct	0.000	0.021	1 (0.96-1.04)	9.96E-01	9.99E-01	NA
56	SFA_pct	0.000	0.020	1 (0.96-1.04)	9.99E-01	9.99E-01	NA
57	LA_pct	-0.010	0.020	0.99 (0.95-1.03)	6.15E-01	7.73E-01	NA
58	DHA_pct	0.009	0.021	1.01 (0.97-1.05)	6.70E-01	8.21E-01	NA
59	PUFA_by_MUFA	0.001	0.021	1 (0.96-1.04)	9.72E-01	9.96E-01	NA
60	Omega_6_by_Omega_3	0.034	0.021	1.03 (0.99-1.08)	1.16E-01	2.48E-01	NA
61	Ala	0.012	0.024	1.01 (0.97-1.06)	6.06E-01	7.70E-01	NA
62	Gln	0.001	0.019	1 (0.96-1.04)	9.56E-01	9.96E-01	NA
63	Gly	-0.003	0.018	1 (0.96-1.03)	8.76E-01	9.52E-01	NA
64	His	-0.044	0.019	0.96 (0.92-0.99)	1.78E-02	8.13E-02	NA
65	Total_BCAA	-0.034	0.020	0.97 (0.93-1.01)	9.97E-02	2.26E-01	NA
66	Ile	-0.014	0.020	0.99 (0.95-1.02)	4.61E-01	6.43E-01	NA
67	Leu	-0.032	0.020	0.97 (0.93-1.01)	1.12E-01	2.44E-01	NA
68	Val	-0.040	0.021	0.96 (0.92-1)	5.56E-02	1.56E-01	NA
69	Phe	0.053	0.019	1.05 (1.02-1.1)	5.84E-03	4.59E-02	NA
70	Tyr	0.029	0.020	1.03 (0.99-1.07)	1.38E-01	2.81E-01	NA
71	Glucose	-0.015	0.019	0.99 (0.95-1.02)	4.22E-01	6.08E-01	NA
72	Lactate	0.040	0.020	1.04 (1-1.08)	4.77E-02	1.44E-01	NA
73	Pyruvate	0.020	0.020	1.02 (0.98-1.06)	3.22E-01	5.30E-01	NA
74	Citrate	-0.011	0.021	0.99 (0.95-1.03)	5.87E-01	7.70E-01	NA
75	bOHbutyrate	-0.001	0.021	1 (0.96-1.04)	9.43E-01	9.96E-01	NA
76	Acetate	-0.013	0.017	0.99 (0.96-1.02)	4.50E-01	6.33E-01	NA
77	Acetoacetate	-0.017	0.019	0.98 (0.95-1.02)	3.72E-01	5.70E-01	NA
78	Acetone	0.000	0.019	1 (0.96-1.04)	9.95E-01	9.99E-01	NA
79	Creatinine	0.038	0.022	1.04 (0.99-1.09)	8.92E-02	2.12E-01	NA
80	Albumin	-0.136	0.019	0.87 (0.84-0.91)	4.21E-13	1.05E-10	-14%
81	GlycA	0.045	0.021	1.05 (1-1.09)	3.12E-02	1.10E-01	NA
82	XXL_VLDL_P	-0.049	0.021	0.95 (0.91-0.99)	2.00E-02	8.30E-02	NA
83	XXL_VLDL_L	-0.055	0.020	0.95 (0.91-0.98)	6.33E-03	4.59E-02	NA
84	XXL_VLDL_PL	-0.049	0.021	0.95 (0.91-0.99)	2.18E-02	8.48E-02	NA
85	XXL_VLDL_C	-0.048	0.021	0.95 (0.92-0.99)	2.13E-02	8.48E-02	NA

		Model 7					
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***
86	XXL_VLDL_CE	-0.037	0.021	0.96 (0.92-1)	7.51E-02	1.81E-01	NA
87	XXL_VLDL_FC	-0.060	0.021	0.94 (0.9-0.98)	3.31E-03	3.58E-02	NA
88	XXL_VLDL_TG	-0.048	0.019	0.95 (0.92-0.99)	1.00E-02	5.79E-02	NA
89	XL_VLDL_P	-0.043	0.021	0.96 (0.92-1)	4.30E-02	1.39E-01	NA
90	XL_VLDL_L	-0.044	0.021	0.96 (0.92-1)	3.52E-02	1.17E-01	NA
91	XL_VLDL_PL	-0.040	0.021	0.96 (0.92-1)	6.07E-02	1.60E-01	NA
92	XL_VLDL_C	-0.030	0.021	0.97 (0.93-1.01)	1.52E-01	2.97E-01	NA
93	XL_VLDL_CE	-0.021	0.021	0.98 (0.94-1.02)	3.20E-01	5.30E-01	NA
94	XL_VLDL_FC	-0.038	0.021	0.96 (0.92-1)	6.66E-02	1.68E-01	NA
95	XL_VLDL_TG	-0.050	0.021	0.95 (0.91-0.99)	1.67E-02	7.86E-02	NA
96	L_VLDL_P	-0.041	0.021	0.96 (0.92-1)	4.73E-02	1.44E-01	NA
97	L_VLDL_L	-0.043	0.020	0.96 (0.92-1)	3.50E-02	1.17E-01	NA
98	L_VLDL_PL	-0.040	0.021	0.96 (0.92-1)	5.85E-02	1.58E-01	NA
99	L_VLDL_C	-0.033	0.021	0.97 (0.93-1.01)	1.11E-01	2.44E-01	NA
100	L_VLDL_CE	-0.026	0.020	0.97 (0.94-1.01)	2.02E-01	3.75E-01	NA
101	L_VLDL_FC	-0.039	0.021	0.96 (0.92-1)	6.43E-02	1.67E-01	NA
102	L_VLDL_TG	-0.049	0.020	0.95 (0.92-0.99)	1.60E-02	7.80E-02	NA
103	M_VLDL_P	-0.014	0.020	0.99 (0.95-1.03)	4.90E-01	6.64E-01	NA
104	M_VLDL_L	-0.022	0.020	0.98 (0.94-1.02)	2.60E-01	4.59E-01	NA
105	M_VLDL_PL	-0.008	0.020	0.99 (0.96-1.03)	7.04E-01	8.27E-01	NA
106	M_VLDL_C	0.019	0.020	1.02 (0.98-1.06)	3.27E-01	5.32E-01	NA
107	M_VLDL_CE	0.037	0.019	1.04 (1-1.08)	5.90E-02	1.58E-01	NA
108	M_VLDL_FC	0.000	0.020	1 (0.96-1.04)	9.80E-01	9.96E-01	NA
109	M_VLDL_TG	-0.044	0.020	0.96 (0.92-1)	2.68E-02	9.91E-02	NA
110	S_VLDL_P	-0.016	0.020	0.98 (0.95-1.02)	4.14E-01	6.08E-01	NA
111	S_VLDL_L	-0.015	0.020	0.99 (0.95-1.03)	4.62E-01	6.43E-01	NA
112	S_VLDL_PL	-0.001	0.020	1 (0.96-1.04)	9.50E-01	9.96E-01	NA
113	S_VLDL_C	0.010	0.020	1.01 (0.97-1.05)	6.05E-01	7.70E-01	NA
114	S_VLDL_CE	0.010	0.020	1.01 (0.97-1.05)	6.09E-01	7.70E-01	NA
115	S_VLDL_FC	0.010	0.020	1.01 (0.97-1.05)	6.07E-01	7.70E-01	NA
116	S_VLDL_TG	-0.037	0.020	0.96 (0.93-1)	6.55E-02	1.68E-01	NA
117	XS_VLDL_P	0.040	0.020	1.04 (1-1.08)	4.51E-02	1.42E-01	NA
118	XS_VLDL_L	0.038	0.020	1.04 (1-1.08)	5.56E-02	1.56E-01	NA
119	XS_VLDL_PL	0.040	0.020	1.04 (1-1.08)	4.56E-02	1.42E-01	NA
120	XS_VLDL_C	0.054	0.020	1.06 (1.02-1.1)	7.36E-03	4.59E-02	NA
121	XS_VLDL_CE	0.058	0.020	1.06 (1.02-1.1)	4.47E-03	4.04E-02	NA
122	XS_VLDL_FC	0.045	0.020	1.05 (1.01-1.09)	2.56E-02	9.65E-02	NA
123	XS_VLDL_TG	-0.006	0.020	0.99 (0.96-1.03)	7.66E-01	8.72E-01	NA
124	IDL_P	0.037	0.020	1.04 (1-1.08)	6.11E-02	1.60E-01	NA
125	IDL_L	0.042	0.020	1.04 (1-1.09)	3.95E-02	1.29E-01	NA
126	IDL_PL	0.048	0.020	1.05 (1.01-1.09)	1.90E-02	8.14E-02	NA
127	IDL_C	0.040	0.020	1.04 (1-1.08)	4.90E-02	1.44E-01	NA
128	IDL_CE	0.037	0.020	1.04 (1-1.08)	6.90E-02	1.72E-01	NA
129	IDL_FC	0.047	0.020	1.05 (1.01-1.09)	2.04E-02	8.34E-02	NA

		Model 7					
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***
130	IDL_TG	0.017	0.020	1.02 (0.98-1.06)	3.75E-01	5.70E-01	NA
131	L_LDL_P	0.023	0.019	1.02 (0.99-1.06)	2.29E-01	4.13E-01	NA
132	L_LDL_L	0.020	0.020	1.02 (0.98-1.06)	3.17E-01	5.30E-01	NA
133	L_LDL_PL	0.021	0.020	1.02 (0.98-1.06)	2.91E-01	5.00E-01	NA
134	L_LDL_C	0.019	0.020	1.02 (0.98-1.06)	3.39E-01	5.38E-01	NA
135	L_LDL_CE	0.015	0.020	1.02 (0.98-1.06)	4.42E-01	6.28E-01	NA
136	L_LDL_FC	0.029	0.020	1.03 (0.99-1.07)	1.47E-01	2.94E-01	NA
137	L_LDL_TG	0.017	0.020	1.02 (0.98-1.06)	3.76E-01	5.70E-01	NA
138	M_LDL_P	0.008	0.020	1.01 (0.97-1.05)	6.84E-01	8.21E-01	NA
139	M_LDL_L	0.001	0.020	1 (0.96-1.04)	9.74E-01	9.96E-01	NA
140	M_LDL_PL	0.005	0.020	1.01 (0.97-1.05)	7.84E-01	8.80E-01	NA
141	M_LDL_C	0.000	0.020	1 (0.96-1.04)	9.80E-01	9.96E-01	NA
142	M_LDL_CE	-0.007	0.020	0.99 (0.96-1.03)	7.38E-01	8.50E-01	NA
143	M_LDL_FC	0.016	0.020	1.02 (0.98-1.06)	4.05E-01	6.00E-01	NA
144	M_LDL_TG	-0.006	0.020	0.99 (0.96-1.03)	7.71E-01	8.72E-01	NA
145	S_LDL_P	0.014	0.020	1.01 (0.98-1.05)	4.87E-01	6.64E-01	NA
146	S_LDL_L	0.011	0.020	1.01 (0.97-1.05)	5.73E-01	7.59E-01	NA
147	S_LDL_PL	0.028	0.019	1.03 (0.99-1.07)	1.49E-01	2.94E-01	NA
148	S_LDL_C	0.008	0.020	1.01 (0.97-1.05)	6.93E-01	8.21E-01	NA
149	S_LDL_CE	-0.001	0.020	1 (0.96-1.04)	9.76E-01	9.96E-01	NA
150	S_LDL_FC	0.028	0.019	1.03 (0.99-1.07)	1.43E-01	2.89E-01	NA
151	S_LDL_TG	-0.027	0.020	0.97 (0.94-1.01)	1.60E-01	3.09E-01	NA
152	XL_HDL_P	0.052	0.022	1.05 (1.01-1.1)	1.83E-02	8.13E-02	NA
153	XL_HDL_L	0.059	0.022	1.06 (1.02-1.11)	7.34E-03	4.59E-02	NA
154	XL_HDL_PL	0.056	0.022	1.06 (1.01-1.1)	1.05E-02	5.79E-02	NA
155	XL_HDL_C	0.064	0.022	1.07 (1.02-1.11)	3.71E-03	3.85E-02	NA
156	XL_HDL_CE	0.056	0.022	1.06 (1.01-1.11)	1.16E-02	6.12E-02	NA
157	XL_HDL_FC	0.085	0.021	1.09 (1.04-1.14)	6.63E-05	8.25E-03	NA
158	XL_HDL_TG	-0.006	0.019	0.99 (0.96-1.03)	7.58E-01	8.70E-01	NA
159	L_HDL_P	0.004	0.023	1 (0.96-1.05)	8.48E-01	9.35E-01	NA
160	L_HDL_L	0.005	0.023	1.01 (0.96-1.05)	8.27E-01	9.19E-01	NA
161	L_HDL_PL	0.000	0.023	1 (0.96-1.05)	9.87E-01	9.99E-01	NA
162	L_HDL_C	0.012	0.022	1.01 (0.97-1.06)	5.94E-01	7.70E-01	NA
163	L_HDL_CE	0.009	0.022	1.01 (0.97-1.05)	6.76E-01	8.21E-01	NA
164	L_HDL_FC	0.021	0.022	1.02 (0.98-1.07)	3.42E-01	5.38E-01	NA
165	L_HDL_TG	-0.016	0.020	0.98 (0.95-1.02)	4.20E-01	6.08E-01	NA
166	M_HDL_P	-0.056	0.021	0.95 (0.91-0.99)	8.55E-03	5.07E-02	NA
167	M_HDL_L	-0.057	0.021	0.95 (0.91-0.99)	7.22E-03	4.59E-02	NA
168	M_HDL_PL	-0.058	0.021	0.94 (0.91-0.98)	5.51E-03	4.57E-02	NA
169	M_HDL_C	-0.050	0.021	0.95 (0.91-0.99)	1.89E-02	8.14E-02	NA
170	M_HDL_CE	-0.052	0.021	0.95 (0.91-0.99)	1.55E-02	7.72E-02	NA
171	M_HDL_FC	-0.043	0.022	0.96 (0.92-1)	4.83E-02	1.44E-01	NA
172	M_HDL_TG	-0.038	0.020	0.96 (0.93-1)	5.85E-02	1.58E-01	NA
173	S_HDL_P	-0.069	0.020	0.93 (0.9-0.97)	5.95E-04	1.23E-02	NA

		Model 7					
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***
174	S_HDL_L	-0.074	0.020	0.93 (0.89-0.97)	2.62E-04	1.09E-02	NA
175	S_HDL_PL	-0.075	0.020	0.93 (0.89-0.97)	2.22E-04	1.09E-02	NA
176	S_HDL_C	-0.066	0.020	0.94 (0.9-0.97)	1.03E-03	1.50E-02	NA
177	S_HDL_CE	-0.070	0.020	0.93 (0.9-0.97)	4.69E-04	1.19E-02	NA
178	S_HDL_FC	-0.044	0.020	0.96 (0.92-1)	2.94E-02	1.06E-01	NA
179	S_HDL_TG	-0.037	0.021	0.96 (0.93-1)	7.34E-02	1.79E-01	NA
180	XXL_VLDL_PL_pct	-0.014	0.023	0.99 (0.94-1.03)	5.34E-01	7.11E-01	NA
181	XXL_VLDL_C_pct	0.056	0.020	1.06 (1.02-1.1)	5.14E-03	4.41E-02	NA
182	XXL_VLDL_CE_pct	0.068	0.020	1.07 (1.03-1.11)	9.00E-04	1.48E-02	NA
183	XXL_VLDL_FC_pct	0.022	0.020	1.02 (0.98-1.06)	2.83E-01	4.89E-01	NA
184	XXL_VLDL_TG_pct	-0.027	0.017	0.97 (0.94-1.01)	1.25E-01	2.58E-01	NA
185	XL_VLDL_PL_pct	0.008	0.022	1.01 (0.97-1.05)	7.30E-01	8.50E-01	NA
186	XL_VLDL_C_pct	0.062	0.021	1.06 (1.02-1.11)	2.93E-03	3.32E-02	NA
187	XL_VLDL_CE_pct	0.063	0.021	1.07 (1.02-1.11)	2.38E-03	2.83E-02	NA
188	XL_VLDL_FC_pct	0.048	0.021	1.05 (1.01-1.09)	2.15E-02	8.48E-02	NA
189	XL_VLDL_TG_pct	-0.059	0.021	0.94 (0.91-0.98)	4.49E-03	4.04E-02	NA
190	L_VLDL_PL_pct	-0.011	0.022	0.99 (0.95-1.03)	6.34E-01	7.93E-01	NA
191	L_VLDL_C_pct	0.048	0.020	1.05 (1.01-1.09)	1.34E-02	6.95E-02	NA
192	L_VLDL_CE_pct	0.053	0.020	1.05 (1.01-1.1)	8.29E-03	5.03E-02	NA
193	L_VLDL_FC_pct	0.029	0.020	1.03 (0.99-1.07)	1.54E-01	3.00E-01	NA
194	L_VLDL_TG_pct	-0.037	0.019	0.96 (0.93-1)	5.73E-02	1.58E-01	NA
195	M_VLDL_PL_pct	0.069	0.020	1.07 (1.03-1.11)	6.41E-04	1.23E-02	NA
196	M_VLDL_C_pct	0.069	0.020	1.07 (1.03-1.11)	6.20E-04	1.23E-02	NA
197	M_VLDL_CE_pct	0.071	0.020	1.07 (1.03-1.12)	4.03E-04	1.19E-02	NA
198	M_VLDL_FC_pct	0.068	0.020	1.07 (1.03-1.11)	7.75E-04	1.38E-02	NA
199	M_VLDL_TG_pct	-0.064	0.021	0.94 (0.9-0.98)	1.84E-03	2.29E-02	NA
200	S_VLDL_PL_pct	0.052	0.020	1.05 (1.01-1.1)	1.05E-02	5.79E-02	NA
201	S_VLDL_C_pct	0.063	0.020	1.07 (1.03-1.11)	1.32E-03	1.79E-02	NA
202	S_VLDL_CE_pct	0.064	0.019	1.07 (1.03-1.11)	9.48E-04	1.48E-02	NA
203	S_VLDL_FC_pct	0.054	0.020	1.06 (1.02-1.1)	7.38E-03	4.59E-02	NA
204	S_VLDL_TG_pct	-0.054	0.020	0.95 (0.91-0.99)	6.78E-03	4.59E-02	NA
205	XS_VLDL_PL_pct	0.022	0.021	1.02 (0.98-1.07)	2.76E-01	4.84E-01	NA
206	XS_VLDL_C_pct	0.055	0.020	1.06 (1.02-1.1)	6.15E-03	4.59E-02	NA
207	XS_VLDL_CE_pct	0.055	0.020	1.06 (1.02-1.1)	7.23E-03	4.59E-02	NA
208	XS_VLDL_FC_pct	0.048	0.019	1.05 (1.01-1.09)	1.15E-02	6.12E-02	NA
209	XS_VLDL_TG_pct	-0.059	0.021	0.94 (0.91-0.98)	4.00E-03	3.99E-02	NA
210	IDL_PL_pct	0.020	0.019	1.02 (0.98-1.06)	2.97E-01	5.07E-01	NA
211	IDL_C_pct	0.010	0.019	1.01 (0.97-1.05)	5.83E-01	7.68E-01	NA
212	IDL_CE_pct	-0.004	0.019	1 (0.96-1.03)	8.31E-01	9.20E-01	NA
213	IDL_FC_pct	0.033	0.018	1.03 (1-1.07)	7.74E-02	1.85E-01	NA
214	IDL_TG_pct	-0.020	0.020	0.98 (0.94-1.02)	3.25E-01	5.32E-01	NA
215	L_LDL_PL_pct	0.003	0.018	1 (0.97-1.04)	8.90E-01	9.64E-01	NA
216	L_LDL_C_pct	0.001	0.020	1 (0.96-1.04)	9.46E-01	9.96E-01	NA
217	L_LDL_CE_pct	-0.036	0.018	0.97 (0.93-1)	5.35E-02	1.53E-01	NA

		Model 7					
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**% change***
218	L_LDL_FC_pct	0.042	0.020	1.04 (1-1.08)	3.34E-02	1.16E-01	NA
219	L_LDL_TG_pct	-0.001	0.020	1 (0.96-1.04)	9.54E-01	9.96E-01	NA
220	M_LDL_PL_pct	0.037	0.019	1.04 (1-1.08)	5.33E-02	1.53E-01	NA
221	M_LDL_C_pct	-0.014	0.019	0.99 (0.95-1.02)	4.68E-01	6.47E-01	NA
222	M_LDL_CE_pct	-0.056	0.020	0.95 (0.91-0.98)	4.54E-03	4.04E-02	NA
223	M_LDL_FC_pct	0.046	0.020	1.05 (1.01-1.09)	1.93E-02	8.14E-02	NA
224	M_LDL_TG_pct	-0.008	0.020	0.99 (0.96-1.03)	6.89E-01	8.21E-01	NA
225	S_LDL_PL_pct	0.070	0.020	1.07 (1.03-1.12)	4.79E-04	1.19E-02	NA
226	S_LDL_C_pct	-0.023	0.020	0.98 (0.94-1.02)	2.41E-01	4.32E-01	NA
227	S_LDL_CE_pct	-0.063	0.020	0.94 (0.9-0.98)	1.36E-03	1.79E-02	NA
228	S_LDL_FC_pct	0.042	0.019	1.04 (1.01-1.08)	2.71E-02	9.91E-02	NA
229	S_LDL_TG_pct	-0.042	0.020	0.96 (0.92-1)	3.10E-02	1.10E-01	NA
230	XL_HDL_PL_pct	0.037	0.020	1.04 (1-1.08)	6.67E-02	1.68E-01	NA
231	XL_HDL_C_pct	-0.012	0.019	0.99 (0.95-1.03)	5.34E-01	7.11E-01	NA
232	XL_HDL_CE_pct	-0.013	0.018	0.99 (0.95-1.02)	4.73E-01	6.51E-01	NA
233	XL_HDL_FC_pct	-0.001	0.022	1 (0.96-1.04)	9.64E-01	9.96E-01	NA
234	XL_HDL_TG_pct	-0.049	0.020	0.95 (0.92-0.99)	1.64E-02	7.86E-02	NA
235	L_HDL_PL_pct	-0.034	0.020	0.97 (0.93-1.01)	9.25E-02	2.13E-01	NA
236	L_HDL_C_pct	0.028	0.019	1.03 (0.99-1.07)	1.45E-01	2.92E-01	NA
237	L_HDL_CE_pct	0.016	0.019	1.02 (0.98-1.05)	3.99E-01	5.95E-01	NA
238	L_HDL_FC_pct	0.072	0.020	1.08 (1.03-1.12)	3.34E-04	1.19E-02	NA
239	L_HDL_TG_pct	-0.018	0.020	0.98 (0.94-1.02)	3.84E-01	5.76E-01	NA
240	M_HDL_PL_pct	0.004	0.021	1 (0.96-1.05)	8.62E-01	9.41E-01	NA
241	M_HDL_C_pct	0.003	0.020	1 (0.96-1.04)	8.94E-01	9.64E-01	NA
242	M_HDL_CE_pct	0.001	0.019	1 (0.96-1.04)	9.74E-01	9.96E-01	NA
243	M_HDL_FC_pct	0.010	0.022	1.01 (0.97-1.06)	6.61E-01	8.19E-01	NA
244	M_HDL_TG_pct	-0.007	0.021	0.99 (0.95-1.03)	7.34E-01	8.50E-01	NA
245	S_HDL_PL_pct	-0.020	0.020	0.98 (0.94-1.02)	3.32E-01	5.32E-01	NA
246	S_HDL_C_pct	0.016	0.019	1.02 (0.98-1.06)	4.16E-01	6.08E-01	NA
247	S_HDL_CE_pct	-0.003	0.019	1 (0.96-1.03)	8.54E-01	9.37E-01	NA
248	S_HDL_FC_pct	0.077	0.020	1.08 (1.04-1.12)	1.31E-04	9.36E-03	NA
249	S_HDL_TG_pct	-0.006	0.021	0.99 (0.95-1.04)	7.74E-01	8.72E-01	NA

* Percentage change in beta from previous model.

SE, standard error. HR, hazard ratio. CI, confidence intervals. FDR, false discovery rate. NA, not applicable.

Abbreviations are in line with **Supplementary Table 3.1**.

Supplementary Table 4.5 Association of individual metabolites with incident ischaemic stroke in the lipid modifying agent nonusers and users in model 7

		Lipid modifying agent nonusers (n = 158,429)					Lipid modifying agent users (n = 26,135)				
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**
1	Total_C	0.003	0.022	1 (0.96-1.05)	9.09E-01	9.77E-01	0.019	0.045	1.02 (0.93-1.11)	6.74E-01	8.64E-01
2	non_HDL_C	0.007	0.022	1.01 (0.96-1.05)	7.62E-01	9.04E-01	0.038	0.043	1.04 (0.95-1.13)	3.81E-01	8.05E-01
3	Remnant_C	0.005	0.022	1.01 (0.96-1.05)	8.09E-01	9.33E-01	0.053	0.043	1.05 (0.97-1.15)	2.21E-01	6.48E-01
4	VLDL_C	-0.015	0.023	0.99 (0.94-1.03)	5.21E-01	7.19E-01	0.031	0.041	1.03 (0.95-1.12)	4.52E-01	8.05E-01
5	Clinical_LDL_C	0.014	0.022	1.01 (0.97-1.06)	5.13E-01	7.19E-01	0.037	0.043	1.04 (0.95-1.13)	3.93E-01	8.05E-01
6	LDL_C	0.007	0.022	1.01 (0.97-1.05)	7.39E-01	8.94E-01	0.023	0.043	1.02 (0.94-1.11)	5.96E-01	8.05E-01
7	HDL_C	-0.015	0.025	0.99 (0.94-1.04)	5.51E-01	7.46E-01	-0.065	0.048	0.94 (0.85-1.03)	1.78E-01	5.58E-01
8	Total_TG	-0.047	0.023	0.95 (0.91-1)	3.82E-02	1.94E-01	-0.024	0.041	0.98 (0.9-1.06)	5.59E-01	8.05E-01
9	VLDL_TG	-0.055	0.023	0.95 (0.91-0.99)	1.79E-02	1.44E-01	-0.034	0.042	0.97 (0.89-1.05)	4.18E-01	8.05E-01
10	LDL_TG	-0.003	0.022	1 (0.96-1.04)	8.88E-01	9.65E-01	0.029	0.041	1.03 (0.95-1.12)	4.70E-01	8.05E-01
11	HDL_TG	-0.035	0.022	0.97 (0.93-1.01)	1.20E-01	3.15E-01	-0.023	0.042	0.98 (0.9-1.06)	5.84E-01	8.05E-01
12	Total_PL	-0.023	0.023	0.98 (0.93-1.02)	3.18E-01	5.65E-01	-0.010	0.045	0.99 (0.91-1.08)	8.32E-01	9.09E-01
13	VLDL_PL	-0.031	0.023	0.97 (0.93-1.01)	1.81E-01	4.11E-01	0.007	0.041	1.01 (0.93-1.09)	8.69E-01	9.41E-01
14	LDL_PL	0.010	0.022	1.01 (0.97-1.06)	6.46E-01	8.13E-01	0.036	0.043	1.04 (0.95-1.13)	4.07E-01	8.05E-01
15	HDL_PL	-0.032	0.025	0.97 (0.92-1.02)	1.96E-01	4.33E-01	-0.067	0.047	0.94 (0.85-1.02)	1.48E-01	4.92E-01
16	Total_CE	0.004	0.022	1 (0.96-1.05)	8.62E-01	9.59E-01	0.014	0.046	1.01 (0.93-1.11)	7.60E-01	8.80E-01
17	VLDL_CE	-0.005	0.022	1 (0.95-1.04)	8.32E-01	9.50E-01	0.049	0.042	1.05 (0.97-1.14)	2.44E-01	6.75E-01
18	LDL_CE	0.002	0.022	1 (0.96-1.05)	9.23E-01	9.80E-01	0.018	0.043	1.02 (0.94-1.11)	6.68E-01	8.62E-01
19	HDL_CE	-0.017	0.025	0.98 (0.94-1.03)	4.89E-01	6.96E-01	-0.073	0.048	0.93 (0.85-1.02)	1.25E-01	4.63E-01
20	Total_FC	-0.001	0.022	1 (0.96-1.04)	9.73E-01	9.95E-01	0.031	0.045	1.03 (0.95-1.13)	4.87E-01	8.05E-01
21	VLDL_FC	-0.027	0.023	0.97 (0.93-1.02)	2.39E-01	4.77E-01	0.009	0.041	1.01 (0.93-1.09)	8.29E-01	9.09E-01
22	LDL_FC	0.021	0.022	1.02 (0.98-1.07)	3.33E-01	5.85E-01	0.034	0.044	1.04 (0.95-1.13)	4.31E-01	8.05E-01
23	HDL_FC	-0.005	0.026	1 (0.95-1.05)	8.48E-01	9.58E-01	-0.030	0.048	0.97 (0.88-1.07)	5.32E-01	8.05E-01
24	Total_L	-0.022	0.022	0.98 (0.94-1.02)	3.34E-01	5.85E-01	-0.003	0.043	1 (0.92-1.09)	9.43E-01	9.79E-01
25	VLDL_L	-0.040	0.023	0.96 (0.92-1.01)	7.89E-02	2.45E-01	-0.011	0.041	0.99 (0.91-1.07)	7.88E-01	9.00E-01
26	LDL_L	0.007	0.022	1.01 (0.97-1.05)	7.36E-01	8.94E-01	0.027	0.043	1.03 (0.94-1.12)	5.32E-01	8.05E-01
27	HDL_L	-0.028	0.025	0.97 (0.93-1.02)	2.70E-01	5.10E-01	-0.068	0.047	0.93 (0.85-1.03)	1.48E-01	4.92E-01
28	Total_P	-0.041	0.024	0.96 (0.92-1.01)	8.48E-02	2.52E-01	-0.093	0.046	0.91 (0.83-1)	4.36E-02	2.93E-01
29	VLDL_P	-0.018	0.023	0.98 (0.94-1.03)	4.26E-01	6.67E-01	0.027	0.041	1.03 (0.95-1.11)	5.11E-01	8.05E-01
30	LDL_P	0.006	0.022	1.01 (0.96-1.05)	7.68E-01	9.06E-01	0.053	0.042	1.05 (0.97-1.15)	2.09E-01	6.26E-01
31	HDL_P	-0.045	0.024	0.96 (0.91-1)	6.28E-02	2.19E-01	-0.107	0.046	0.9 (0.82-0.98)	1.90E-02	2.47E-01
32	VLDL_size	-0.076	0.024	0.93 (0.88-0.97)	1.38E-03	1.12E-01	-0.092	0.042	0.91 (0.84-0.99)	2.92E-02	2.50E-01

		Lipid modifying agent nonusers (n = 158,429)					Lipid modifying agent users (n = 26,135)				
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**
33	LDL_size	0.016	0.021	1.02 (0.97-1.06)	4.58E-01	6.82E-01	0.084	0.043	1.09 (1-1.18)	4.91E-02	3.14E-01
34	HDL_size	0.022	0.026	1.02 (0.97-1.08)	3.99E-01	6.40E-01	0.019	0.048	1.02 (0.93-1.12)	6.86E-01	8.71E-01
35	Phosphoglyc	-0.030	0.023	0.97 (0.93-1.02)	2.01E-01	4.33E-01	-0.032	0.045	0.97 (0.89-1.06)	4.83E-01	8.05E-01
36	TG_by_PG	-0.043	0.024	0.96 (0.91-1)	7.40E-02	2.33E-01	-0.015	0.044	0.99 (0.9-1.08)	7.40E-01	8.78E-01
37	Cholines	-0.027	0.024	0.97 (0.93-1.02)	2.47E-01	4.85E-01	-0.026	0.046	0.98 (0.89-1.07)	5.76E-01	8.05E-01
38	Phosphatidylc	-0.030	0.024	0.97 (0.93-1.02)	2.02E-01	4.33E-01	-0.020	0.045	0.98 (0.9-1.07)	6.56E-01	8.51E-01
39	Sphingomyelins	0.032	0.024	1.03 (0.99-1.08)	1.79E-01	4.09E-01	0.029	0.047	1.03 (0.94-1.13)	5.35E-01	8.05E-01
40	ApoB	0.007	0.022	1.01 (0.97-1.05)	7.48E-01	8.96E-01	0.056	0.042	1.06 (0.97-1.15)	1.80E-01	5.58E-01
41	ApoA1	-0.037	0.025	0.96 (0.92-1.01)	1.32E-01	3.36E-01	-0.091	0.047	0.91 (0.83-1)	5.06E-02	3.15E-01
42	ApoB_by_ApoA1	0.025	0.023	1.03 (0.98-1.07)	2.71E-01	5.10E-01	0.091	0.041	1.1 (1.01-1.19)	2.62E-02	2.47E-01
43	Total_FA	-0.032	0.022	0.97 (0.93-1.01)	1.49E-01	3.58E-01	-0.022	0.041	0.98 (0.9-1.06)	5.99E-01	8.05E-01
44	Unsaturation	-0.002	0.024	1 (0.95-1.05)	9.34E-01	9.80E-01	-0.043	0.045	0.96 (0.88-1.05)	3.49E-01	7.97E-01
45	Omega_3	-0.034	0.023	0.97 (0.92-1.01)	1.44E-01	3.54E-01	-0.074	0.043	0.93 (0.85-1.01)	8.84E-02	3.89E-01
46	Omega_6	-0.040	0.022	0.96 (0.92-1)	7.09E-02	2.33E-01	-0.010	0.044	0.99 (0.91-1.08)	8.22E-01	9.09E-01
47	PUFA	-0.038	0.022	0.96 (0.92-1.01)	8.48E-02	2.52E-01	-0.026	0.044	0.97 (0.9-1.06)	5.54E-01	8.05E-01
48	MUFA	-0.024	0.023	0.98 (0.93-1.02)	2.82E-01	5.24E-01	-0.015	0.041	0.99 (0.91-1.07)	7.13E-01	8.78E-01
49	SFA	-0.028	0.022	0.97 (0.93-1.02)	2.19E-01	4.50E-01	-0.022	0.041	0.98 (0.9-1.06)	5.90E-01	8.05E-01
50	LA	-0.046	0.022	0.96 (0.92-1)	3.28E-02	1.85E-01	0.004	0.043	1 (0.92-1.09)	9.23E-01	9.72E-01
51	DHA	0.000	0.024	1 (0.96-1.05)	9.97E-01	9.97E-01	-0.048	0.045	0.95 (0.87-1.04)	2.94E-01	7.28E-01
52	Omega_3_pct	-0.023	0.024	0.98 (0.93-1.03)	3.39E-01	5.91E-01	-0.083	0.045	0.92 (0.84-1.01)	6.68E-02	3.54E-01
53	Omega_6_pct	0.002	0.022	1 (0.96-1.05)	9.26E-01	9.80E-01	0.029	0.042	1.03 (0.95-1.12)	4.88E-01	8.05E-01
54	PUFA_pct	0.002	0.022	1 (0.96-1.05)	9.37E-01	9.80E-01	0.005	0.043	1.01 (0.92-1.09)	9.04E-01	9.62E-01
55	MUFA_pct	-0.002	0.024	1 (0.95-1.05)	9.42E-01	9.81E-01	0.002	0.045	1 (0.92-1.09)	9.67E-01	9.83E-01
56	SFA_pct	0.003	0.022	1 (0.96-1.05)	8.82E-01	9.65E-01	-0.014	0.042	0.99 (0.91-1.07)	7.48E-01	8.78E-01
57	LA_pct	-0.026	0.022	0.98 (0.93-1.02)	2.39E-01	4.77E-01	0.044	0.043	1.05 (0.96-1.14)	3.15E-01	7.40E-01
58	DHA_pct	0.021	0.024	1.02 (0.97-1.07)	3.86E-01	6.23E-01	-0.033	0.045	0.97 (0.89-1.06)	4.65E-01	8.05E-01
59	PUFA_by_MUFA	0.002	0.023	1 (0.96-1.05)	9.37E-01	9.80E-01	0.002	0.044	1 (0.92-1.09)	9.71E-01	9.83E-01
60	Omega_6_by_Omega_3	0.021	0.024	1.02 (0.98-1.07)	3.67E-01	6.11E-01	0.081	0.043	1.08 (1-1.18)	6.18E-02	3.54E-01
61	Ala	0.031	0.027	1.03 (0.98-1.09)	2.55E-01	4.96E-01	-0.052	0.051	0.95 (0.86-1.05)	3.06E-01	7.40E-01
62	Gln	0.008	0.022	1.01 (0.97-1.05)	7.31E-01	8.94E-01	-0.023	0.041	0.98 (0.9-1.06)	5.80E-01	8.05E-01
63	Gly	-0.007	0.019	0.99 (0.96-1.03)	7.09E-01	8.74E-01	0.013	0.043	1.01 (0.93-1.1)	7.60E-01	8.80E-01
64	His	-0.048	0.021	0.95 (0.91-0.99)	2.15E-02	1.53E-01	-0.027	0.040	0.97 (0.9-1.05)	4.90E-01	8.05E-01
65	Total_BCAA	-0.039	0.023	0.96 (0.92-1.01)	9.48E-02	2.74E-01	-0.023	0.043	0.98 (0.9-1.06)	5.93E-01	8.05E-01
66	Ile	-0.020	0.022	0.98 (0.94-1.02)	3.76E-01	6.19E-01	-0.003	0.041	1 (0.92-1.08)	9.38E-01	9.77E-01
67	Leu	-0.036	0.023	0.97 (0.92-1.01)	1.19E-01	3.15E-01	-0.025	0.042	0.98 (0.9-1.06)	5.56E-01	8.05E-01

		Lipid modifying agent nonusers (n = 158,429)					Lipid modifying agent users (n = 26,135)				
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**
68	Val	-0.045	0.024	0.96 (0.91-1)	5.40E-02	2.17E-01	-0.027	0.044	0.97 (0.89-1.06)	5.30E-01	8.05E-01
69	Phe	0.048	0.022	1.05 (1.01-1.1)	2.71E-02	1.65E-01	0.069	0.041	1.07 (0.99-1.16)	8.85E-02	3.89E-01
70	Tyr	0.044	0.022	1.05 (1-1.09)	4.88E-02	2.10E-01	-0.027	0.041	0.97 (0.9-1.06)	5.16E-01	8.05E-01
71	Glucose	-0.015	0.022	0.99 (0.94-1.03)	4.79E-01	6.96E-01	-0.014	0.042	0.99 (0.91-1.07)	7.45E-01	8.78E-01
72	Lactate	0.058	0.023	1.06 (1.01-1.11)	1.12E-02	1.25E-01	-0.026	0.042	0.98 (0.9-1.06)	5.47E-01	8.05E-01
73	Pyruvate	0.020	0.022	1.02 (0.98-1.07)	3.79E-01	6.21E-01	0.022	0.042	1.02 (0.94-1.11)	6.08E-01	8.05E-01
74	Citrate	-0.001	0.023	1 (0.95-1.05)	9.53E-01	9.81E-01	-0.046	0.044	0.96 (0.88-1.04)	2.95E-01	7.28E-01
75	bOHbutyrate	0.003	0.024	1 (0.96-1.05)	9.11E-01	9.77E-01	-0.020	0.042	0.98 (0.9-1.06)	6.33E-01	8.29E-01
76	Acetate	-0.023	0.019	0.98 (0.94-1.01)	2.10E-01	4.40E-01	0.026	0.042	1.03 (0.95-1.11)	5.42E-01	8.05E-01
77	Acetoacetate	-0.022	0.022	0.98 (0.94-1.02)	3.17E-01	5.65E-01	-0.002	0.040	1 (0.92-1.08)	9.69E-01	9.83E-01
78	Acetone	0.012	0.021	1.01 (0.97-1.06)	5.72E-01	7.46E-01	-0.045	0.041	0.96 (0.88-1.04)	2.82E-01	7.10E-01
79	Creatinine	0.023	0.026	1.02 (0.97-1.08)	3.59E-01	6.08E-01	0.076	0.046	1.08 (0.99-1.18)	1.02E-01	4.22E-01
80	Albumin	-0.124	0.021	0.88 (0.85-0.92)	5.99E-09	1.49E-06	-0.177	0.039	0.84 (0.78-0.91)	6.82E-06	1.70E-03
81	GlycA	0.043	0.024	1.04 (1-1.09)	7.21E-02	2.33E-01	0.038	0.044	1.04 (0.95-1.13)	3.81E-01	8.05E-01
82	XXL_VLDL_P	-0.056	0.024	0.95 (0.9-0.99)	1.78E-02	1.44E-01	-0.030	0.043	0.97 (0.89-1.06)	4.86E-01	8.05E-01
83	XXL_VLDL_L	-0.063	0.023	0.94 (0.9-0.98)	6.64E-03	1.12E-01	-0.039	0.042	0.96 (0.89-1.05)	3.60E-01	8.05E-01
84	XXL_VLDL_PL	-0.054	0.024	0.95 (0.9-0.99)	2.32E-02	1.56E-01	-0.035	0.045	0.97 (0.89-1.05)	4.38E-01	8.05E-01
85	XXL_VLDL_C	-0.053	0.024	0.95 (0.91-0.99)	2.64E-02	1.65E-01	-0.040	0.043	0.96 (0.88-1.04)	3.43E-01	7.92E-01
86	XXL_VLDL_CE	-0.040	0.024	0.96 (0.92-1.01)	9.42E-02	2.74E-01	-0.036	0.043	0.97 (0.89-1.05)	3.98E-01	8.05E-01
87	XXL_VLDL_FC	-0.067	0.023	0.94 (0.89-0.98)	3.77E-03	1.12E-01	-0.043	0.043	0.96 (0.88-1.04)	3.15E-01	7.40E-01
88	XXL_VLDL_TG	-0.060	0.022	0.94 (0.9-0.98)	5.43E-03	1.12E-01	-0.025	0.040	0.98 (0.9-1.05)	5.29E-01	8.05E-01
89	XL_VLDL_P	-0.049	0.024	0.95 (0.91-1)	4.29E-02	2.02E-01	-0.030	0.043	0.97 (0.89-1.06)	4.77E-01	8.05E-01
90	XL_VLDL_L	-0.050	0.024	0.95 (0.91-1)	3.57E-02	1.85E-01	-0.032	0.042	0.97 (0.89-1.05)	4.51E-01	8.05E-01
91	XL_VLDL_PL	-0.047	0.024	0.95 (0.91-1)	5.52E-02	2.18E-01	-0.026	0.042	0.97 (0.9-1.06)	5.38E-01	8.05E-01
92	XL_VLDL_C	-0.038	0.024	0.96 (0.92-1.01)	1.15E-01	3.11E-01	-0.014	0.042	0.99 (0.91-1.07)	7.37E-01	8.78E-01
93	XL_VLDL_CE	-0.029	0.024	0.97 (0.93-1.02)	2.22E-01	4.54E-01	-0.003	0.042	1 (0.92-1.08)	9.48E-01	9.79E-01
94	XL_VLDL_FC	-0.046	0.024	0.96 (0.91-1)	5.60E-02	2.18E-01	-0.023	0.042	0.98 (0.9-1.06)	5.77E-01	8.05E-01
95	XL_VLDL_TG	-0.056	0.024	0.95 (0.9-0.99)	1.87E-02	1.45E-01	-0.038	0.043	0.96 (0.89-1.05)	3.82E-01	8.05E-01
96	L_VLDL_P	-0.047	0.024	0.95 (0.91-1)	4.48E-02	2.04E-01	-0.028	0.042	0.97 (0.9-1.06)	4.98E-01	8.05E-01
97	L_VLDL_L	-0.050	0.024	0.95 (0.91-1)	3.37E-02	1.85E-01	-0.030	0.042	0.97 (0.89-1.05)	4.69E-01	8.05E-01
98	L_VLDL_PL	-0.047	0.024	0.95 (0.91-1)	5.05E-02	2.10E-01	-0.023	0.042	0.98 (0.9-1.06)	5.84E-01	8.05E-01
99	L_VLDL_C	-0.041	0.024	0.96 (0.92-1.01)	8.36E-02	2.52E-01	-0.015	0.042	0.99 (0.91-1.07)	7.10E-01	8.78E-01
100	L_VLDL_CE	-0.036	0.023	0.97 (0.92-1.01)	1.24E-01	3.22E-01	-0.003	0.041	1 (0.92-1.08)	9.36E-01	9.77E-01
10	L_VLDL_FC	-0.045	0.024	0.96 (0.91-1)	6.06E-02	2.18E-01	-0.026	0.042	0.97 (0.9-1.06)	5.35E-01	8.05E-01

		Lipid modifying agent nonusers (n = 158,429)					Lipid modifying agent users (n = 26,135)				
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**
1											
102	L_VLDL_TG	-0.055	0.023	0.95 (0.91-0.99)	1.76E-02	1.44E-01	-0.037	0.042	0.96 (0.89-1.05)	3.70E-01	8.05E-01
103	M_VLDL_P	-0.023	0.023	0.98 (0.94-1.02)	2.99E-01	5.51E-01	0.007	0.041	1.01 (0.93-1.09)	8.58E-01	9.33E-01
104	M_VLDL_L	-0.031	0.023	0.97 (0.93-1.01)	1.73E-01	4.04E-01	-0.005	0.041	1 (0.92-1.08)	8.97E-01	9.59E-01
105	M_VLDL_PL	-0.018	0.022	0.98 (0.94-1.03)	4.23E-01	6.66E-01	0.015	0.041	1.02 (0.94-1.1)	7.16E-01	8.78E-01
106	M_VLDL_C	0.006	0.022	1.01 (0.96-1.05)	7.99E-01	9.30E-01	0.051	0.042	1.05 (0.97-1.14)	2.28E-01	6.52E-01
107	M_VLDL_CE	0.020	0.021	1.02 (0.98-1.06)	3.43E-01	5.93E-01	0.074	0.043	1.08 (0.99-1.17)	8.62E-02	3.89E-01
108	M_VLDL_FC	-0.011	0.022	0.99 (0.95-1.03)	6.38E-01	8.13E-01	0.025	0.041	1.03 (0.95-1.11)	5.46E-01	8.05E-01
109	M_VLDL_TG	-0.050	0.023	0.95 (0.91-0.99)	2.73E-02	1.65E-01	-0.033	0.041	0.97 (0.89-1.05)	4.23E-01	8.05E-01
110	S_VLDL_P	-0.027	0.023	0.97 (0.93-1.02)	2.37E-01	4.77E-01	0.012	0.041	1.01 (0.93-1.1)	7.78E-01	8.93E-01
111	S_VLDL_L	-0.025	0.023	0.98 (0.93-1.02)	2.66E-01	5.09E-01	0.013	0.041	1.01 (0.94-1.1)	7.52E-01	8.79E-01
112	S_VLDL_PL	-0.013	0.023	0.99 (0.94-1.03)	5.63E-01	7.46E-01	0.029	0.041	1.03 (0.95-1.12)	4.81E-01	8.05E-01
113	S_VLDL_C	-0.003	0.023	1 (0.95-1.04)	8.91E-01	9.65E-01	0.046	0.041	1.05 (0.97-1.14)	2.63E-01	6.92E-01
114	S_VLDL_CE	-0.004	0.023	1 (0.95-1.04)	8.70E-01	9.61E-01	0.049	0.041	1.05 (0.97-1.14)	2.38E-01	6.66E-01
115	S_VLDL_FC	-0.002	0.022	1 (0.96-1.04)	9.27E-01	9.80E-01	0.041	0.042	1.04 (0.96-1.13)	3.20E-01	7.44E-01
116	S_VLDL_TG	-0.044	0.023	0.96 (0.92-1)	5.03E-02	2.10E-01	-0.017	0.042	0.98 (0.91-1.07)	6.77E-01	8.64E-01
117	XS_VLDL_P	0.020	0.022	1.02 (0.98-1.07)	3.66E-01	6.11E-01	0.096	0.042	1.1 (1.01-1.2)	2.23E-02	2.47E-01
118	XS_VLDL_L	0.017	0.022	1.02 (0.97-1.06)	4.59E-01	6.82E-01	0.101	0.042	1.11 (1.02-1.2)	1.65E-02	2.47E-01
119	XS_VLDL_PL	0.017	0.022	1.02 (0.97-1.06)	4.52E-01	6.82E-01	0.107	0.042	1.11 (1.03-1.21)	1.02E-02	2.17E-01
120	XS_VLDL_C	0.030	0.022	1.03 (0.99-1.08)	1.74E-01	4.04E-01	0.120	0.044	1.13 (1.04-1.23)	6.01E-03	2.17E-01
121	XS_VLDL_CE	0.033	0.022	1.03 (0.99-1.08)	1.34E-01	3.36E-01	0.124	0.044	1.13 (1.04-1.24)	5.25E-03	2.17E-01
122	XS_VLDL_FC	0.022	0.022	1.02 (0.98-1.07)	3.16E-01	5.65E-01	0.109	0.042	1.12 (1.03-1.21)	1.04E-02	2.17E-01
123	XS_VLDL_TG	-0.018	0.022	0.98 (0.94-1.03)	4.17E-01	6.62E-01	0.033	0.041	1.03 (0.95-1.12)	4.22E-01	8.05E-01

		Lipid modifying agent nonusers (n = 158,429)					Lipid modifying agent users (n = 26,135)				
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**
124	IDL_P	0.022	0.022	1.02 (0.98-1.07)	3.15E-01	5.65E-01	0.079	0.043	1.08 (1-1.18)	6.68E-02	3.54E-01
125	IDL_L	0.029	0.022	1.03 (0.99-1.08)	1.99E-01	4.33E-01	0.076	0.045	1.08 (0.99-1.18)	9.07E-02	3.89E-01
126	IDL_PL	0.031	0.022	1.03 (0.99-1.08)	1.59E-01	3.76E-01	0.092	0.045	1.1 (1-1.2)	4.00E-02	2.85E-01
127	IDL_C	0.028	0.022	1.03 (0.99-1.07)	1.99E-01	4.33E-01	0.068	0.045	1.07 (0.98-1.17)	1.32E-01	4.63E-01
128	IDL_CE	0.026	0.022	1.03 (0.98-1.07)	2.42E-01	4.79E-01	0.064	0.045	1.07 (0.98-1.17)	1.60E-01	5.16E-01
129	IDL_FC	0.034	0.022	1.04 (0.99-1.08)	1.20E-01	3.15E-01	0.078	0.045	1.08 (0.99-1.18)	8.09E-02	3.89E-01
130	IDL_TG	0.004	0.022	1 (0.96-1.05)	8.50E-01	9.58E-01	0.059	0.041	1.06 (0.98-1.15)	1.53E-01	5.01E-01
131	L_LDL_P	0.009	0.021	1.01 (0.97-1.05)	6.69E-01	8.38E-01	0.064	0.042	1.07 (0.98-1.16)	1.26E-01	4.63E-01
132	L_LDL_L	0.014	0.022	1.01 (0.97-1.06)	5.23E-01	7.19E-01	0.032	0.044	1.03 (0.95-1.13)	4.67E-01	8.05E-01
133	L_LDL_PL	0.014	0.022	1.01 (0.97-1.06)	5.30E-01	7.21E-01	0.038	0.044	1.04 (0.95-1.13)	3.84E-01	8.05E-01
134	L_LDL_C	0.014	0.022	1.01 (0.97-1.06)	5.18E-01	7.19E-01	0.027	0.044	1.03 (0.94-1.12)	5.32E-01	8.05E-01
135	L_LDL_CE	0.010	0.022	1.01 (0.97-1.06)	6.40E-01	8.13E-01	0.024	0.044	1.02 (0.94-1.12)	5.80E-01	8.05E-01
136	L_LDL_FC	0.024	0.022	1.03 (0.98-1.07)	2.61E-01	5.04E-01	0.036	0.044	1.04 (0.95-1.13)	4.18E-01	8.05E-01
137	L_LDL_TG	0.007	0.022	1.01 (0.97-1.05)	7.36E-01	8.94E-01	0.045	0.041	1.05 (0.97-1.13)	2.68E-01	6.92E-01
138	M_LDL_P	0.000	0.022	1 (0.96-1.04)	9.91E-01	9.95E-01	0.027	0.042	1.03 (0.95-1.12)	5.26E-01	8.05E-01
139	M_LDL_L	-0.005	0.022	1 (0.95-1.04)	8.26E-01	9.47E-01	0.011	0.042	1.01 (0.93-1.1)	7.94E-01	9.03E-01
140	M_LDL_PL	0.000	0.022	1 (0.96-1.05)	9.85E-01	9.95E-01	0.015	0.042	1.02 (0.94-1.1)	7.19E-01	8.78E-01
14	M_LDL_C	-0.006	0.022	0.99 (0.95-1.04)	7.91E-01	9.24E-01	0.009	0.042	1.01 (0.93-1.1)	8.26E-01	9.09E-01

		Lipid modifying agent nonusers (n = 158,429)					Lipid modifying agent users (n = 26,135)				
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**
1											
142	M_LDL_CE	-0.013	0.023	0.99 (0.95-1.03)	5.72E-01	7.46E-01	0.004	0.042	1 (0.93-1.09)	9.18E-01	9.72E-01
143	M_LDL_FC	0.013	0.022	1.01 (0.97-1.06)	5.56E-01	7.46E-01	0.022	0.043	1.02 (0.94-1.11)	6.04E-01	8.05E-01
144	M_LDL_TG	-0.013	0.022	0.99 (0.95-1.03)	5.72E-01	7.46E-01	0.010	0.041	1.01 (0.93-1.09)	8.03E-01	9.09E-01
145	S_LDL_P	0.004	0.022	1 (0.96-1.05)	8.59E-01	9.59E-01	0.037	0.042	1.04 (0.96-1.13)	3.68E-01	8.05E-01
146	S_LDL_L	0.001	0.022	1 (0.96-1.05)	9.53E-01	9.81E-01	0.036	0.042	1.04 (0.96-1.12)	3.94E-01	8.05E-01
147	S_LDL_PL	0.015	0.022	1.02 (0.97-1.06)	4.78E-01	6.96E-01	0.063	0.041	1.07 (0.98-1.16)	1.26E-01	4.63E-01
148	S_LDL_C	0.000	0.022	1 (0.96-1.04)	9.87E-01	9.95E-01	0.027	0.042	1.03 (0.95-1.12)	5.21E-01	8.05E-01
149	S_LDL_CE	-0.008	0.022	0.99 (0.95-1.04)	7.05E-01	8.73E-01	0.017	0.042	1.02 (0.94-1.1)	6.92E-01	8.75E-01
150	S_LDL_FC	0.020	0.021	1.02 (0.98-1.06)	3.53E-01	6.02E-01	0.052	0.042	1.05 (0.97-1.14)	2.23E-01	6.48E-01
151	S_LDL_TG	-0.035	0.022	0.97 (0.92-1.01)	1.13E-01	3.10E-01	-0.009	0.041	0.99 (0.92-1.07)	8.22E-01	9.09E-01
152	XL_HDL_P	0.041	0.025	1.04 (0.99-1.09)	9.95E-02	2.85E-01	0.082	0.047	1.09 (0.99-1.19)	7.86E-02	3.89E-01
153	XL_HDL_L	0.045	0.025	1.05 (1-1.1)	7.36E-02	2.33E-01	0.100	0.046	1.11 (1.01-1.21)	3.06E-02	2.50E-01
154	XL_HDL_PL	0.040	0.025	1.04 (0.99-1.09)	1.09E-01	3.06E-01	0.101	0.046	1.11 (1.01-1.21)	2.68E-02	2.47E-01
155	XL_HDL_C	0.051	0.025	1.05 (1-1.11)	3.97E-02	1.98E-01	0.100	0.046	1.11 (1.01-1.21)	3.12E-02	2.50E-01
156	XL_HDL_CE	0.045	0.025	1.05 (1-1.1)	7.10E-02	2.33E-01	0.086	0.047	1.09 (1-1.19)	6.47E-02	3.54E-01
157	XL_HDL_FC	0.067	0.024	1.07 (1.02-1.12)	4.94E-03	1.12E-01	0.134	0.045	1.14 (1.05-1.25)	2.89E-03	2.17E-01
158	XL_HDL_TG	-0.013	0.022	0.99 (0.95-1.03)	5.59E-01	7.46E-01	0.010	0.041	1.01 (0.93-1.09)	8.07E-01	9.09E-01

		Lipid modifying agent nonusers (n = 158,429)					Lipid modifying agent users (n = 26,135)				
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**
159	L_HDL_P	0.008	0.026	1.01 (0.96-1.06)	7.56E-01	9.00E-01	-0.008	0.047	0.99 (0.91-1.09)	8.73E-01	9.41E-01
160	L_HDL_L	0.008	0.026	1.01 (0.96-1.06)	7.45E-01	8.96E-01	-0.006	0.048	0.99 (0.91-1.09)	8.97E-01	9.59E-01
161	L_HDL_PL	0.004	0.026	1 (0.96-1.06)	8.62E-01	9.59E-01	-0.013	0.047	0.99 (0.9-1.08)	7.77E-01	8.93E-01
162	L_HDL_C	0.015	0.025	1.02 (0.97-1.07)	5.67E-01	7.46E-01	0.004	0.047	1 (0.92-1.1)	9.25E-01	9.72E-01
163	L_HDL_CE	0.013	0.025	1.01 (0.96-1.07)	6.02E-01	7.81E-01	-0.002	0.047	1 (0.91-1.09)	9.71E-01	9.83E-01
164	L_HDL_FC	0.019	0.025	1.02 (0.97-1.07)	4.46E-01	6.81E-01	0.026	0.047	1.03 (0.94-1.13)	5.77E-01	8.05E-01
165	L_HDL_TG	-0.016	0.023	0.98 (0.94-1.03)	4.86E-01	6.96E-01	-0.023	0.042	0.98 (0.9-1.06)	5.91E-01	8.05E-01
166	M_HDL_P	-0.043	0.024	0.96 (0.91-1.01)	7.99E-02	2.46E-01	-0.106	0.045	0.9 (0.82-0.98)	1.96E-02	2.47E-01
167	M_HDL_L	-0.045	0.024	0.96 (0.91-1)	6.06E-02	2.18E-01	-0.100	0.045	0.91 (0.83-0.99)	2.63E-02	2.47E-01
168	M_HDL_PL	-0.049	0.024	0.95 (0.91-1)	4.16E-02	1.99E-01	-0.095	0.045	0.91 (0.83-0.99)	3.40E-02	2.57E-01
169	M_HDL_C	-0.035	0.024	0.97 (0.92-1.01)	1.47E-01	3.54E-01	-0.103	0.045	0.9 (0.83-0.99)	2.29E-02	2.47E-01
170	M_HDL_CE	-0.036	0.024	0.97 (0.92-1.01)	1.38E-01	3.43E-01	-0.107	0.045	0.9 (0.82-0.98)	1.82E-02	2.47E-01
171	M_HDL_FC	-0.031	0.025	0.97 (0.92-1.02)	2.09E-01	4.40E-01	-0.086	0.046	0.92 (0.84-1)	5.98E-02	3.54E-01
172	M_HDL_TG	-0.041	0.022	0.96 (0.92-1)	6.63E-02	2.24E-01	-0.030	0.042	0.97 (0.89-1.05)	4.75E-01	8.05E-01
173	S_HDL_P	-0.057	0.023	0.95 (0.9-0.99)	1.35E-02	1.29E-01	-0.118	0.043	0.89 (0.82-0.97)	6.25E-03	2.17E-01
174	S_HDL_L	-0.065	0.023	0.94 (0.9-0.98)	4.93E-03	1.12E-01	-0.111	0.043	0.9 (0.82-0.97)	1.00E-02	2.17E-01
175	S_HDL_PL	-0.067	0.023	0.94 (0.89-0.98)	3.69E-03	1.12E-01	-0.108	0.043	0.9 (0.83-0.98)	1.28E-02	2.46E-01
17	S_HDL_C	-0.052	0.023	0.95 (0.91-0.99)	2.21E-02	1.53E-01	-0.118	0.043	0.89 (0.82-0.97)	6.16E-03	2.17E-01

		Lipid modifying agent nonusers (n = 158,429)					Lipid modifying agent users (n = 26,135)				
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**
6											
177	S_HDL_CE	-0.054	0.023	0.95 (0.91-0.99)	1.68E-02	1.44E-01	-0.128	0.043	0.88 (0.81-0.96)	2.77E-03	2.17E-01
178	S_HDL_FC	-0.037	0.023	0.96 (0.92-1.01)	1.06E-01	3.00E-01	-0.075	0.044	0.93 (0.85-1.01)	8.58E-02	3.89E-01
179	S_HDL_TG	-0.043	0.023	0.96 (0.92-1)	6.67E-02	2.24E-01	-0.022	0.044	0.98 (0.9-1.07)	6.14E-01	8.09E-01
180	XXL_VLDL_PL_pct	-0.018	0.025	0.98 (0.94-1.03)	4.74E-01	6.96E-01	-0.001	0.049	1 (0.91-1.1)	9.79E-01	9.85E-01
181	XXL_VLDL_C_pct	0.066	0.023	1.07 (1.02-1.12)	3.31E-03	1.12E-01	0.023	0.042	1.02 (0.94-1.11)	5.85E-01	8.05E-01
182	XXL_VLDL_CE_pct	0.082	0.023	1.09 (1.04-1.14)	4.16E-04	5.18E-02	0.024	0.043	1.02 (0.94-1.12)	5.78E-01	8.05E-01
183	XXL_VLDL_FC_pct	0.024	0.023	1.02 (0.98-1.07)	3.07E-01	5.62E-01	0.019	0.042	1.02 (0.94-1.11)	6.56E-01	8.51E-01
184	XXL_VLDL_TG_pct	-0.037	0.020	0.96 (0.93-1)	6.33E-02	2.19E-01	-0.009	0.040	0.99 (0.92-1.07)	8.27E-01	9.09E-01
185	XL_VLDL_PL_pct	0.004	0.026	1 (0.95-1.06)	8.84E-01	9.65E-01	0.015	0.043	1.02 (0.93-1.1)	7.36E-01	8.78E-01
186	XL_VLDL_C_pct	0.061	0.023	1.06 (1.02-1.11)	8.81E-03	1.12E-01	0.068	0.045	1.07 (0.98-1.17)	1.29E-01	4.63E-01
187	XL_VLDL_CE_pct	0.062	0.023	1.06 (1.02-1.11)	7.25E-03	1.12E-01	0.068	0.044	1.07 (0.98-1.17)	1.23E-01	4.63E-01
188	XL_VLDL_FC_pct	0.044	0.023	1.05 (1-1.09)	5.69E-02	2.18E-01	0.060	0.045	1.06 (0.97-1.16)	1.81E-01	5.58E-01
189	XL_VLDL_TG_pct	-0.060	0.023	0.94 (0.9-0.98)	7.49E-03	1.12E-01	-0.050	0.046	0.95 (0.87-1.04)	2.70E-01	6.92E-01
190	L_VLDL_PL_pct	-0.019	0.025	0.98 (0.93-1.03)	4.60E-01	6.82E-01	0.016	0.046	1.02 (0.93-1.11)	7.25E-01	8.78E-01
191	L_VLDL_C_pct	0.045	0.022	1.05 (1-1.09)	4.16E-02	1.99E-01	0.063	0.042	1.07 (0.98-1.16)	1.38E-01	4.71E-01
192	L_VLDL_CE_pct	0.047	0.022	1.05 (1-1.1)	3.57E-02	1.85E-01	0.074	0.044	1.08 (0.99-1.17)	8.93E-02	3.89E-01
193	L_VLDL_FC_pct	0.030	0.023	1.03 (0.98-1.08)	2.01E-01	4.33E-01	0.028	0.042	1.03 (0.95-1.12)	5.06E-01	8.05E-01

		Lipid modifying agent nonusers (n = 158,429)					Lipid modifying agent users (n = 26,135)				
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**
194	L_VLDL_TG_pct	-0.035	0.022	0.97 (0.93-1.01)	1.12E-01	3.10E-01	-0.046	0.041	0.96 (0.88-1.04)	2.67E-01	6.92E-01
195	M_VLDL_PL_pct	0.057	0.022	1.06 (1.01-1.11)	9.75E-03	1.16E-01	0.098	0.044	1.1 (1.01-1.2)	2.58E-02	2.47E-01
196	M_VLDL_C_pct	0.059	0.022	1.06 (1.02-1.11)	8.20E-03	1.12E-01	0.095	0.044	1.1 (1.01-1.2)	3.11E-02	2.50E-01
197	M_VLDL_CE_pct	0.060	0.022	1.06 (1.02-1.11)	6.86E-03	1.12E-01	0.098	0.044	1.1 (1.01-1.2)	2.67E-02	2.47E-01
198	M_VLDL_FC_pct	0.058	0.022	1.06 (1.02-1.11)	9.00E-03	1.12E-01	0.094	0.044	1.1 (1.01-1.2)	3.33E-02	2.57E-01
199	M_VLDL_TG_pct	-0.058	0.022	0.94 (0.9-0.99)	8.69E-03	1.12E-01	-0.079	0.045	0.92 (0.85-1.01)	8.05E-02	3.89E-01
200	S_VLDL_PL_pct	0.046	0.022	1.05 (1-1.09)	3.56E-02	1.85E-01	0.067	0.044	1.07 (0.98-1.17)	1.32E-01	4.63E-01
201	S_VLDL_C_pct	0.054	0.022	1.06 (1.01-1.1)	1.19E-02	1.25E-01	0.089	0.043	1.09 (1.01-1.19)	3.81E-02	2.79E-01
202	S_VLDL_CE_pct	0.054	0.022	1.06 (1.01-1.1)	1.26E-02	1.25E-01	0.096	0.042	1.1 (1.01-1.2)	2.22E-02	2.47E-01
203	S_VLDL_FC_pct	0.048	0.022	1.05 (1.01-1.1)	2.74E-02	1.65E-01	0.068	0.044	1.07 (0.98-1.17)	1.24E-01	4.63E-01
204	S_VLDL_TG_pct	-0.050	0.022	0.95 (0.91-0.99)	2.11E-02	1.53E-01	-0.068	0.044	0.94 (0.86-1.02)	1.25E-01	4.63E-01
205	XS_VLDL_PL_pct	0.006	0.023	1.01 (0.96-1.05)	8.07E-01	9.33E-01	0.079	0.045	1.08 (0.99-1.18)	8.05E-02	3.89E-01
206	XS_VLDL_C_pct	0.044	0.022	1.05 (1-1.09)	4.51E-02	2.04E-01	0.084	0.045	1.09 (1-1.19)	6.00E-02	3.54E-01
207	XS_VLDL_CE_pct	0.043	0.022	1.04 (1-1.09)	5.17E-02	2.11E-01	0.083	0.045	1.09 (1-1.19)	6.41E-02	3.54E-01
208	XS_VLDL_FC_pct	0.039	0.021	1.04 (1-1.08)	6.12E-02	2.18E-01	0.072	0.041	1.07 (0.99-1.17)	8.21E-02	3.89E-01
209	XS_VLDL_TG_pct	-0.048	0.022	0.95 (0.91-1)	2.78E-02	1.65E-01	-0.090	0.045	0.91 (0.84-1)	4.57E-02	3.00E-01
210	IDL_PL_pct	0.008	0.021	1.01 (0.97-1.05)	6.99E-01	8.70E-01	0.063	0.041	1.07 (0.98-1.15)	1.29E-01	4.63E-01
211	IDL_C_pct	0.013	0.021	1.01 (0.97-1.06)	5.19E-01	7.19E-01	-0.002	0.042	1 (0.92-1.08)	9.58E-01	9.83E-01

		Lipid modifying agent nonusers (n = 158,429)					Lipid modifying agent users (n = 26,135)				
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**
212	IDL_CE_pct	0.000	0.021	1 (0.96-1.04)	9.83E-01	9.95E-01	-0.021	0.042	0.98 (0.9-1.06)	6.08E-01	8.05E-01
213	IDL_FC_pct	0.032	0.021	1.03 (0.99-1.08)	1.31E-01	3.36E-01	0.036	0.040	1.04 (0.96-1.12)	3.67E-01	8.05E-01
214	IDL_TG_pct	-0.020	0.021	0.98 (0.94-1.02)	3.49E-01	5.99E-01	-0.014	0.045	0.99 (0.9-1.08)	7.46E-01	8.78E-01
215	L_LDL_PL_pct	-0.004	0.021	1 (0.96-1.04)	8.42E-01	9.57E-01	0.030	0.039	1.03 (0.95-1.11)	4.42E-01	8.05E-01
216	L_LDL_C_pct	0.010	0.022	1.01 (0.97-1.05)	6.46E-01	8.13E-01	-0.032	0.044	0.97 (0.89-1.06)	4.60E-01	8.05E-01
217	L_LDL_CE_pct	-0.031	0.021	0.97 (0.93-1.01)	1.46E-01	3.54E-01	-0.058	0.039	0.94 (0.87-1.02)	1.38E-01	4.71E-01
218	L_LDL_FC_pct	0.046	0.022	1.05 (1-1.09)	3.27E-02	1.85E-01	0.027	0.043	1.03 (0.94-1.12)	5.36E-01	8.05E-01
219	L_LDL_TG_pct	-0.006	0.021	0.99 (0.95-1.04)	7.90E-01	9.24E-01	0.017	0.044	1.02 (0.93-1.11)	7.07E-01	8.78E-01
220	M_LDL_PL_pct	0.041	0.022	1.04 (1-1.09)	5.84E-02	2.18E-01	0.033	0.041	1.03 (0.95-1.12)	4.29E-01	8.05E-01
221	M_LDL_C_pct	-0.014	0.021	0.99 (0.95-1.03)	4.87E-01	6.96E-01	-0.015	0.042	0.99 (0.91-1.07)	7.26E-01	8.78E-01
222	M_LDL_CE_pct	-0.062	0.022	0.94 (0.9-0.98)	6.07E-03	1.12E-01	-0.046	0.041	0.96 (0.88-1.04)	2.59E-01	6.92E-01
223	M_LDL_FC_pct	0.052	0.022	1.05 (1.01-1.1)	1.99E-02	1.50E-01	0.033	0.042	1.03 (0.95-1.12)	4.35E-01	8.05E-01
224	M_LDL_TG_pct	-0.010	0.021	0.99 (0.95-1.03)	6.43E-01	8.13E-01	-0.001	0.043	1 (0.92-1.09)	9.90E-01	9.90E-01
225	S_LDL_PL_pct	0.061	0.023	1.06 (1.02-1.11)	8.06E-03	1.12E-01	0.110	0.042	1.12 (1.03-1.21)	9.35E-03	2.17E-01
226	S_LDL_C_pct	-0.014	0.022	0.99 (0.94-1.03)	5.26E-01	7.19E-01	-0.055	0.043	0.95 (0.87-1.03)	1.97E-01	6.00E-01
227	S_LDL_CE_pct	-0.057	0.023	0.95 (0.9-0.99)	1.22E-02	1.25E-01	-0.094	0.041	0.91 (0.84-0.99)	2.26E-02	2.47E-01
228	S_LDL_FC_pct	0.043	0.022	1.04 (1-1.09)	4.59E-02	2.04E-01	0.044	0.041	1.05 (0.96-1.13)	2.81E-01	7.10E-01
22	S_LDL_TG_pct	-0.042	0.022	0.96 (0.92-1)	5.05E-02	2.10E-01	-0.043	0.043	0.96 (0.88-1.04)	3.12E-01	7.40E-01

		Lipid modifying agent nonusers (n = 158,429)					Lipid modifying agent users (n = 26,135)				
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**
9											
230	XL_HDL_PL_pct	0.019	0.023	1.02 (0.97-1.07)	4.15E-01	6.62E-01	0.089	0.044	1.09 (1-1.19)	4.31E-02	2.93E-01
231	XL_HDL_C_pct	0.001	0.022	1 (0.96-1.05)	9.48E-01	9.81E-01	-0.050	0.041	0.95 (0.88-1.03)	2.24E-01	6.48E-01
232	XL_HDL_CE_pct	0.001	0.021	1 (0.96-1.04)	9.75E-01	9.95E-01	-0.053	0.040	0.95 (0.88-1.03)	1.80E-01	5.58E-01
233	XL_HDL_FC_pct	0.004	0.025	1 (0.96-1.06)	8.73E-01	9.61E-01	-0.015	0.045	0.99 (0.9-1.08)	7.46E-01	8.78E-01
234	XL_HDL_TG_pct	-0.044	0.023	0.96 (0.91-1)	6.02E-02	2.18E-01	-0.070	0.043	0.93 (0.86-1.01)	1.01E-01	4.22E-01
235	L_HDL_PL_pct	-0.029	0.022	0.97 (0.93-1.01)	1.78E-01	4.09E-01	-0.045	0.044	0.96 (0.88-1.04)	3.05E-01	7.40E-01
236	L_HDL_C_pct	0.028	0.022	1.03 (0.99-1.07)	2.04E-01	4.34E-01	0.032	0.042	1.03 (0.95-1.12)	4.46E-01	8.05E-01
237	L_HDL_CE_pct	0.019	0.021	1.02 (0.98-1.06)	3.68E-01	6.11E-01	0.009	0.041	1.01 (0.93-1.09)	8.19E-01	9.09E-01
238	L_HDL_FC_pct	0.055	0.023	1.06 (1.01-1.1)	1.53E-02	1.41E-01	0.118	0.045	1.13 (1.03-1.23)	8.00E-03	2.17E-01
239	L_HDL_TG_pct	-0.020	0.023	0.98 (0.94-1.03)	3.83E-01	6.23E-01	-0.014	0.043	0.99 (0.91-1.07)	7.41E-01	8.78E-01
240	M_HDL_PL_pct	-0.015	0.023	0.99 (0.94-1.03)	5.15E-01	7.19E-01	0.068	0.044	1.07 (0.98-1.17)	1.22E-01	4.63E-01
241	M_HDL_C_pct	0.018	0.022	1.02 (0.97-1.06)	4.34E-01	6.72E-01	-0.048	0.042	0.95 (0.88-1.04)	2.55E-01	6.89E-01
242	M_HDL_CE_pct	0.015	0.022	1.02 (0.97-1.06)	4.88E-01	6.96E-01	-0.047	0.041	0.95 (0.88-1.03)	2.47E-01	6.77E-01
243	M_HDL_FC_pct	0.019	0.025	1.02 (0.97-1.07)	4.50E-01	6.82E-01	-0.027	0.048	0.97 (0.89-1.07)	5.68E-01	8.05E-01
244	M_HDL_TG_pct	-0.018	0.023	0.98 (0.94-1.03)	4.33E-01	6.72E-01	0.031	0.044	1.03 (0.95-1.13)	4.75E-01	8.05E-01
245	S_HDL_PL_pct	-0.025	0.023	0.98 (0.93-1.02)	2.77E-01	5.19E-01	-0.001	0.042	1 (0.92-1.09)	9.81E-01	9.85E-01
246	S_HDL_C_pct	0.027	0.022	1.03 (0.99-1.07)	2.13E-01	4.43E-01	-0.023	0.041	0.98 (0.9-1.06)	5.84E-01	8.05E-01

		Lipid modifying agent nonusers (n = 158,429)					Lipid modifying agent users (n = 26,135)				
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**
247	S_HDL_CE_pct	0.010	0.021	1.01 (0.97-1.05)	6.47E-01	8.13E-01	-0.048	0.041	0.95 (0.88-1.03)	2.32E-01	6.57E-01
248	S_HDL_FC_pct	0.070	0.022	1.07 (1.03-1.12)	1.94E-03	1.12E-01	0.099	0.044	1.1 (1.01-1.2)	2.46E-02	2.47E-01
249	S_HDL_TG_pct	-0.018	0.023	0.98 (0.94-1.03)	4.38E-01	6.73E-01	0.036	0.045	1.04 (0.95-1.13)	4.15E-01	8.05E-01

SE, standard error. HR, hazard ratio. CI, confidence intervals. FDR, false discovery rate.
Abbreviations are in line with **Supplementary Table 3.1**.

Supplementary Table 4.6 Association of individual metabolites with incident ischaemic stroke in competing risks regression in model 7 (n = 184,564)

ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR
1	Total_C	0.016	0.020	1.02 (0.98-1.06)	4.40E-01	6.12E-01
2	non_HDL_C	0.024	0.020	1.02 (0.99-1.06)	2.20E-01	3.89E-01
3	Remnant_C	0.026	0.020	1.03 (0.99-1.07)	1.90E-01	3.58E-01
4	VLDL_C	0.005	0.020	1 (0.97-1.04)	8.10E-01	8.92E-01
5	Clinical_LDL_C	0.030	0.020	1.03 (0.99-1.07)	1.20E-01	2.74E-01
6	LDL_C	0.022	0.020	1.02 (0.98-1.06)	2.70E-01	4.45E-01
7	HDL_C	-0.025	0.023	0.98 (0.93-1.02)	2.80E-01	4.53E-01
8	Total_TG	-0.035	0.020	0.97 (0.93-1)	8.10E-02	2.20E-01
9	VLDL_TG	-0.042	0.020	0.96 (0.92-1)	3.90E-02	1.37E-01
10	LDL_TG	0.008	0.020	1.01 (0.97-1.05)	6.90E-01	8.18E-01
11	HDL_TG	-0.030	0.020	0.97 (0.93-1.01)	1.40E-01	2.98E-01
12	Total_PL	-0.012	0.021	0.99 (0.95-1.03)	5.60E-01	7.15E-01
13	VLDL_PL	-0.013	0.020	0.99 (0.95-1.03)	5.00E-01	6.62E-01
14	LDL_PL	0.026	0.019	1.03 (0.99-1.07)	1.80E-01	3.45E-01
15	HDL_PL	-0.039	0.023	0.96 (0.92-1.01)	8.60E-02	2.23E-01
16	Total_CE	0.016	0.020	1.02 (0.98-1.06)	4.40E-01	6.12E-01
17	VLDL_CE	0.016	0.020	1.02 (0.98-1.06)	4.10E-01	5.94E-01
18	LDL_CE	0.017	0.020	1.02 (0.98-1.06)	3.80E-01	5.67E-01
19	HDL_CE	-0.029	0.023	0.97 (0.93-1.02)	2.10E-01	3.84E-01
20	Total_FC	0.016	0.020	1.02 (0.98-1.06)	4.30E-01	6.08E-01
21	VLDL_FC	-0.009	0.020	0.99 (0.95-1.03)	6.30E-01	7.80E-01
22	LDL_FC	0.034	0.020	1.03 (1-1.08)	8.20E-02	2.20E-01
23	HDL_FC	-0.010	0.024	0.99 (0.95-1.04)	6.70E-01	8.06E-01
24	Total_L	-0.008	0.020	0.99 (0.95-1.03)	7.00E-01	8.26E-01
25	VLDL_L	-0.025	0.020	0.98 (0.94-1.01)	2.10E-01	3.84E-01
26	LDL_L	0.023	0.020	1.02 (0.98-1.06)	2.50E-01	4.26E-01
27	HDL_L	-0.036	0.023	0.96 (0.92-1.01)	1.20E-01	2.74E-01
28	Total_P	-0.044	0.022	0.96 (0.92-1)	4.20E-02	1.41E-01

ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR
29	VLDL_P	0.000	0.020	1 (0.96-1.04)	9.90E-01	1.00E+00
30	LDL_P	0.026	0.019	1.03 (0.99-1.07)	1.70E-01	3.41E-01
31	HDL_P	-0.051	0.022	0.95 (0.91-0.99)	1.90E-02	9.86E-02
32	VLDL_size	-0.070	0.021	0.93 (0.9-0.97)	6.80E-04	2.49E-02
33	LDL_size	0.028	0.020	1.03 (0.99-1.07)	1.50E-01	3.09E-01
34	HDL_size	0.013	0.024	1.01 (0.97-1.06)	5.70E-01	7.24E-01
35	Phosphoglyc	-0.023	0.021	0.98 (0.94-1.02)	2.70E-01	4.45E-01
36	TG_by_PG	-0.032	0.021	0.97 (0.93-1.01)	1.40E-01	2.98E-01
37	Cholines	-0.020	0.021	0.98 (0.94-1.02)	3.60E-01	5.57E-01
38	Phosphatidylc	-0.021	0.021	0.98 (0.94-1.02)	3.40E-01	5.36E-01
39	Sphingomyelins	0.036	0.021	1.04 (0.99-1.08)	9.40E-02	2.39E-01
40	ApoB	0.028	0.019	1.03 (0.99-1.07)	1.50E-01	3.09E-01
41	ApoA1	-0.046	0.023	0.96 (0.91-1)	4.40E-02	1.42E-01
42	ApoB_by_ApoA1	0.047	0.020	1.05 (1.01-1.09)	1.90E-02	9.86E-02
43	Total_FA	-0.022	0.020	0.98 (0.94-1.02)	2.70E-01	4.45E-01
44	Unsaturation	-0.004	0.021	1 (0.96-1.04)	8.30E-01	9.02E-01
45	Omega_3	-0.032	0.021	0.97 (0.93-1.01)	1.40E-01	2.98E-01
46	Omega_6	-0.023	0.020	0.98 (0.94-1.02)	2.40E-01	4.12E-01
47	PUFA	-0.024	0.020	0.98 (0.94-1.01)	2.20E-01	3.89E-01
48	MUFA	-0.017	0.020	0.98 (0.95-1.02)	4.00E-01	5.89E-01
49	SFA	-0.020	0.020	0.98 (0.94-1.02)	3.00E-01	4.82E-01
50	LA	-0.025	0.020	0.98 (0.94-1.01)	2.00E-01	3.74E-01
51	DHA	-0.001	0.021	1 (0.96-1.04)	9.60E-01	9.80E-01
52	Omega_3_pct	-0.026	0.022	0.97 (0.93-1.02)	2.30E-01	3.98E-01
53	Omega_6_pct	0.008	0.020	1.01 (0.97-1.05)	6.90E-01	8.18E-01
54	PUFA_pct	0.005	0.020	1.01 (0.97-1.05)	8.00E-01	8.89E-01
55	MUFA_pct	-0.002	0.021	1 (0.96-1.04)	9.40E-01	9.67E-01
56	SFA_pct	-0.004	0.020	1 (0.96-1.03)	8.30E-01	9.02E-01
57	LA_pct	-0.006	0.020	0.99 (0.96-1.03)	7.70E-01	8.72E-01
58	DHA_pct	0.014	0.021	1.01 (0.97-1.06)	5.10E-01	6.65E-01

ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR
59	PUFA_by_MUFA	0.003	0.021	1 (0.96-1.04)	8.70E-01	9.14E-01
60	Omega_6_by_Omega_3	0.026	0.022	1.03 (0.98-1.07)	2.30E-01	3.98E-01
61	Ala	0.016	0.024	1.02 (0.97-1.07)	5.10E-01	6.65E-01
62	Gln	0.006	0.020	1.01 (0.97-1.05)	7.50E-01	8.57E-01
63	Gly	-0.003	0.018	1 (0.96-1.03)	8.60E-01	9.07E-01
64	His	-0.039	0.019	0.96 (0.93-1)	3.80E-02	1.35E-01
65	Total_BCAA	-0.027	0.021	0.97 (0.93-1.01)	1.90E-01	3.58E-01
66	Ile	-0.012	0.020	0.99 (0.95-1.03)	5.60E-01	7.15E-01
67	Leu	-0.025	0.020	0.98 (0.94-1.02)	2.20E-01	3.89E-01
68	Val	-0.032	0.021	0.97 (0.93-1.01)	1.30E-01	2.86E-01
69	Phe	0.051	0.019	1.05 (1.01-1.09)	8.10E-03	5.93E-02
70	Tyr	0.028	0.019	1.03 (0.99-1.07)	1.50E-01	3.09E-01
71	Glucose	-0.015	0.020	0.99 (0.95-1.02)	4.60E-01	6.29E-01
72	Lactate	0.039	0.020	1.04 (1-1.08)	5.10E-02	1.61E-01
73	Pyruvate	0.017	0.020	1.02 (0.98-1.06)	3.90E-01	5.78E-01
74	Citrate	-0.012	0.020	0.99 (0.95-1.03)	5.60E-01	7.15E-01
75	bOHbutyrate	-0.005	0.022	0.99 (0.95-1.04)	8.10E-01	8.92E-01
76	Acetate	-0.012	0.016	0.99 (0.96-1.02)	4.60E-01	6.29E-01
77	Acetoacetate	-0.020	0.019	0.98 (0.95-1.02)	2.80E-01	4.53E-01
78	Acetone	-0.002	0.019	1 (0.96-1.04)	9.30E-01	9.61E-01
79	Creatinine	0.040	0.023	1.04 (0.99-1.09)	8.90E-02	2.28E-01
80	Albumin	-0.126	0.019	0.88 (0.85-0.92)	2.20E-11	5.48E-09
81	GlycA	0.037	0.021	1.04 (1-1.08)	8.00E-02	2.20E-01
82	XXL_VLDL_P	-0.047	0.021	0.95 (0.92-0.99)	2.50E-02	1.07E-01
83	XXL_VLDL_L	-0.053	0.020	0.95 (0.91-0.99)	9.20E-03	6.38E-02
84	XXL_VLDL_PL	-0.047	0.021	0.95 (0.92-0.99)	2.40E-02	1.07E-01
85	XXL_VLDL_C	-0.045	0.021	0.96 (0.92-1)	3.00E-02	1.20E-01
86	XXL_VLDL_CE	-0.034	0.021	0.97 (0.93-1.01)	1.00E-01	2.47E-01
87	XXL_VLDL_FC	-0.058	0.021	0.94 (0.91-0.98)	5.00E-03	4.45E-02
88	XXL_VLDL_TG	-0.045	0.018	0.96 (0.92-0.99)	1.10E-02	6.68E-02

ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR
89	XL_VLDL_P	-0.037	0.021	0.96 (0.92-1)	8.30E-02	2.20E-01
90	XL_VLDL_L	-0.038	0.021	0.96 (0.92-1)	7.00E-02	2.03E-01
91	XL_VLDL_PL	-0.034	0.021	0.97 (0.93-1.01)	1.10E-01	2.61E-01
92	XL_VLDL_C	-0.023	0.021	0.98 (0.94-1.02)	2.60E-01	4.40E-01
93	XL_VLDL_CE	-0.013	0.021	0.99 (0.95-1.03)	5.10E-01	6.65E-01
94	XL_VLDL_FC	-0.032	0.021	0.97 (0.93-1.01)	1.20E-01	2.74E-01
95	XL_VLDL_TG	-0.044	0.021	0.96 (0.92-1)	3.60E-02	1.30E-01
96	L_VLDL_P	-0.034	0.021	0.97 (0.93-1.01)	1.00E-01	2.47E-01
97	L_VLDL_L	-0.036	0.021	0.96 (0.93-1)	8.20E-02	2.20E-01
98	L_VLDL_PL	-0.033	0.021	0.97 (0.93-1.01)	1.20E-01	2.74E-01
99	L_VLDL_C	-0.026	0.021	0.97 (0.94-1.01)	2.10E-01	3.84E-01
100	L_VLDL_CE	-0.019	0.020	0.98 (0.94-1.02)	3.50E-01	5.45E-01
101	L_VLDL_FC	-0.031	0.021	0.97 (0.93-1.01)	1.30E-01	2.86E-01
102	L_VLDL_TG	-0.041	0.020	0.96 (0.92-1)	4.40E-02	1.42E-01
103	M_VLDL_P	-0.006	0.020	0.99 (0.96-1.03)	7.80E-01	8.75E-01
104	M_VLDL_L	-0.014	0.020	0.99 (0.95-1.02)	4.80E-01	6.53E-01
105	M_VLDL_PL	0.000	0.020	1 (0.96-1.04)	9.90E-01	1.00E+00
106	M_VLDL_C	0.026	0.019	1.03 (0.99-1.07)	1.80E-01	3.45E-01
107	M_VLDL_CE	0.042	0.019	1.04 (1.01-1.08)	2.50E-02	1.07E-01
108	M_VLDL_FC	0.008	0.019	1.01 (0.97-1.05)	6.80E-01	8.14E-01
109	M_VLDL_TG	-0.036	0.020	0.96 (0.93-1)	7.10E-02	2.03E-01
110	S_VLDL_P	-0.010	0.020	0.99 (0.95-1.03)	6.00E-01	7.55E-01
111	S_VLDL_L	-0.009	0.020	0.99 (0.95-1.03)	6.70E-01	8.06E-01
112	S_VLDL_PL	0.006	0.020	1.01 (0.97-1.05)	7.80E-01	8.75E-01
113	S_VLDL_C	0.016	0.020	1.02 (0.98-1.06)	4.10E-01	5.94E-01
114	S_VLDL_CE	0.016	0.020	1.02 (0.98-1.06)	4.30E-01	6.08E-01
115	S_VLDL_FC	0.017	0.019	1.02 (0.98-1.06)	3.80E-01	5.67E-01
116	S_VLDL_TG	-0.032	0.020	0.97 (0.93-1.01)	1.10E-01	2.61E-01
117	XS_VLDL_P	0.041	0.020	1.04 (1-1.08)	3.60E-02	1.30E-01
118	XS_VLDL_L	0.039	0.020	1.04 (1-1.08)	4.70E-02	1.50E-01

ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR
119	XS_VLDL_PL	0.040	0.020	1.04 (1-1.08)	4.30E-02	1.42E-01
120	XS_VLDL_C	0.056	0.020	1.06 (1.02-1.1)	5.00E-03	4.45E-02
121	XS_VLDL_CE	0.059	0.020	1.06 (1.02-1.1)	2.90E-03	4.25E-02
122	XS_VLDL_FC	0.045	0.020	1.05 (1.01-1.09)	2.10E-02	1.03E-01
123	XS_VLDL_TG	-0.005	0.020	0.99 (0.96-1.03)	7.90E-01	8.82E-01
124	IDL_P	0.043	0.020	1.04 (1-1.08)	3.00E-02	1.20E-01
125	IDL_L	0.047	0.020	1.05 (1.01-1.09)	1.90E-02	9.86E-02
126	IDL_PL	0.052	0.020	1.05 (1.01-1.1)	1.00E-02	6.38E-02
127	IDL_C	0.046	0.020	1.05 (1.01-1.09)	2.30E-02	1.06E-01
128	IDL_CE	0.043	0.020	1.04 (1-1.09)	3.30E-02	1.25E-01
129	IDL_FC	0.052	0.020	1.05 (1.01-1.09)	9.70E-03	6.38E-02
130	IDL_TG	0.017	0.020	1.02 (0.98-1.06)	3.80E-01	5.67E-01
131	L_LDL_P	0.030	0.019	1.03 (0.99-1.07)	1.10E-01	2.61E-01
132	L_LDL_L	0.028	0.020	1.03 (0.99-1.07)	1.50E-01	3.09E-01
133	L_LDL_PL	0.029	0.020	1.03 (0.99-1.07)	1.30E-01	2.86E-01
134	L_LDL_C	0.028	0.020	1.03 (0.99-1.07)	1.60E-01	3.27E-01
135	L_LDL_CE	0.024	0.020	1.02 (0.99-1.06)	2.20E-01	3.89E-01
136	L_LDL_FC	0.037	0.020	1.04 (1-1.08)	6.40E-02	1.94E-01
137	L_LDL_TG	0.018	0.020	1.02 (0.98-1.06)	3.50E-01	5.45E-01
138	M_LDL_P	0.016	0.020	1.02 (0.98-1.06)	4.10E-01	5.94E-01
139	M_LDL_L	0.011	0.020	1.01 (0.97-1.05)	5.90E-01	7.46E-01
140	M_LDL_PL	0.015	0.020	1.02 (0.98-1.06)	4.30E-01	6.08E-01
141	M_LDL_C	0.010	0.020	1.01 (0.97-1.05)	6.20E-01	7.76E-01
142	M_LDL_CE	0.003	0.020	1 (0.97-1.04)	8.60E-01	9.07E-01
143	M_LDL_FC	0.026	0.019	1.03 (0.99-1.07)	1.80E-01	3.45E-01
144	M_LDL_TG	-0.003	0.020	1 (0.96-1.04)	8.90E-01	9.31E-01
145	S_LDL_P	0.021	0.019	1.02 (0.98-1.06)	2.70E-01	4.45E-01
146	S_LDL_L	0.020	0.019	1.02 (0.98-1.06)	3.10E-01	4.95E-01
147	S_LDL_PL	0.035	0.019	1.04 (1-1.08)	6.50E-02	1.95E-01
148	S_LDL_C	0.017	0.019	1.02 (0.98-1.06)	3.80E-01	5.67E-01

ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR
149	S_LDL_CE	0.009	0.019	1.01 (0.97-1.05)	6.50E-01	7.97E-01
150	S_LDL_FC	0.036	0.019	1.04 (1-1.08)	5.70E-02	1.77E-01
151	S_LDL_TG	-0.024	0.020	0.98 (0.94-1.01)	2.30E-01	3.98E-01
152	XL_HDL_P	0.047	0.023	1.05 (1-1.1)	4.00E-02	1.38E-01
153	XL_HDL_L	0.051	0.022	1.05 (1.01-1.1)	2.20E-02	1.03E-01
154	XL_HDL_PL	0.047	0.022	1.05 (1-1.09)	3.30E-02	1.25E-01
155	XL_HDL_C	0.057	0.022	1.06 (1.01-1.11)	1.00E-02	6.38E-02
156	XL_HDL_CE	0.050	0.022	1.05 (1.01-1.1)	2.60E-02	1.08E-01
157	XL_HDL_FC	0.077	0.022	1.08 (1.04-1.13)	3.30E-04	2.49E-02
158	XL_HDL_TG	-0.006	0.019	0.99 (0.96-1.03)	7.50E-01	8.57E-01
159	L_HDL_P	0.000	0.023	1 (0.96-1.05)	1.00E+00	1.00E+00
160	L_HDL_L	0.000	0.023	1 (0.96-1.05)	1.00E+00	1.00E+00
161	L_HDL_PL	-0.005	0.023	1 (0.95-1.04)	8.40E-01	9.02E-01
162	L_HDL_C	0.007	0.023	1.01 (0.96-1.05)	7.50E-01	8.57E-01
163	L_HDL_CE	0.005	0.023	1 (0.96-1.05)	8.30E-01	9.02E-01
164	L_HDL_FC	0.016	0.023	1.02 (0.97-1.06)	5.00E-01	6.62E-01
165	L_HDL_TG	-0.018	0.020	0.98 (0.94-1.02)	3.80E-01	5.67E-01
166	M_HDL_P	-0.053	0.022	0.95 (0.91-0.99)	1.60E-02	8.85E-02
167	M_HDL_L	-0.055	0.022	0.95 (0.91-0.99)	1.30E-02	7.53E-02
168	M_HDL_PL	-0.056	0.022	0.95 (0.91-0.99)	9.90E-03	6.38E-02
169	M_HDL_C	-0.048	0.022	0.95 (0.91-1)	3.10E-02	1.23E-01
170	M_HDL_CE	-0.049	0.022	0.95 (0.91-0.99)	2.60E-02	1.08E-01
171	M_HDL_FC	-0.040	0.022	0.96 (0.92-1)	7.10E-02	2.03E-01
172	M_HDL_TG	-0.036	0.020	0.96 (0.93-1)	7.10E-02	2.03E-01
173	S_HDL_P	-0.058	0.020	0.94 (0.91-0.98)	4.60E-03	4.45E-02
174	S_HDL_L	-0.065	0.021	0.94 (0.9-0.98)	1.80E-03	3.20E-02
175	S_HDL_PL	-0.067	0.021	0.93 (0.9-0.97)	1.30E-03	2.49E-02
176	S_HDL_C	-0.056	0.021	0.95 (0.91-0.98)	6.80E-03	5.13E-02
177	S_HDL_CE	-0.059	0.020	0.94 (0.91-0.98)	3.50E-03	4.45E-02
178	S_HDL_FC	-0.036	0.021	0.96 (0.93-1)	8.40E-02	2.20E-01

ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR
179	S_HDL_TG	-0.033	0.021	0.97 (0.93-1.01)	1.10E-01	2.61E-01
180	XXL_VLDL_PL_pct	-0.015	0.023	0.98 (0.94-1.03)	5.00E-01	6.62E-01
181	XXL_VLDL_C_pct	0.055	0.020	1.06 (1.02-1.1)	5.70E-03	4.81E-02
182	XXL_VLDL_CE_pct	0.067	0.020	1.07 (1.03-1.11)	9.20E-04	2.49E-02
183	XXL_VLDL_FC_pct	0.019	0.020	1.02 (0.98-1.06)	3.40E-01	5.36E-01
184	XXL_VLDL_TG_pct	-0.025	0.017	0.98 (0.94-1.01)	1.30E-01	2.86E-01
185	XL_VLDL_PL_pct	0.008	0.023	1.01 (0.96-1.05)	7.30E-01	8.49E-01
186	XL_VLDL_C_pct	0.060	0.021	1.06 (1.02-1.11)	4.40E-03	4.45E-02
187	XL_VLDL_CE_pct	0.061	0.021	1.06 (1.02-1.11)	3.20E-03	4.43E-02
188	XL_VLDL_FC_pct	0.044	0.022	1.05 (1-1.09)	4.10E-02	1.40E-01
189	XL_VLDL_TG_pct	-0.056	0.021	0.95 (0.91-0.98)	6.70E-03	5.13E-02
190	L_VLDL_PL_pct	-0.010	0.023	0.99 (0.95-1.04)	6.60E-01	8.06E-01
191	L_VLDL_C_pct	0.046	0.020	1.05 (1.01-1.09)	2.10E-02	1.03E-01
192	L_VLDL_CE_pct	0.050	0.020	1.05 (1.01-1.09)	1.40E-02	7.92E-02
193	L_VLDL_FC_pct	0.028	0.021	1.03 (0.99-1.07)	1.80E-01	3.45E-01
194	L_VLDL_TG_pct	-0.034	0.019	0.97 (0.93-1)	8.00E-02	2.20E-01
195	M_VLDL_PL_pct	0.068	0.020	1.07 (1.03-1.11)	6.40E-04	2.49E-02
196	M_VLDL_C_pct	0.068	0.020	1.07 (1.03-1.11)	7.60E-04	2.49E-02
197	M_VLDL_CE_pct	0.069	0.020	1.07 (1.03-1.11)	4.40E-04	2.49E-02
198	M_VLDL_FC_pct	0.067	0.020	1.07 (1.03-1.11)	8.90E-04	2.49E-02
199	M_VLDL_TG_pct	-0.063	0.021	0.94 (0.9-0.98)	2.30E-03	3.58E-02
200	S_VLDL_PL_pct	0.055	0.020	1.06 (1.02-1.1)	6.80E-03	5.13E-02
201	S_VLDL_C_pct	0.064	0.020	1.07 (1.03-1.11)	1.10E-03	2.49E-02
202	S_VLDL_CE_pct	0.063	0.019	1.07 (1.03-1.11)	1.00E-03	2.49E-02
203	S_VLDL_FC_pct	0.057	0.020	1.06 (1.02-1.1)	4.80E-03	4.45E-02
204	S_VLDL_TG_pct	-0.056	0.020	0.95 (0.91-0.98)	4.90E-03	4.45E-02
205	XS_VLDL_PL_pct	0.014	0.021	1.01 (0.97-1.06)	5.00E-01	6.62E-01
206	XS_VLDL_C_pct	0.057	0.020	1.06 (1.02-1.1)	4.90E-03	4.45E-02
207	XS_VLDL_CE_pct	0.056	0.020	1.06 (1.02-1.1)	5.80E-03	4.81E-02
208	XS_VLDL_FC_pct	0.048	0.019	1.05 (1.01-1.09)	1.10E-02	6.68E-02

ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR
209	XS_VLDL_TG_pct	-0.060	0.021	0.94 (0.9-0.98)	3.70E-03	4.45E-02
210	IDL_PL_pct	0.012	0.019	1.01 (0.97-1.05)	5.30E-01	6.87E-01
211	IDL_C_pct	0.017	0.019	1.02 (0.98-1.06)	3.80E-01	5.67E-01
212	IDL_CE_pct	0.004	0.019	1 (0.97-1.04)	8.50E-01	9.04E-01
213	IDL_FC_pct	0.032	0.018	1.03 (1-1.07)	7.80E-02	2.20E-01
214	IDL_TG_pct	-0.025	0.020	0.98 (0.94-1.01)	2.20E-01	3.89E-01
215	L_LDL_PL_pct	0.001	0.018	1 (0.97-1.04)	9.50E-01	9.73E-01
216	L_LDL_C_pct	0.010	0.020	1.01 (0.97-1.05)	6.40E-01	7.89E-01
217	L_LDL_CE_pct	-0.027	0.019	0.97 (0.94-1.01)	1.40E-01	2.98E-01
218	L_LDL_FC_pct	0.043	0.020	1.04 (1-1.09)	3.20E-02	1.25E-01
219	L_LDL_TG_pct	-0.009	0.020	0.99 (0.95-1.03)	6.70E-01	8.06E-01
220	M_LDL_PL_pct	0.036	0.020	1.04 (1-1.08)	6.40E-02	1.94E-01
221	M_LDL_C_pct	-0.007	0.019	0.99 (0.96-1.03)	7.20E-01	8.42E-01
222	M_LDL_CE_pct	-0.050	0.020	0.95 (0.92-0.99)	1.20E-02	7.11E-02
223	M_LDL_FC_pct	0.046	0.020	1.05 (1.01-1.09)	2.10E-02	1.03E-01
224	M_LDL_TG_pct	-0.015	0.020	0.98 (0.95-1.02)	4.40E-01	6.12E-01
225	S_LDL_PL_pct	0.062	0.020	1.06 (1.02-1.11)	2.00E-03	3.32E-02
226	S_LDL_C_pct	-0.013	0.019	0.99 (0.95-1.03)	5.00E-01	6.62E-01
227	S_LDL_CE_pct	-0.055	0.020	0.95 (0.91-0.98)	5.00E-03	4.45E-02
228	S_LDL_FC_pct	0.043	0.019	1.04 (1.01-1.08)	2.50E-02	1.07E-01
229	S_LDL_TG_pct	-0.045	0.020	0.96 (0.92-0.99)	2.20E-02	1.03E-01
230	XL_HDL_PL_pct	0.028	0.020	1.03 (0.99-1.07)	1.70E-01	3.41E-01
231	XL_HDL_C_pct	-0.004	0.019	1 (0.96-1.03)	8.40E-01	9.02E-01
232	XL_HDL_CE_pct	-0.006	0.017	0.99 (0.96-1.03)	7.20E-01	8.42E-01
233	XL_HDL_FC_pct	0.004	0.022	1 (0.96-1.05)	8.50E-01	9.04E-01
234	XL_HDL_TG_pct	-0.043	0.021	0.96 (0.92-1)	3.40E-02	1.26E-01
235	L_HDL_PL_pct	-0.033	0.020	0.97 (0.93-1.01)	1.00E-01	2.47E-01
236	L_HDL_C_pct	0.026	0.020	1.03 (0.99-1.07)	1.80E-01	3.45E-01
237	L_HDL_CE_pct	0.015	0.019	1.02 (0.98-1.05)	4.30E-01	6.08E-01
238	L_HDL_FC_pct	0.067	0.021	1.07 (1.03-1.11)	1.30E-03	2.49E-02

ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR
239	L_HDL_TG_pct	-0.015	0.021	0.98 (0.95-1.03)	4.60E-01	6.29E-01
240	M_HDL_PL_pct	0.000	0.021	1 (0.96-1.04)	1.00E+00	1.00E+00
241	M_HDL_C_pct	0.004	0.020	1 (0.97-1.04)	8.40E-01	9.02E-01
242	M_HDL_CE_pct	0.002	0.019	1 (0.96-1.04)	9.20E-01	9.58E-01
243	M_HDL_FC_pct	0.011	0.023	1.01 (0.97-1.06)	6.30E-01	7.80E-01
244	M_HDL_TG_pct	-0.007	0.021	0.99 (0.95-1.03)	7.40E-01	8.57E-01
245	S_HDL_PL_pct	-0.027	0.020	0.97 (0.94-1.01)	1.80E-01	3.45E-01
246	S_HDL_C_pct	0.021	0.019	1.02 (0.98-1.06)	2.80E-01	4.53E-01
247	S_HDL_CE_pct	0.002	0.019	1 (0.97-1.04)	9.30E-01	9.61E-01
248	S_HDL_FC_pct	0.076	0.020	1.08 (1.04-1.12)	1.80E-04	2.24E-02
249	S_HDL_TG_pct	-0.006	0.021	0.99 (0.95-1.04)	7.70E-01	8.72E-01

SE, standard error. HR, hazard ratio. CI, confidence intervals. FDR, false discovery rate. Abbreviations are in line with **Supplementary Table 3.1**.

Supplementary Table 4.7 Association of individual metabolites with incident ischaemic stroke after exclusion of the first 3 years of follow-up in model 7 (n = 183,189)

ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR
1	Total_C	0.002	0.021	1 (0.96-1.05)	9.26E-01	9.65E-01
2	non_HDL_C	0.009	0.021	1.01 (0.97-1.05)	6.57E-01	8.26E-01
3	Remnant_C	0.010	0.021	1.01 (0.97-1.05)	6.20E-01	8.00E-01
4	VLDL_C	-0.008	0.021	0.99 (0.95-1.03)	7.09E-01	8.80E-01
5	Clinical_LDL_C	0.015	0.021	1.02 (0.98-1.06)	4.60E-01	6.94E-01
6	LDL_C	0.008	0.021	1.01 (0.97-1.05)	7.11E-01	8.80E-01
7	HDL_C	-0.023	0.024	0.98 (0.93-1.02)	3.29E-01	5.53E-01
8	Total_TG	-0.045	0.021	0.96 (0.92-1)	3.24E-02	1.46E-01
9	VLDL_TG	-0.052	0.021	0.95 (0.91-0.99)	1.36E-02	9.53E-02
10	LDL_TG	0.000	0.021	1 (0.96-1.04)	9.94E-01	9.97E-01
11	HDL_TG	-0.034	0.021	0.97 (0.93-1.01)	1.04E-01	2.59E-01
12	Total_PL	-0.022	0.022	0.98 (0.94-1.02)	3.02E-01	5.19E-01
13	VLDL_PL	-0.025	0.021	0.98 (0.94-1.02)	2.28E-01	4.33E-01
14	LDL_PL	0.012	0.021	1.01 (0.97-1.05)	5.60E-01	7.74E-01
15	HDL_PL	-0.036	0.023	0.96 (0.92-1.01)	1.17E-01	2.75E-01
16	Total_CE	0.002	0.021	1 (0.96-1.05)	9.15E-01	9.61E-01
17	VLDL_CE	0.003	0.021	1 (0.96-1.05)	8.68E-01	9.40E-01
18	LDL_CE	0.003	0.021	1 (0.96-1.05)	8.76E-01	9.40E-01
19	HDL_CE	-0.025	0.023	0.98 (0.93-1.02)	2.78E-01	5.02E-01
20	Total_FC	0.001	0.021	1 (0.96-1.04)	9.56E-01	9.75E-01
21	VLDL_FC	-0.022	0.021	0.98 (0.94-1.02)	2.96E-01	5.12E-01
22	LDL_FC	0.020	0.021	1.02 (0.98-1.06)	3.43E-01	5.70E-01
23	HDL_FC	-0.012	0.024	0.99 (0.94-1.04)	6.08E-01	7.96E-01
24	Total_L	-0.021	0.021	0.98 (0.94-1.02)	3.21E-01	5.43E-01
25	VLDL_L	-0.037	0.021	0.96 (0.93-1.01)	8.13E-02	2.25E-01
26	LDL_L	0.009	0.021	1.01 (0.97-1.05)	6.82E-01	8.53E-01
27	HDL_L	-0.033	0.023	0.97 (0.92-1.01)	1.54E-01	3.40E-01
28	Total_P	-0.047	0.022	0.96 (0.91-1)	3.64E-02	1.54E-01

ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR
29	VLDL_P	-0.012	0.021	0.99 (0.95-1.03)	5.79E-01	7.79E-01
30	LDL_P	0.013	0.021	1.01 (0.97-1.06)	5.40E-01	7.68E-01
31	HDL_P	-0.051	0.022	0.95 (0.91-0.99)	2.27E-02	1.26E-01
32	VLDL_size	-0.079	0.022	0.92 (0.89-0.97)	2.91E-04	3.62E-02
33	LDL_size	0.028	0.020	1.03 (0.99-1.07)	1.66E-01	3.47E-01
34	HDL_size	0.015	0.024	1.02 (0.97-1.07)	5.28E-01	7.56E-01
35	Phosphoglyc	-0.031	0.022	0.97 (0.93-1.01)	1.51E-01	3.38E-01
36	TG_by_PG	-0.040	0.022	0.96 (0.92-1)	7.38E-02	2.16E-01
37	Cholines	-0.029	0.022	0.97 (0.93-1.02)	1.92E-01	3.87E-01
38	Phosphatidylc	-0.029	0.022	0.97 (0.93-1.01)	1.86E-01	3.83E-01
39	Sphingomyelins	0.026	0.022	1.03 (0.98-1.07)	2.44E-01	4.50E-01
40	ApoB	0.013	0.021	1.01 (0.97-1.06)	5.26E-01	7.56E-01
41	ApoA1	-0.043	0.023	0.96 (0.92-1)	5.89E-02	1.90E-01
42	ApoB_by_ApoA1	0.033	0.021	1.03 (0.99-1.08)	1.16E-01	2.75E-01
43	Total_FA	-0.032	0.021	0.97 (0.93-1.01)	1.22E-01	2.84E-01
44	Unsaturation	-0.005	0.022	1 (0.95-1.04)	8.35E-01	9.40E-01
45	Omega_3	-0.039	0.022	0.96 (0.92-1)	7.53E-02	2.18E-01
46	Omega_6	-0.036	0.021	0.97 (0.93-1.01)	8.64E-02	2.29E-01
47	PUFA	-0.036	0.021	0.97 (0.93-1.01)	8.22E-02	2.25E-01
48	MUFA	-0.025	0.021	0.98 (0.94-1.02)	2.31E-01	4.33E-01
49	SFA	-0.029	0.021	0.97 (0.93-1.01)	1.57E-01	3.40E-01
50	LA	-0.038	0.021	0.96 (0.92-1)	6.26E-02	1.94E-01
51	DHA	-0.004	0.022	1 (0.95-1.04)	8.43E-01	9.40E-01
52	Omega_3_pct	-0.029	0.023	0.97 (0.93-1.02)	2.04E-01	3.99E-01
53	Omega_6_pct	0.009	0.020	1.01 (0.97-1.05)	6.47E-01	8.22E-01
54	PUFA_pct	0.006	0.021	1.01 (0.97-1.05)	7.62E-01	9.21E-01
55	MUFA_pct	-0.004	0.022	1 (0.95-1.04)	8.51E-01	9.40E-01
56	SFA_pct	-0.005	0.020	1 (0.96-1.04)	8.21E-01	9.37E-01
57	LA_pct	-0.012	0.021	0.99 (0.95-1.03)	5.75E-01	7.78E-01
58	DHA_pct	0.017	0.022	1.02 (0.97-1.06)	4.52E-01	6.87E-01

ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR
59	PUFA_by_MUFA	0.005	0.022	1.01 (0.96-1.05)	8.01E-01	9.29E-01
60	Omega_6_by_Omega_3	0.029	0.022	1.03 (0.99-1.08)	1.96E-01	3.91E-01
61	Ala	0.005	0.025	1.01 (0.96-1.06)	8.41E-01	9.40E-01
62	Gln	-0.004	0.020	1 (0.96-1.04)	8.62E-01	9.40E-01
63	Gly	-0.003	0.019	1 (0.96-1.03)	8.71E-01	9.40E-01
64	His	-0.043	0.020	0.96 (0.92-1)	2.81E-02	1.43E-01
65	Total_BCAA	-0.034	0.021	0.97 (0.93-1.01)	1.11E-01	2.66E-01
66	Ile	-0.015	0.021	0.99 (0.95-1.03)	4.80E-01	7.16E-01
67	Leu	-0.029	0.021	0.97 (0.93-1.01)	1.65E-01	3.47E-01
68	Val	-0.043	0.022	0.96 (0.92-1)	5.14E-02	1.85E-01
69	Phe	0.043	0.020	1.04 (1-1.09)	3.15E-02	1.46E-01
70	Tyr	0.017	0.021	1.02 (0.98-1.06)	4.13E-01	6.42E-01
71	Glucose	-0.015	0.020	0.99 (0.95-1.03)	4.50E-01	6.87E-01
72	Lactate	0.040	0.021	1.04 (1-1.08)	5.93E-02	1.90E-01
73	Pyruvate	0.018	0.021	1.02 (0.98-1.06)	3.90E-01	6.23E-01
74	Citrate	-0.013	0.022	0.99 (0.95-1.03)	5.53E-01	7.73E-01
75	bOHbutyrate	0.000	0.022	1 (0.96-1.04)	9.97E-01	9.97E-01
76	Acetate	-0.015	0.018	0.99 (0.95-1.02)	4.12E-01	6.42E-01
77	Acetoacetate	-0.020	0.020	0.98 (0.94-1.02)	3.09E-01	5.27E-01
78	Acetone	0.002	0.020	1 (0.96-1.04)	9.33E-01	9.66E-01
79	Creatinine	0.031	0.024	1.03 (0.99-1.08)	1.92E-01	3.87E-01
80	Albumin	-0.134	0.020	0.87 (0.84-0.91)	9.21E-12	2.29E-09
81	GlycA	0.038	0.022	1.04 (1-1.09)	8.37E-02	2.27E-01
82	XXL_VLDL_P	-0.055	0.022	0.95 (0.91-0.99)	1.16E-02	9.53E-02
83	XXL_VLDL_L	-0.062	0.021	0.94 (0.9-0.98)	3.78E-03	4.38E-02
84	XXL_VLDL_PL	-0.056	0.022	0.95 (0.91-0.99)	1.18E-02	9.53E-02
85	XXL_VLDL_C	-0.053	0.022	0.95 (0.91-0.99)	1.49E-02	9.53E-02
86	XXL_VLDL_CE	-0.042	0.022	0.96 (0.92-1)	5.67E-02	1.90E-01
87	XXL_VLDL_FC	-0.066	0.022	0.94 (0.9-0.98)	2.12E-03	4.38E-02
88	XXL_VLDL_TG	-0.057	0.019	0.95 (0.91-0.98)	3.28E-03	4.38E-02

ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR
89	XL_VLDL_P	-0.047	0.022	0.95 (0.91-1)	3.28E-02	1.46E-01
90	XL_VLDL_L	-0.049	0.022	0.95 (0.91-0.99)	2.70E-02	1.40E-01
91	XL_VLDL_PL	-0.044	0.022	0.96 (0.92-1)	4.73E-02	1.75E-01
92	XL_VLDL_C	-0.035	0.022	0.97 (0.93-1.01)	1.10E-01	2.66E-01
93	XL_VLDL_CE	-0.026	0.022	0.98 (0.93-1.02)	2.40E-01	4.45E-01
94	XL_VLDL_FC	-0.043	0.022	0.96 (0.92-1)	4.77E-02	1.75E-01
95	XL_VLDL_TG	-0.054	0.022	0.95 (0.91-0.99)	1.40E-02	9.53E-02
96	L_VLDL_P	-0.045	0.022	0.96 (0.92-1)	3.72E-02	1.54E-01
97	L_VLDL_L	-0.048	0.022	0.95 (0.91-1)	2.70E-02	1.40E-01
98	L_VLDL_PL	-0.044	0.022	0.96 (0.92-1)	4.73E-02	1.75E-01
99	L_VLDL_C	-0.037	0.022	0.96 (0.92-1.01)	8.80E-02	2.29E-01
100	L_VLDL_CE	-0.030	0.021	0.97 (0.93-1.01)	1.60E-01	3.40E-01
101	L_VLDL_FC	-0.043	0.022	0.96 (0.92-1)	5.20E-02	1.85E-01
102	L_VLDL_TG	-0.053	0.021	0.95 (0.91-0.99)	1.20E-02	9.53E-02
103	M_VLDL_P	-0.018	0.021	0.98 (0.94-1.02)	3.81E-01	6.13E-01
104	M_VLDL_L	-0.027	0.021	0.97 (0.94-1.01)	1.93E-01	3.87E-01
105	M_VLDL_PL	-0.013	0.021	0.99 (0.95-1.03)	5.43E-01	7.68E-01
106	M_VLDL_C	0.013	0.021	1.01 (0.97-1.06)	5.29E-01	7.56E-01
107	M_VLDL_CE	0.030	0.020	1.03 (0.99-1.07)	1.43E-01	3.23E-01
108	M_VLDL_FC	-0.005	0.021	1 (0.96-1.04)	8.10E-01	9.29E-01
109	M_VLDL_TG	-0.048	0.021	0.95 (0.92-0.99)	2.26E-02	1.26E-01
110	S_VLDL_P	-0.020	0.021	0.98 (0.94-1.02)	3.34E-01	5.59E-01
111	S_VLDL_L	-0.019	0.021	0.98 (0.94-1.02)	3.68E-01	5.98E-01
112	S_VLDL_PL	-0.006	0.021	0.99 (0.95-1.04)	7.79E-01	9.29E-01
113	S_VLDL_C	0.005	0.021	1.01 (0.97-1.05)	7.99E-01	9.29E-01
114	S_VLDL_CE	0.005	0.021	1.01 (0.97-1.05)	8.00E-01	9.29E-01
115	S_VLDL_FC	0.005	0.021	1.01 (0.97-1.05)	8.08E-01	9.29E-01
116	S_VLDL_TG	-0.040	0.021	0.96 (0.92-1)	6.04E-02	1.90E-01
117	XS_VLDL_P	0.030	0.021	1.03 (0.99-1.07)	1.53E-01	3.40E-01
118	XS_VLDL_L	0.028	0.021	1.03 (0.99-1.07)	1.77E-01	3.68E-01

ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR
119	XS_VLDL_PL	0.029	0.021	1.03 (0.99-1.07)	1.58E-01	3.40E-01
120	XS_VLDL_C	0.043	0.021	1.04 (1-1.09)	4.26E-02	1.66E-01
121	XS_VLDL_CE	0.046	0.021	1.05 (1.01-1.09)	2.98E-02	1.45E-01
122	XS_VLDL_FC	0.034	0.021	1.04 (0.99-1.08)	1.03E-01	2.59E-01
123	XS_VLDL_TG	-0.010	0.021	0.99 (0.95-1.03)	6.15E-01	8.00E-01
124	IDL_P	0.027	0.021	1.03 (0.99-1.07)	2.03E-01	3.99E-01
125	IDL_L	0.032	0.021	1.03 (0.99-1.08)	1.34E-01	3.09E-01
126	IDL_PL	0.038	0.021	1.04 (1-1.08)	7.39E-02	2.16E-01
127	IDL_C	0.030	0.021	1.03 (0.99-1.08)	1.59E-01	3.40E-01
128	IDL_CE	0.027	0.021	1.03 (0.99-1.07)	2.11E-01	4.07E-01
129	IDL_FC	0.038	0.021	1.04 (1-1.08)	7.37E-02	2.16E-01
130	IDL_TG	0.011	0.021	1.01 (0.97-1.05)	5.83E-01	7.81E-01
131	L_LDL_P	0.017	0.020	1.02 (0.98-1.06)	4.12E-01	6.42E-01
132	L_LDL_L	0.014	0.021	1.01 (0.97-1.06)	4.98E-01	7.33E-01
133	L_LDL_PL	0.015	0.021	1.02 (0.97-1.06)	4.68E-01	7.02E-01
134	L_LDL_C	0.013	0.021	1.01 (0.97-1.06)	5.21E-01	7.56E-01
135	L_LDL_CE	0.010	0.021	1.01 (0.97-1.05)	6.28E-01	8.02E-01
136	L_LDL_FC	0.022	0.021	1.02 (0.98-1.07)	2.89E-01	5.11E-01
137	L_LDL_TG	0.012	0.021	1.01 (0.97-1.05)	5.74E-01	7.78E-01
138	M_LDL_P	0.004	0.021	1 (0.96-1.05)	8.63E-01	9.40E-01
139	M_LDL_L	-0.003	0.021	1 (0.96-1.04)	8.86E-01	9.47E-01
140	M_LDL_PL	0.001	0.021	1 (0.96-1.04)	9.48E-01	9.75E-01
141	M_LDL_C	-0.004	0.021	1 (0.96-1.04)	8.54E-01	9.40E-01
142	M_LDL_CE	-0.009	0.021	0.99 (0.95-1.03)	6.51E-01	8.23E-01
143	M_LDL_FC	0.012	0.021	1.01 (0.97-1.05)	5.70E-01	7.78E-01
144	M_LDL_TG	-0.011	0.021	0.99 (0.95-1.03)	6.01E-01	7.92E-01
145	S_LDL_P	0.007	0.021	1.01 (0.97-1.05)	7.51E-01	9.21E-01
146	S_LDL_L	0.005	0.021	1.01 (0.97-1.05)	8.02E-01	9.29E-01
147	S_LDL_PL	0.021	0.020	1.02 (0.98-1.06)	2.95E-01	5.12E-01
148	S_LDL_C	0.002	0.021	1 (0.96-1.04)	9.09E-01	9.59E-01

ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR
149	S_LDL_CE	-0.006	0.021	0.99 (0.96-1.04)	7.81E-01	9.29E-01
150	S_LDL_FC	0.023	0.020	1.02 (0.98-1.07)	2.64E-01	4.79E-01
151	S_LDL_TG	-0.033	0.021	0.97 (0.93-1.01)	1.11E-01	2.66E-01
152	XL_HDL_P	0.037	0.023	1.04 (0.99-1.09)	1.10E-01	2.66E-01
153	XL_HDL_L	0.044	0.023	1.05 (1-1.09)	5.78E-02	1.90E-01
154	XL_HDL_PL	0.041	0.023	1.04 (1-1.09)	7.61E-02	2.18E-01
155	XL_HDL_C	0.049	0.023	1.05 (1-1.1)	3.38E-02	1.46E-01
156	XL_HDL_CE	0.043	0.023	1.04 (1-1.09)	6.66E-02	2.02E-01
157	XL_HDL_FC	0.066	0.022	1.07 (1.02-1.12)	3.33E-03	4.38E-02
158	XL_HDL_TG	-0.014	0.020	0.99 (0.95-1.03)	4.91E-01	7.28E-01
159	L_HDL_P	-0.001	0.024	1 (0.95-1.05)	9.72E-01	9.88E-01
160	L_HDL_L	0.000	0.024	1 (0.96-1.05)	9.87E-01	9.94E-01
161	L_HDL_PL	-0.003	0.024	1 (0.95-1.04)	8.92E-01	9.49E-01
162	L_HDL_C	0.006	0.023	1.01 (0.96-1.05)	7.88E-01	9.29E-01
163	L_HDL_CE	0.005	0.023	1.01 (0.96-1.05)	8.40E-01	9.40E-01
164	L_HDL_FC	0.013	0.024	1.01 (0.97-1.06)	5.94E-01	7.88E-01
165	L_HDL_TG	-0.022	0.021	0.98 (0.94-1.02)	2.95E-01	5.12E-01
166	M_HDL_P	-0.049	0.023	0.95 (0.91-1)	2.87E-02	1.43E-01
167	M_HDL_L	-0.050	0.022	0.95 (0.91-0.99)	2.63E-02	1.40E-01
168	M_HDL_PL	-0.051	0.022	0.95 (0.91-0.99)	2.11E-02	1.25E-01
169	M_HDL_C	-0.043	0.023	0.96 (0.92-1)	5.86E-02	1.90E-01
170	M_HDL_CE	-0.043	0.022	0.96 (0.92-1)	5.48E-02	1.90E-01
171	M_HDL_FC	-0.039	0.023	0.96 (0.92-1.01)	8.75E-02	2.29E-01
172	M_HDL_TG	-0.039	0.021	0.96 (0.92-1)	6.32E-02	1.94E-01
173	S_HDL_P	-0.059	0.021	0.94 (0.9-0.98)	5.10E-03	5.52E-02
174	S_HDL_L	-0.065	0.021	0.94 (0.9-0.98)	2.46E-03	4.38E-02
175	S_HDL_PL	-0.066	0.021	0.94 (0.9-0.98)	2.06E-03	4.38E-02
176	S_HDL_C	-0.056	0.021	0.95 (0.91-0.99)	8.55E-03	8.19E-02
177	S_HDL_CE	-0.058	0.021	0.94 (0.91-0.98)	5.89E-03	6.11E-02
178	S_HDL_FC	-0.040	0.021	0.96 (0.92-1)	5.97E-02	1.90E-01

ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR
179	S_HDL_TG	-0.038	0.022	0.96 (0.92-1)	7.74E-02	2.19E-01
180	XXL_VLDL_PL_pct	-0.020	0.023	0.98 (0.94-1.03)	3.93E-01	6.23E-01
181	XXL_VLDL_C_pct	0.062	0.021	1.06 (1.02-1.11)	2.81E-03	4.38E-02
182	XXL_VLDL_CE_pct	0.075	0.021	1.08 (1.03-1.12)	5.03E-04	4.17E-02
183	XXL_VLDL_FC_pct	0.027	0.021	1.03 (0.99-1.07)	2.05E-01	3.99E-01
184	XXL_VLDL_TG_pct	-0.035	0.018	0.97 (0.93-1)	4.51E-02	1.73E-01
185	XL_VLDL_PL_pct	0.006	0.023	1.01 (0.96-1.05)	7.86E-01	9.29E-01
186	XL_VLDL_C_pct	0.063	0.022	1.07 (1.02-1.11)	3.87E-03	4.38E-02
187	XL_VLDL_CE_pct	0.064	0.022	1.07 (1.02-1.11)	2.98E-03	4.38E-02
188	XL_VLDL_FC_pct	0.047	0.022	1.05 (1-1.09)	3.39E-02	1.46E-01
189	XL_VLDL_TG_pct	-0.058	0.022	0.94 (0.9-0.99)	7.36E-03	7.33E-02
190	L_VLDL_PL_pct	-0.012	0.023	0.99 (0.94-1.04)	6.17E-01	8.00E-01
191	L_VLDL_C_pct	0.051	0.021	1.05 (1.01-1.1)	1.35E-02	9.53E-02
192	L_VLDL_CE_pct	0.055	0.021	1.06 (1.01-1.1)	9.06E-03	8.36E-02
193	L_VLDL_FC_pct	0.032	0.021	1.03 (0.99-1.08)	1.41E-01	3.23E-01
194	L_VLDL_TG_pct	-0.039	0.020	0.96 (0.92-1)	5.67E-02	1.90E-01
195	M_VLDL_PL_pct	0.066	0.021	1.07 (1.03-1.11)	1.84E-03	4.38E-02
196	M_VLDL_C_pct	0.065	0.021	1.07 (1.02-1.11)	1.99E-03	4.38E-02
197	M_VLDL_CE_pct	0.066	0.021	1.07 (1.03-1.11)	1.55E-03	4.38E-02
198	M_VLDL_FC_pct	0.065	0.021	1.07 (1.02-1.11)	2.24E-03	4.38E-02
199	M_VLDL_TG_pct	-0.062	0.022	0.94 (0.9-0.98)	3.77E-03	4.38E-02
200	S_VLDL_PL_pct	0.051	0.021	1.05 (1.01-1.1)	1.76E-02	1.10E-01
201	S_VLDL_C_pct	0.061	0.021	1.06 (1.02-1.11)	3.35E-03	4.38E-02
202	S_VLDL_CE_pct	0.061	0.020	1.06 (1.02-1.11)	2.57E-03	4.38E-02
203	S_VLDL_FC_pct	0.052	0.021	1.05 (1.01-1.1)	1.48E-02	9.53E-02
204	S_VLDL_TG_pct	-0.052	0.021	0.95 (0.91-0.99)	1.30E-02	9.53E-02
205	XS_VLDL_PL_pct	0.017	0.022	1.02 (0.98-1.06)	4.20E-01	6.46E-01
206	XS_VLDL_C_pct	0.050	0.021	1.05 (1.01-1.1)	1.88E-02	1.14E-01
207	XS_VLDL_CE_pct	0.049	0.021	1.05 (1.01-1.1)	2.18E-02	1.26E-01
208	XS_VLDL_FC_pct	0.043	0.020	1.04 (1-1.09)	3.15E-02	1.46E-01

ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR
209	XS_VLDL_TG_pct	-0.054	0.022	0.95 (0.91-0.99)	1.30E-02	9.53E-02
210	IDL_PL_pct	0.024	0.020	1.02 (0.99-1.07)	2.32E-01	4.33E-01
211	IDL_C_pct	0.005	0.020	1.01 (0.97-1.05)	8.01E-01	9.29E-01
212	IDL_CE_pct	-0.010	0.020	0.99 (0.95-1.03)	5.95E-01	7.88E-01
213	IDL_FC_pct	0.033	0.019	1.03 (1-1.07)	9.35E-02	2.38E-01
214	IDL_TG_pct	-0.017	0.021	0.98 (0.94-1.02)	4.16E-01	6.43E-01
215	L_LDL_PL_pct	0.003	0.019	1 (0.97-1.04)	8.75E-01	9.40E-01
216	L_LDL_C_pct	0.000	0.021	1 (0.96-1.04)	9.85E-01	9.94E-01
217	L_LDL_CE_pct	-0.033	0.019	0.97 (0.93-1.01)	9.01E-02	2.31E-01
218	L_LDL_FC_pct	0.036	0.021	1.04 (1-1.08)	7.96E-02	2.23E-01
219	L_LDL_TG_pct	-0.002	0.021	1 (0.96-1.04)	9.35E-01	9.66E-01
220	M_LDL_PL_pct	0.034	0.020	1.04 (1-1.08)	8.83E-02	2.29E-01
221	M_LDL_C_pct	-0.012	0.020	0.99 (0.95-1.03)	5.47E-01	7.70E-01
222	M_LDL_CE_pct	-0.051	0.021	0.95 (0.91-0.99)	1.33E-02	9.53E-02
223	M_LDL_FC_pct	0.043	0.021	1.04 (1-1.09)	3.89E-02	1.57E-01
224	M_LDL_TG_pct	-0.010	0.021	0.99 (0.95-1.03)	6.25E-01	8.02E-01
225	S_LDL_PL_pct	0.068	0.021	1.07 (1.03-1.12)	1.27E-03	4.38E-02
226	S_LDL_C_pct	-0.022	0.021	0.98 (0.94-1.02)	2.87E-01	5.11E-01
227	S_LDL_CE_pct	-0.062	0.021	0.94 (0.9-0.98)	2.95E-03	4.38E-02
228	S_LDL_FC_pct	0.041	0.020	1.04 (1-1.08)	4.10E-02	1.62E-01
229	S_LDL_TG_pct	-0.044	0.021	0.96 (0.92-1)	3.33E-02	1.46E-01
230	XL_HDL_PL_pct	0.024	0.021	1.03 (0.98-1.07)	2.46E-01	4.51E-01
231	XL_HDL_C_pct	-0.003	0.020	1 (0.96-1.04)	8.67E-01	9.40E-01
232	XL_HDL_CE_pct	-0.006	0.019	0.99 (0.96-1.03)	7.62E-01	9.21E-01
233	XL_HDL_FC_pct	0.003	0.023	1 (0.96-1.05)	8.97E-01	9.50E-01
234	XL_HDL_TG_pct	-0.044	0.022	0.96 (0.92-1)	3.92E-02	1.57E-01
235	L_HDL_PL_pct	-0.026	0.021	0.98 (0.94-1.02)	2.22E-01	4.25E-01
236	L_HDL_C_pct	0.022	0.020	1.02 (0.98-1.06)	2.86E-01	5.11E-01
237	L_HDL_CE_pct	0.013	0.020	1.01 (0.98-1.05)	5.01E-01	7.34E-01
238	L_HDL_FC_pct	0.051	0.021	1.05 (1.01-1.1)	1.46E-02	9.53E-02

ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR
239	L_HDL_TG_pct	-0.019	0.021	0.98 (0.94-1.02)	3.70E-01	5.98E-01
240	M_HDL_PL_pct	0.001	0.022	1 (0.96-1.05)	9.55E-01	9.75E-01
241	M_HDL_C_pct	0.006	0.021	1.01 (0.97-1.05)	7.62E-01	9.21E-01
242	M_HDL_CE_pct	0.006	0.020	1.01 (0.97-1.05)	7.62E-01	9.21E-01
243	M_HDL_FC_pct	0.002	0.023	1 (0.96-1.05)	9.26E-01	9.65E-01
244	M_HDL_TG_pct	-0.013	0.022	0.99 (0.95-1.03)	5.58E-01	7.74E-01
245	S_HDL_PL_pct	-0.019	0.021	0.98 (0.94-1.02)	3.57E-01	5.85E-01
246	S_HDL_C_pct	0.019	0.020	1.02 (0.98-1.06)	3.50E-01	5.78E-01
247	S_HDL_CE_pct	0.003	0.020	1 (0.97-1.04)	8.70E-01	9.40E-01
248	S_HDL_FC_pct	0.063	0.021	1.07 (1.02-1.11)	3.13E-03	4.38E-02
249	S_HDL_TG_pct	-0.012	0.022	0.99 (0.95-1.03)	5.74E-01	7.78E-01

SE, standard error. HR, hazard ratio. CI, confidence intervals. FDR, false discovery rate. Abbreviations are in line with **Supplementary Table 3.1**.

Supplementary Table 4.8 Association of individual metabolites with incident ischaemic stroke in the White ethnic group and the others in model 7

		The White ethnic group (n = 175,904)					The others (n = 8,660)				
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**
1	Total_C	0.010	0.021	1.01 (0.97-1.05)	6.20E-01	7.64E-01	-0.045	0.107	0.96 (0.78-1.18)	6.70E-01	9.60E-01
2	non_HDL_C	0.019	0.020	1.02 (0.98-1.06)	3.45E-01	5.34E-01	-0.072	0.104	0.93 (0.76-1.14)	4.91E-01	9.51E-01
3	Remnant_C	0.022	0.020	1.02 (0.98-1.06)	2.83E-01	4.70E-01	-0.060	0.104	0.94 (0.77-1.16)	5.66E-01	9.51E-01
4	VLDL_C	0.000	0.020	1 (0.96-1.04)	9.82E-01	9.88E-01	-0.057	0.101	0.95 (0.78-1.15)	5.75E-01	9.51E-01
5	Clinical_LDL_C	0.026	0.020	1.03 (0.99-1.07)	2.02E-01	3.68E-01	-0.097	0.105	0.91 (0.74-1.11)	3.53E-01	9.51E-01
6	LDL_C	0.016	0.020	1.02 (0.98-1.06)	4.25E-01	6.19E-01	-0.084	0.104	0.92 (0.75-1.13)	4.19E-01	9.51E-01
7	HDL_C	-0.030	0.023	0.97 (0.93-1.02)	1.96E-01	3.62E-01	0.097	0.125	1.1 (0.86-1.41)	4.38E-01	9.51E-01
8	Total_TG	-0.042	0.020	0.96 (0.92-1)	3.91E-02	1.19E-01	-0.007	0.098	0.99 (0.82-1.2)	9.45E-01	9.93E-01
9	VLDL_TG	-0.050	0.021	0.95 (0.91-0.99)	1.57E-02	7.09E-02	-0.019	0.100	0.98 (0.81-1.19)	8.47E-01	9.87E-01
10	LDL_TG	0.004	0.020	1 (0.97-1.04)	8.53E-01	9.16E-01	0.045	0.096	1.05 (0.87-1.26)	6.39E-01	9.60E-01
11	HDL_TG	-0.033	0.020	0.97 (0.93-1.01)	9.86E-02	2.19E-01	0.025	0.092	1.03 (0.86-1.23)	7.88E-01	9.82E-01
12	Total_PL	-0.019	0.021	0.98 (0.94-1.02)	3.78E-01	5.64E-01	0.002	0.104	1 (0.82-1.23)	9.85E-01	9.93E-01
13	VLDL_PL	-0.019	0.020	0.98 (0.94-1.02)	3.60E-01	5.50E-01	-0.050	0.098	0.95 (0.78-1.15)	6.14E-01	9.51E-01
14	LDL_PL	0.021	0.020	1.02 (0.98-1.06)	2.87E-01	4.70E-01	-0.093	0.105	0.91 (0.74-1.12)	3.80E-01	9.51E-01
15	HDL_PL	-0.044	0.022	0.96 (0.92-1)	5.11E-02	1.36E-01	0.098	0.118	1.1 (0.88-1.39)	4.05E-01	9.51E-01
16	Total_CE	0.010	0.021	1.01 (0.97-1.05)	6.23E-01	7.64E-01	-0.048	0.108	0.95 (0.77-1.18)	6.55E-01	9.60E-01
17	VLDL_CE	0.012	0.020	1.01 (0.97-1.05)	5.42E-01	7.00E-01	-0.053	0.103	0.95 (0.78-1.16)	6.09E-01	9.51E-01
18	LDL_CE	0.011	0.020	1.01 (0.97-1.05)	5.95E-01	7.41E-01	-0.078	0.103	0.93 (0.76-1.13)	4.53E-01	9.51E-01
19	HDL_CE	-0.033	0.023	0.97 (0.93-1.01)	1.46E-01	2.82E-01	0.083	0.125	1.09 (0.85-1.39)	5.08E-01	9.51E-01
20	Total_FC	0.010	0.021	1.01 (0.97-1.05)	6.18E-01	7.64E-01	-0.039	0.105	0.96 (0.78-1.18)	7.08E-01	9.80E-01
21	VLDL_FC	-0.015	0.020	0.99 (0.95-1.03)	4.70E-01	6.50E-01	-0.059	0.099	0.94 (0.78-1.14)	5.48E-01	9.51E-01
22	LDL_FC	0.030	0.020	1.03 (0.99-1.07)	1.34E-01	2.70E-01	-0.100	0.106	0.91 (0.74-1.12)	3.47E-01	9.51E-01
23	HDL_FC	-0.014	0.023	0.99 (0.94-1.03)	5.31E-01	7.00E-01	0.136	0.122	1.15 (0.9-1.46)	2.66E-01	9.51E-01
24	Total_L	-0.015	0.020	0.99 (0.95-1.03)	4.65E-01	6.50E-01	-0.024	0.100	0.98 (0.8-1.19)	8.08E-01	9.82E-01
25	VLDL_L	-0.031	0.020	0.97 (0.93-1.01)	1.25E-01	2.58E-01	-0.040	0.099	0.96 (0.79-1.17)	6.87E-01	9.60E-01
26	LDL_L	0.017	0.020	1.02 (0.98-1.06)	4.01E-01	5.94E-01	-0.079	0.104	0.92 (0.75-1.13)	4.47E-01	9.51E-01
27	HDL_L	-0.041	0.023	0.96 (0.92-1)	7.35E-02	1.75E-01	0.102	0.121	1.11 (0.87-1.4)	3.98E-01	9.51E-01
28	Total_P	-0.052	0.022	0.95 (0.91-0.99)	1.52E-02	7.09E-02	-0.018	0.113	0.98 (0.79-1.23)	8.75E-01	9.90E-01
29	VLDL_P	-0.005	0.020	1 (0.96-1.04)	8.19E-01	9.07E-01	-0.033	0.099	0.97 (0.8-1.18)	7.39E-01	9.82E-01
30	LDL_P	0.021	0.020	1.02 (0.98-1.06)	2.87E-01	4.70E-01	-0.057	0.105	0.95 (0.77-1.16)	5.87E-01	9.51E-01
31	HDL_P	-0.059	0.022	0.94 (0.9-0.98)	6.23E-03	3.97E-02	-0.006	0.115	0.99 (0.79-1.25)	9.59E-01	9.93E-01
32	VLDL_size	-0.079	0.021	0.92 (0.89-0.96)	1.67E-04	9.10E-03	-0.076	0.105	0.93 (0.75-1.14)	4.68E-01	9.51E-01

		The White ethnic group (n = 175,904)					The others (n = 8,660)				
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**
33	LDL_size	0.036	0.020	1.04 (1-1.08)	6.69E-02	1.67E-01	-0.124	0.105	0.88 (0.72-1.09)	2.39E-01	9.51E-01
34	HDL_size	0.014	0.023	1.01 (0.97-1.06)	5.46E-01	7.01E-01	0.232	0.124	1.26 (0.99-1.61)	6.14E-02	9.51E-01
35	Phosphoglyc	-0.029	0.021	0.97 (0.93-1.01)	1.65E-01	3.15E-01	-0.001	0.105	1 (0.81-1.23)	9.96E-01	9.98E-01
36	TG_by_PG	-0.037	0.022	0.96 (0.92-1.01)	8.55E-02	1.94E-01	-0.009	0.108	0.99 (0.8-1.23)	9.37E-01	9.93E-01
37	Cholines	-0.026	0.021	0.97 (0.94-1.02)	2.25E-01	3.95E-01	0.000	0.108	1 (0.81-1.23)	9.98E-01	9.98E-01
38	Phosphatidylc	-0.028	0.021	0.97 (0.93-1.01)	1.88E-01	3.48E-01	0.026	0.106	1.03 (0.83-1.26)	8.04E-01	9.82E-01
39	Sphingomyelins	0.035	0.022	1.04 (0.99-1.08)	1.09E-01	2.33E-01	-0.003	0.113	1 (0.8-1.24)	9.77E-01	9.93E-01
40	ApoB	0.023	0.020	1.02 (0.98-1.07)	2.49E-01	4.26E-01	-0.064	0.105	0.94 (0.76-1.15)	5.42E-01	9.51E-01
41	ApoA1	-0.052	0.022	0.95 (0.91-0.99)	1.95E-02	7.84E-02	0.064	0.119	1.07 (0.85-1.35)	5.90E-01	9.51E-01
42	ApoB_by_ApoA1	0.046	0.021	1.05 (1.01-1.09)	2.40E-02	9.06E-02	-0.093	0.111	0.91 (0.73-1.13)	4.03E-01	9.51E-01
43	Total_FA	-0.027	0.020	0.97 (0.94-1.01)	1.72E-01	3.24E-01	-0.041	0.097	0.96 (0.79-1.16)	6.74E-01	9.60E-01
44	Unsaturation	-0.007	0.021	0.99 (0.95-1.04)	7.30E-01	8.43E-01	-0.118	0.107	0.89 (0.72-1.1)	2.72E-01	9.51E-01
45	Omega_3	-0.039	0.021	0.96 (0.92-1)	6.29E-02	1.61E-01	-0.081	0.105	0.92 (0.75-1.13)	4.38E-01	9.51E-01
46	Omega_6	-0.031	0.020	0.97 (0.93-1.01)	1.21E-01	2.57E-01	-0.058	0.095	0.94 (0.78-1.14)	5.43E-01	9.51E-01
47	PUFA	-0.033	0.020	0.97 (0.93-1.01)	1.04E-01	2.26E-01	-0.071	0.097	0.93 (0.77-1.13)	4.69E-01	9.51E-01
48	MUFA	-0.020	0.020	0.98 (0.94-1.02)	3.31E-01	5.19E-01	-0.040	0.100	0.96 (0.79-1.17)	6.88E-01	9.60E-01
49	SFA	-0.025	0.020	0.98 (0.94-1.01)	2.11E-01	3.81E-01	-0.002	0.095	1 (0.83-1.2)	9.85E-01	9.93E-01
50	LA	-0.035	0.020	0.97 (0.93-1)	8.11E-02	1.85E-01	-0.024	0.092	0.98 (0.81-1.17)	7.93E-01	9.82E-01
51	DHA	-0.008	0.021	0.99 (0.95-1.03)	7.03E-01	8.33E-01	-0.024	0.105	0.98 (0.8-1.2)	8.22E-01	9.84E-01
52	Omega_3_pct	-0.033	0.022	0.97 (0.93-1.01)	1.37E-01	2.73E-01	-0.073	0.106	0.93 (0.76-1.15)	4.94E-01	9.51E-01
53	Omega_6_pct	0.007	0.020	1.01 (0.97-1.05)	7.27E-01	8.43E-01	-0.013	0.102	0.99 (0.81-1.2)	8.96E-01	9.90E-01
54	PUFA_pct	0.002	0.020	1 (0.96-1.04)	9.12E-01	9.50E-01	-0.042	0.104	0.96 (0.78-1.18)	6.88E-01	9.60E-01
55	MUFA_pct	0.001	0.021	1 (0.96-1.04)	9.52E-01	9.71E-01	-0.028	0.111	0.97 (0.78-1.21)	8.01E-01	9.82E-01
56	SFA_pct	-0.004	0.020	1 (0.96-1.04)	8.45E-01	9.15E-01	0.107	0.088	1.11 (0.94-1.32)	2.20E-01	9.51E-01
57	LA_pct	-0.013	0.020	0.99 (0.95-1.03)	5.11E-01	6.84E-01	0.024	0.101	1.02 (0.84-1.25)	8.11E-01	9.82E-01
58	DHA_pct	0.010	0.022	1.01 (0.97-1.05)	6.52E-01	7.88E-01	0.003	0.104	1 (0.82-1.23)	9.76E-01	9.93E-01
59	PUFA_by_MUFA	0.000	0.021	1 (0.96-1.04)	9.84E-01	9.88E-01	-0.006	0.109	0.99 (0.8-1.23)	9.57E-01	9.93E-01
60	Omega_6_by_Omega_3	0.032	0.022	1.03 (0.99-1.08)	1.45E-01	2.82E-01	0.061	0.104	1.06 (0.87-1.3)	5.55E-01	9.51E-01
61	Ala	0.019	0.024	1.02 (0.97-1.07)	4.39E-01	6.32E-01	-0.154	0.125	0.86 (0.67-1.1)	2.17E-01	9.51E-01
62	Gln	0.004	0.020	1 (0.97-1.04)	8.23E-01	9.07E-01	-0.095	0.097	0.91 (0.75-1.1)	3.26E-01	9.51E-01
63	Gly	-0.003	0.018	1 (0.96-1.03)	8.78E-01	9.39E-01	-0.009	0.095	0.99 (0.82-1.19)	9.26E-01	9.93E-01
64	His	-0.045	0.019	0.96 (0.92-0.99)	1.74E-02	7.62E-02	-0.004	0.104	1 (0.81-1.22)	9.68E-01	9.93E-01
65	Total_BCAA	-0.032	0.021	0.97 (0.93-1.01)	1.27E-01	2.61E-01	-0.082	0.114	0.92 (0.74-1.15)	4.74E-01	9.51E-01
66	Ile	-0.015	0.020	0.99 (0.95-1.02)	4.46E-01	6.35E-01	0.016	0.115	1.02 (0.81-1.27)	8.89E-01	9.90E-01
67	Leu	-0.030	0.020	0.97 (0.93-1.01)	1.44E-01	2.82E-01	-0.097	0.113	0.91 (0.73-1.13)	3.95E-01	9.51E-01

		The White ethnic group (n = 175,904)					The others (n = 8,660)				
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**
68	Val	-0.037	0.021	0.96 (0.92-1)	8.00E-02	1.84E-01	-0.104	0.114	0.9 (0.72-1.13)	3.60E-01	9.51E-01
69	Phe	0.058	0.020	1.06 (1.02-1.1)	3.25E-03	3.11E-02	-0.091	0.101	0.91 (0.75-1.11)	3.68E-01	9.51E-01
70	Tyr	0.036	0.020	1.04 (1-1.08)	7.14E-02	1.73E-01	-0.201	0.107	0.82 (0.66-1.01)	6.08E-02	9.51E-01
71	Glucose	-0.021	0.020	0.98 (0.94-1.02)	2.89E-01	4.70E-01	0.090	0.094	1.09 (0.91-1.32)	3.38E-01	9.51E-01
72	Lactate	0.045	0.020	1.05 (1.01-1.09)	2.78E-02	9.85E-02	-0.110	0.103	0.9 (0.73-1.1)	2.85E-01	9.51E-01
73	Pyruvate	0.025	0.020	1.03 (0.99-1.07)	2.16E-01	3.82E-01	-0.147	0.099	0.86 (0.71-1.05)	1.36E-01	9.51E-01
74	Citrate	-0.012	0.021	0.99 (0.95-1.03)	5.72E-01	7.21E-01	0.017	0.110	1.02 (0.82-1.26)	8.75E-01	9.90E-01
75	bOHbutyrate	-0.001	0.022	1 (0.96-1.04)	9.81E-01	9.88E-01	-0.021	0.109	0.98 (0.79-1.21)	8.46E-01	9.87E-01
76	Acetate	-0.016	0.017	0.98 (0.95-1.02)	3.38E-01	5.26E-01	0.092	0.127	1.1 (0.85-1.41)	4.71E-01	9.51E-01
77	Acetoacetate	-0.017	0.020	0.98 (0.95-1.02)	3.75E-01	5.63E-01	-0.023	0.099	0.98 (0.81-1.19)	8.19E-01	9.84E-01
78	Acetone	-0.004	0.019	1 (0.96-1.03)	8.40E-01	9.14E-01	0.163	0.108	1.18 (0.95-1.46)	1.31E-01	9.51E-01
79	Creatinine	0.042	0.023	1.04 (1-1.09)	6.62E-02	1.67E-01	-0.035	0.116	0.97 (0.77-1.21)	7.63E-01	9.82E-01
80	Albumin	-0.134	0.019	0.87 (0.84-0.91)	2.07E-12	5.15E-10	-0.150	0.096	0.86 (0.71-1.04)	1.15E-01	9.51E-01
81	GlycA	0.050	0.021	1.05 (1.01-1.1)	1.90E-02	7.80E-02	-0.082	0.108	0.92 (0.75-1.14)	4.44E-01	9.51E-01
82	XXL_VLDL_P	-0.047	0.021	0.95 (0.92-1)	2.81E-02	9.85E-02	-0.097	0.094	0.91 (0.76-1.09)	3.02E-01	9.51E-01
83	XXL_VLDL_L	-0.055	0.021	0.95 (0.91-0.99)	8.37E-03	4.74E-02	-0.091	0.098	0.91 (0.75-1.11)	3.52E-01	9.51E-01
84	XXL_VLDL_PL	-0.046	0.022	0.96 (0.92-1)	3.48E-02	1.10E-01	-0.114	0.091	0.89 (0.75-1.07)	2.11E-01	9.51E-01
85	XXL_VLDL_C	-0.048	0.021	0.95 (0.91-0.99)	2.44E-02	9.09E-02	-0.067	0.098	0.94 (0.77-1.13)	4.95E-01	9.51E-01
86	XXL_VLDL_CE	-0.038	0.022	0.96 (0.92-1)	7.72E-02	1.81E-01	-0.045	0.100	0.96 (0.79-1.16)	6.50E-01	9.60E-01
87	XXL_VLDL_FC	-0.060	0.021	0.94 (0.9-0.98)	4.43E-03	3.34E-02	-0.083	0.093	0.92 (0.77-1.11)	3.73E-01	9.51E-01
88	XXL_VLDL_TG	-0.049	0.019	0.95 (0.92-0.99)	9.39E-03	5.08E-02	-0.034	0.084	0.97 (0.82-1.14)	6.83E-01	9.60E-01
89	XL_VLDL_P	-0.044	0.022	0.96 (0.92-1)	4.18E-02	1.23E-01	-0.033	0.102	0.97 (0.79-1.18)	7.48E-01	9.82E-01
90	XL_VLDL_L	-0.045	0.021	0.96 (0.92-1)	3.31E-02	1.08E-01	-0.028	0.102	0.97 (0.8-1.19)	7.80E-01	9.82E-01
91	XL_VLDL_PL	-0.040	0.022	0.96 (0.92-1)	6.46E-02	1.64E-01	-0.049	0.099	0.95 (0.78-1.16)	6.19E-01	9.51E-01
92	XL_VLDL_C	-0.029	0.021	0.97 (0.93-1.01)	1.68E-01	3.19E-01	-0.061	0.103	0.94 (0.77-1.15)	5.54E-01	9.51E-01
93	XL_VLDL_CE	-0.020	0.021	0.98 (0.94-1.02)	3.51E-01	5.39E-01	-0.062	0.104	0.94 (0.77-1.15)	5.53E-01	9.51E-01
94	XL_VLDL_FC	-0.038	0.021	0.96 (0.92-1)	7.39E-02	1.75E-01	-0.058	0.100	0.94 (0.78-1.15)	5.59E-01	9.51E-01
95	XL_VLDL_TG	-0.053	0.021	0.95 (0.91-0.99)	1.30E-02	6.46E-02	-0.002	0.102	1 (0.82-1.22)	9.82E-01	9.93E-01
96	L_VLDL_P	-0.042	0.021	0.96 (0.92-1)	4.58E-02	1.30E-01	-0.032	0.101	0.97 (0.8-1.18)	7.51E-01	9.82E-01
97	L_VLDL_L	-0.045	0.021	0.96 (0.92-1)	3.32E-02	1.08E-01	-0.030	0.100	0.97 (0.8-1.18)	7.61E-01	9.82E-01
98	L_VLDL_PL	-0.041	0.021	0.96 (0.92-1)	5.71E-02	1.48E-01	-0.032	0.101	0.97 (0.8-1.18)	7.49E-01	9.82E-01
99	L_VLDL_C	-0.032	0.021	0.97 (0.93-1.01)	1.23E-01	2.58E-01	-0.060	0.101	0.94 (0.77-1.15)	5.54E-01	9.51E-01
100	L_VLDL_CE	-0.025	0.021	0.98 (0.94-1.02)	2.28E-01	3.97E-01	-0.064	0.099	0.94 (0.77-1.14)	5.21E-01	9.51E-01
10	L_VLDL_FC	-0.039	0.021	0.96 (0.92-1)	6.90E-02	1.69E-01	-0.053	0.103	0.95 (0.78-1.16)	6.09E-01	9.51E-01

		The White ethnic group (n = 175,904)					The others (n = 8,660)				
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**
1											
102	L_VLDL_TG	-0.051	0.021	0.95 (0.91-0.99)	1.32E-02	6.46E-02	-0.012	0.099	0.99 (0.81-1.2)	9.02E-01	9.90E-01
103	M_VLDL_P	-0.012	0.020	0.99 (0.95-1.03)	5.42E-01	7.00E-01	-0.061	0.099	0.94 (0.78-1.14)	5.35E-01	9.51E-01
104	M_VLDL_L	-0.022	0.020	0.98 (0.94-1.02)	2.70E-01	4.55E-01	-0.039	0.097	0.96 (0.8-1.16)	6.90E-01	9.60E-01
105	M_VLDL_PL	-0.006	0.020	0.99 (0.96-1.03)	7.61E-01	8.66E-01	-0.055	0.097	0.95 (0.78-1.15)	5.69E-01	9.51E-01
106	M_VLDL_C	0.022	0.020	1.02 (0.98-1.06)	2.80E-01	4.68E-01	-0.054	0.101	0.95 (0.78-1.15)	5.92E-01	9.51E-01
107	M_VLDL_CE	0.039	0.020	1.04 (1-1.08)	4.74E-02	1.33E-01	-0.040	0.100	0.96 (0.79-1.17)	6.87E-01	9.60E-01
108	M_VLDL_FC	0.002	0.020	1 (0.96-1.04)	9.07E-01	9.49E-01	-0.059	0.099	0.94 (0.78-1.14)	5.50E-01	9.51E-01
109	M_VLDL_TG	-0.046	0.020	0.96 (0.92-0.99)	2.29E-02	8.79E-02	-0.012	0.097	0.99 (0.82-1.2)	9.03E-01	9.90E-01
110	S_VLDL_P	-0.017	0.020	0.98 (0.95-1.02)	4.16E-01	6.09E-01	-0.027	0.099	0.97 (0.8-1.18)	7.90E-01	9.82E-01
111	S_VLDL_L	-0.015	0.020	0.99 (0.95-1.03)	4.69E-01	6.50E-01	-0.030	0.100	0.97 (0.8-1.18)	7.65E-01	9.82E-01
112	S_VLDL_PL	0.000	0.020	1 (0.96-1.04)	9.90E-01	9.90E-01	-0.054	0.101	0.95 (0.78-1.16)	5.93E-01	9.51E-01
113	S_VLDL_C	0.013	0.020	1.01 (0.97-1.05)	5.36E-01	7.00E-01	-0.060	0.104	0.94 (0.77-1.15)	5.63E-01	9.51E-01
114	S_VLDL_CE	0.013	0.020	1.01 (0.97-1.05)	5.40E-01	7.00E-01	-0.059	0.105	0.94 (0.77-1.16)	5.73E-01	9.51E-01
115	S_VLDL_FC	0.012	0.020	1.01 (0.97-1.05)	5.39E-01	7.00E-01	-0.060	0.102	0.94 (0.77-1.15)	5.58E-01	9.51E-01
116	S_VLDL_TG	-0.040	0.020	0.96 (0.92-1)	4.92E-02	1.33E-01	0.018	0.097	1.02 (0.84-1.23)	8.52E-01	9.87E-01
117	XS_VLDL_P	0.042	0.020	1.04 (1-1.09)	4.06E-02	1.22E-01	0.002	0.102	1 (0.82-1.22)	9.83E-01	9.93E-01
118	XS_VLDL_L	0.039	0.020	1.04 (1-1.08)	5.31E-02	1.39E-01	0.014	0.101	1.01 (0.83-1.24)	8.90E-01	9.90E-01
119	XS_VLDL_PL	0.041	0.020	1.04 (1-1.08)	4.38E-02	1.26E-01	0.016	0.099	1.02 (0.84-1.24)	8.68E-01	9.90E-01
120	XS_VLDL_C	0.057	0.021	1.06 (1.02-1.1)	6.03E-03	3.95E-02	0.002	0.107	1 (0.81-1.24)	9.84E-01	9.93E-01
121	XS_VLDL_CE	0.061	0.021	1.06 (1.02-1.11)	3.57E-03	3.30E-02	0.002	0.108	1 (0.81-1.24)	9.83E-01	9.93E-01
122	XS_VLDL_FC	0.046	0.020	1.05 (1.01-1.09)	2.26E-02	8.79E-02	0.004	0.102	1 (0.82-1.23)	9.68E-01	9.93E-01
123	XS_VLDL_TG	-0.009	0.020	0.99 (0.95-1.03)	6.59E-01	7.89E-01	0.054	0.095	1.06 (0.88-1.27)	5.70E-01	9.51E-01

		The White ethnic group (n = 175,904)					The others (n = 8,660)				
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**
124	IDL_P	0.043	0.020	1.04 (1-1.09)	3.60E-02	1.12E-01	-0.105	0.109	0.9 (0.73-1.12)	3.36E-01	9.51E-01
125	IDL_L	0.045	0.021	1.05 (1.01-1.09)	2.88E-02	9.96E-02	-0.039	0.108	0.96 (0.78-1.19)	7.19E-01	9.82E-01
126	IDL_PL	0.050	0.021	1.05 (1.01-1.1)	1.54E-02	7.09E-02	-0.012	0.108	0.99 (0.8-1.22)	9.10E-01	9.90E-01
127	IDL_C	0.044	0.021	1.05 (1-1.09)	3.33E-02	1.08E-01	-0.059	0.109	0.94 (0.76-1.17)	5.90E-01	9.51E-01
128	IDL_CE	0.041	0.021	1.04 (1-1.09)	4.91E-02	1.33E-01	-0.054	0.109	0.95 (0.77-1.17)	6.16E-01	9.51E-01
129	IDL_FC	0.052	0.021	1.05 (1.01-1.1)	1.24E-02	6.43E-02	-0.069	0.110	0.93 (0.75-1.16)	5.33E-01	9.51E-01
130	IDL_TG	0.015	0.020	1.02 (0.98-1.06)	4.56E-01	6.41E-01	0.072	0.096	1.07 (0.89-1.3)	4.56E-01	9.51E-01
131	L_LDL_P	0.026	0.020	1.03 (0.99-1.07)	1.82E-01	3.41E-01	-0.068	0.104	0.94 (0.76-1.15)	5.16E-01	9.51E-01
132	L_LDL_L	0.023	0.020	1.02 (0.98-1.07)	2.50E-01	4.26E-01	-0.073	0.105	0.93 (0.76-1.14)	4.87E-01	9.51E-01
133	L_LDL_PL	0.025	0.020	1.03 (0.99-1.07)	2.14E-01	3.81E-01	-0.094	0.107	0.91 (0.74-1.12)	3.80E-01	9.51E-01
134	L_LDL_C	0.022	0.020	1.02 (0.98-1.06)	2.67E-01	4.52E-01	-0.077	0.105	0.93 (0.75-1.14)	4.64E-01	9.51E-01
135	L_LDL_CE	0.018	0.020	1.02 (0.98-1.06)	3.64E-01	5.52E-01	-0.071	0.104	0.93 (0.76-1.14)	4.98E-01	9.51E-01
136	L_LDL_FC	0.033	0.020	1.03 (0.99-1.08)	1.00E-01	2.21E-01	-0.093	0.107	0.91 (0.74-1.13)	3.88E-01	9.51E-01
137	L_LDL_TG	0.015	0.020	1.02 (0.98-1.06)	4.39E-01	6.32E-01	0.058	0.097	1.06 (0.88-1.28)	5.51E-01	9.51E-01
138	M_LDL_P	0.009	0.020	1.01 (0.97-1.05)	6.42E-01	7.84E-01	-0.048	0.107	0.95 (0.77-1.18)	6.53E-01	9.60E-01
139	M_LDL_L	0.004	0.020	1 (0.97-1.04)	8.54E-01	9.16E-01	-0.092	0.103	0.91 (0.75-1.12)	3.73E-01	9.51E-01
140	M_LDL_PL	0.009	0.020	1.01 (0.97-1.05)	6.51E-01	7.88E-01	-0.103	0.104	0.9 (0.74-1.11)	3.20E-01	9.51E-01
14	M_LDL_C	0.003	0.020	1 (0.96-1.04)	8.94E-01	9.44E-01	-0.096	0.103	0.91 (0.74-1.11)	3.52E-01	9.51E-01

		The White ethnic group (n = 175,904)					The others (n = 8,660)				
ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR	Beta	SE	HR (95% CI)	p value	FDR
1											
142	M_LDL_CE	-0.004	0.020	1 (0.96-1.04)	8.40E-01	9.14E-01	-0.086	0.102	0.92 (0.75-1.12)	4.00E-01	9.51E-01
143	M_LDL_FC	0.021	0.020	1.02 (0.98-1.06)	2.96E-01	4.76E-01	-0.117	0.105	0.89 (0.72-1.09)	2.67E-01	9.51E-01
144	M_LDL_TG	-0.008	0.020	0.99 (0.95-1.03)	6.94E-01	8.26E-01	0.032	0.096	1.03 (0.86-1.25)	7.41E-01	9.82E-01
145	S_LDL_P	0.014	0.020	1.01 (0.98-1.06)	4.91E-01	6.68E-01	-0.012	0.106	0.99 (0.8-1.22)	9.08E-01	9.90E-01
146	S_LDL_L	0.013	0.020	1.01 (0.98-1.05)	4.99E-01	6.72E-01	-0.070	0.103	0.93 (0.76-1.14)	4.93E-01	9.51E-01
147	S_LDL_PL	0.030	0.020	1.03 (0.99-1.07)	1.24E-01	2.58E-01	-0.052	0.101	0.95 (0.78-1.16)	6.09E-01	9.51E-01
148	S_LDL_C	0.011	0.020	1.01 (0.97-1.05)	5.95E-01	7.41E-01	-0.085	0.104	0.92 (0.75-1.13)	4.15E-01	9.51E-01
149	S_LDL_CE	0.002	0.020	1 (0.96-1.04)	9.28E-01	9.59E-01	-0.079	0.103	0.92 (0.76-1.13)	4.40E-01	9.51E-01
150	S_LDL_FC	0.032	0.020	1.03 (0.99-1.07)	1.02E-01	2.22E-01	-0.094	0.105	0.91 (0.74-1.12)	3.71E-01	9.51E-01
151	S_LDL_TG	-0.030	0.020	0.97 (0.93-1.01)	1.33E-01	2.70E-01	0.011	0.097	1.01 (0.84-1.22)	9.10E-01	9.90E-01
152	XL_HDL_P	0.046	0.023	1.05 (1-1.09)	4.41E-02	1.26E-01	0.243	0.123	1.28 (1-1.62)	4.86E-02	9.51E-01
153	XL_HDL_L	0.053	0.023	1.05 (1.01-1.1)	1.91E-02	7.80E-02	0.252	0.125	1.29 (1.01-1.65)	4.45E-02	9.51E-01
154	XL_HDL_PL	0.049	0.022	1.05 (1.01-1.1)	2.71E-02	9.78E-02	0.259	0.125	1.3 (1.02-1.65)	3.77E-02	9.51E-01
155	XL_HDL_C	0.059	0.022	1.06 (1.02-1.11)	9.12E-03	5.05E-02	0.230	0.128	1.26 (0.98-1.62)	7.24E-02	9.51E-01
156	XL_HDL_CE	0.050	0.023	1.05 (1.01-1.1)	2.67E-02	9.78E-02	0.239	0.128	1.27 (0.99-1.63)	6.23E-02	9.51E-01
157	XL_HDL_FC	0.082	0.022	1.09 (1.04-1.13)	1.74E-04	9.10E-03	0.191	0.125	1.21 (0.95-1.55)	1.28E-01	9.51E-01
158	XL_HDL_TG	-0.011	0.020	0.99 (0.95-1.03)	5.73E-01	7.21E-01	0.082	0.090	1.09 (0.91-1.3)	3.59E-01	9.51E-01

		The White ethnic group (n = 175,904)					The others (n = 8,660)				
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**
159	L_HDL_P	-0.003	0.023	1 (0.95-1.04)	8.86E-01	9.42E-01	0.234	0.124	1.26 (0.99-1.61)	5.92E-02	9.51E-01
160	L_HDL_L	-0.002	0.023	1 (0.95-1.04)	9.28E-01	9.59E-01	0.226	0.125	1.25 (0.98-1.6)	7.09E-02	9.51E-01
161	L_HDL_PL	-0.006	0.023	0.99 (0.95-1.04)	7.82E-01	8.85E-01	0.215	0.124	1.24 (0.97-1.58)	8.38E-02	9.51E-01
162	L_HDL_C	0.005	0.023	1.01 (0.96-1.05)	8.20E-01	9.07E-01	0.229	0.127	1.26 (0.98-1.61)	7.09E-02	9.51E-01
163	L_HDL_CE	0.003	0.023	1 (0.96-1.05)	9.02E-01	9.48E-01	0.223	0.127	1.25 (0.97-1.6)	8.09E-02	9.51E-01
164	L_HDL_FC	0.014	0.023	1.01 (0.97-1.06)	5.35E-01	7.00E-01	0.240	0.123	1.27 (1-1.62)	5.19E-02	9.51E-01
165	L_HDL_TG	-0.021	0.021	0.98 (0.94-1.02)	3.13E-01	4.93E-01	0.069	0.088	1.07 (0.9-1.27)	4.36E-01	9.51E-01
166	M_HDL_P	-0.059	0.022	0.94 (0.9-0.98)	6.51E-03	4.02E-02	0.056	0.116	1.06 (0.84-1.33)	6.32E-01	9.59E-01
167	M_HDL_L	-0.060	0.022	0.94 (0.9-0.98)	5.59E-03	3.87E-02	0.052	0.115	1.05 (0.84-1.32)	6.50E-01	9.60E-01
168	M_HDL_PL	-0.061	0.021	0.94 (0.9-0.98)	4.29E-03	3.34E-02	0.049	0.113	1.05 (0.84-1.31)	6.66E-01	9.60E-01
169	M_HDL_C	-0.053	0.022	0.95 (0.91-0.99)	1.56E-02	7.09E-02	0.051	0.119	1.05 (0.83-1.33)	6.71E-01	9.60E-01
170	M_HDL_CE	-0.054	0.022	0.95 (0.91-0.99)	1.31E-02	6.46E-02	0.043	0.119	1.04 (0.83-1.32)	7.16E-01	9.82E-01
171	M_HDL_FC	-0.046	0.022	0.96 (0.92-1)	3.67E-02	1.13E-01	0.075	0.116	1.08 (0.86-1.35)	5.17E-01	9.51E-01
172	M_HDL_TG	-0.040	0.020	0.96 (0.92-1)	5.15E-02	1.36E-01	0.010	0.092	1.01 (0.84-1.21)	9.11E-01	9.90E-01
173	S_HDL_P	-0.067	0.021	0.94 (0.9-0.97)	1.15E-03	1.68E-02	-0.118	0.109	0.89 (0.72-1.1)	2.82E-01	9.51E-01
174	S_HDL_L	-0.073	0.021	0.93 (0.89-0.97)	4.02E-04	9.10E-03	-0.075	0.109	0.93 (0.75-1.15)	4.90E-01	9.51E-01
175	S_HDL_PL	-0.075	0.021	0.93 (0.89-0.97)	2.99E-04	9.10E-03	-0.052	0.108	0.95 (0.77-1.17)	6.30E-01	9.59E-01
17	S_HDL_C	-0.064	0.021	0.94 (0.9-0.98)	1.94E-03	2.01E-02	-0.116	0.110	0.89 (0.72-1.11)	2.91E-01	9.51E-01

		The White ethnic group (n = 175,904)					The others (n = 8,660)				
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**
6											
177	S_HDL_CE	-0.067	0.020	0.94 (0.9-0.97)	1.07E-03	1.66E-02	-0.143	0.110	0.87 (0.7-1.08)	1.95E-01	9.51E-01
178	S_HDL_FC	-0.045	0.021	0.96 (0.92-1)	3.16E-02	1.06E-01	-0.020	0.107	0.98 (0.8-1.21)	8.49E-01	9.87E-01
179	S_HDL_TG	-0.038	0.021	0.96 (0.92-1)	6.91E-02	1.69E-01	-0.007	0.105	0.99 (0.81-1.22)	9.44E-01	9.93E-01
180	XXL_VLDL_PL_pct	-0.008	0.024	0.99 (0.95-1.04)	7.25E-01	8.43E-01	-0.098	0.079	0.91 (0.78-1.06)	2.17E-01	9.51E-01
181	XXL_VLDL_C_pct	0.054	0.020	1.06 (1.01-1.1)	8.03E-03	4.65E-02	0.100	0.090	1.11 (0.93-1.32)	2.67E-01	9.51E-01
182	XXL_VLDL_CE_pct	0.065	0.021	1.07 (1.03-1.11)	1.82E-03	2.01E-02	0.123	0.091	1.13 (0.95-1.35)	1.79E-01	9.51E-01
183	XXL_VLDL_FC_pct	0.021	0.021	1.02 (0.98-1.06)	3.01E-01	4.80E-01	0.046	0.090	1.05 (0.88-1.25)	6.07E-01	9.51E-01
184	XXL_VLDL_TG_pct	-0.030	0.018	0.97 (0.94-1.01)	9.39E-02	2.11E-01	0.017	0.082	1.02 (0.87-1.19)	8.30E-01	9.87E-01
185	XL_VLDL_PL_pct	0.016	0.023	1.02 (0.97-1.06)	4.95E-01	6.71E-01	-0.107	0.077	0.9 (0.77-1.05)	1.67E-01	9.51E-01
186	XL_VLDL_C_pct	0.068	0.021	1.07 (1.03-1.12)	1.42E-03	1.77E-02	-0.063	0.105	0.94 (0.77-1.15)	5.46E-01	9.51E-01
187	XL_VLDL_CE_pct	0.067	0.021	1.07 (1.03-1.12)	1.41E-03	1.77E-02	-0.033	0.103	0.97 (0.79-1.18)	7.47E-01	9.82E-01
188	XL_VLDL_FC_pct	0.058	0.021	1.06 (1.02-1.11)	6.62E-03	4.02E-02	-0.174	0.106	0.84 (0.68-1.04)	1.02E-01	9.51E-01
189	XL_VLDL_TG_pct	-0.068	0.021	0.94 (0.9-0.97)	1.29E-03	1.77E-02	0.127	0.109	1.14 (0.92-1.41)	2.40E-01	9.51E-01
190	L_VLDL_PL_pct	-0.010	0.023	0.99 (0.95-1.04)	6.59E-01	7.89E-01	-0.026	0.094	0.98 (0.81-1.17)	7.86E-01	9.82E-01
191	L_VLDL_C_pct	0.055	0.020	1.06 (1.02-1.1)	5.90E-03	3.95E-02	-0.109	0.103	0.9 (0.73-1.1)	2.91E-01	9.51E-01
192	L_VLDL_CE_pct	0.059	0.020	1.06 (1.02-1.1)	4.25E-03	3.34E-02	-0.078	0.105	0.93 (0.75-1.14)	4.55E-01	9.51E-01
193	L_VLDL_FC_pct	0.037	0.021	1.04 (1-1.08)	7.89E-02	1.84E-01	-0.143	0.101	0.87 (0.71-1.06)	1.58E-01	9.51E-01

		The White ethnic group (n = 175,904)					The others (n = 8,660)				
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**
194	L_VLDL_TG_pct	-0.043	0.020	0.96 (0.92-1)	3.02E-02	1.03E-01	0.095	0.099	1.1 (0.91-1.33)	3.35E-01	9.51E-01
195	M_VLDL_PL_pct	0.075	0.021	1.08 (1.04-1.12)	2.41E-04	9.10E-03	-0.097	0.101	0.91 (0.74-1.11)	3.41E-01	9.51E-01
196	M_VLDL_C_pct	0.073	0.021	1.08 (1.03-1.12)	3.78E-04	9.10E-03	-0.027	0.100	0.97 (0.8-1.19)	7.91E-01	9.82E-01
197	M_VLDL_CE_pct	0.074	0.020	1.08 (1.04-1.12)	2.76E-04	9.10E-03	-0.009	0.098	0.99 (0.82-1.2)	9.30E-01	9.93E-01
198	M_VLDL_FC_pct	0.073	0.021	1.08 (1.03-1.12)	3.61E-04	9.10E-03	-0.071	0.104	0.93 (0.76-1.14)	4.94E-01	9.51E-01
199	M_VLDL_TG_pct	-0.069	0.021	0.93 (0.9-0.97)	9.04E-04	1.59E-02	0.063	0.107	1.07 (0.86-1.31)	5.54E-01	9.51E-01
200	S_VLDL_PL_pct	0.058	0.021	1.06 (1.02-1.1)	5.17E-03	3.68E-02	-0.086	0.103	0.92 (0.75-1.12)	4.03E-01	9.51E-01
201	S_VLDL_C_pct	0.069	0.020	1.07 (1.03-1.12)	5.57E-04	1.15E-02	-0.072	0.099	0.93 (0.77-1.13)	4.69E-01	9.51E-01
202	S_VLDL_CE_pct	0.070	0.020	1.07 (1.03-1.12)	3.78E-04	9.10E-03	-0.072	0.097	0.93 (0.77-1.13)	4.62E-01	9.51E-01
203	S_VLDL_FC_pct	0.059	0.021	1.06 (1.02-1.11)	4.04E-03	3.34E-02	-0.063	0.103	0.94 (0.77-1.15)	5.40E-01	9.51E-01
204	S_VLDL_TG_pct	-0.061	0.020	0.94 (0.9-0.98)	2.70E-03	2.69E-02	0.102	0.104	1.11 (0.9-1.36)	3.27E-01	9.51E-01
205	XS_VLDL_PL_pct	0.022	0.021	1.02 (0.98-1.07)	2.92E-01	4.72E-01	0.030	0.106	1.03 (0.84-1.27)	7.75E-01	9.82E-01
206	XS_VLDL_C_pct	0.060	0.021	1.06 (1.02-1.11)	3.90E-03	3.34E-02	-0.034	0.102	0.97 (0.79-1.18)	7.38E-01	9.82E-01
207	XS_VLDL_CE_pct	0.059	0.021	1.06 (1.02-1.1)	4.85E-03	3.55E-02	-0.025	0.101	0.98 (0.8-1.19)	8.07E-01	9.82E-01
208	XS_VLDL_FC_pct	0.052	0.019	1.05 (1.01-1.09)	6.85E-03	4.06E-02	-0.068	0.101	0.94 (0.77-1.14)	5.04E-01	9.51E-01
209	XS_VLDL_TG_pct	-0.065	0.021	0.94 (0.9-0.98)	1.79E-03	2.01E-02	0.081	0.106	1.08 (0.88-1.33)	4.46E-01	9.51E-01
210	IDL_PL_pct	0.014	0.019	1.01 (0.98-1.05)	4.54E-01	6.41E-01	0.174	0.112	1.19 (0.96-1.48)	1.20E-01	9.51E-01
211	IDL_C_pct	0.016	0.019	1.02 (0.98-1.06)	4.15E-01	6.09E-01	-0.123	0.099	0.88 (0.73-1.07)	2.13E-01	9.51E-01

		The White ethnic group (n = 175,904)					The others (n = 8,660)				
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**
212	IDL_CE_pct	-0.001	0.019	1 (0.96-1.04)	9.79E-01	9.88E-01	-0.095	0.102	0.91 (0.75-1.11)	3.50E-01	9.51E-01
213	IDL_FC_pct	0.038	0.019	1.04 (1-1.08)	4.20E-02	1.23E-01	-0.105	0.095	0.9 (0.75-1.09)	2.73E-01	9.51E-01
214	IDL_TG_pct	-0.025	0.020	0.98 (0.94-1.02)	2.13E-01	3.81E-01	0.115	0.103	1.12 (0.92-1.37)	2.62E-01	9.51E-01
215	L_LDL_PL_pct	0.006	0.019	1.01 (0.97-1.04)	7.32E-01	8.43E-01	-0.107	0.108	0.9 (0.73-1.11)	3.22E-01	9.51E-01
216	L_LDL_C_pct	0.004	0.020	1 (0.97-1.05)	8.27E-01	9.07E-01	-0.080	0.099	0.92 (0.76-1.12)	4.19E-01	9.51E-01
217	L_LDL_CE_pct	-0.037	0.019	0.96 (0.93-1)	4.90E-02	1.33E-01	-0.021	0.100	0.98 (0.81-1.19)	8.35E-01	9.87E-01
218	L_LDL_FC_pct	0.047	0.020	1.05 (1.01-1.09)	1.86E-02	7.80E-02	-0.089	0.103	0.92 (0.75-1.12)	3.87E-01	9.51E-01
219	L_LDL_TG_pct	-0.007	0.020	0.99 (0.96-1.03)	7.35E-01	8.43E-01	0.137	0.100	1.15 (0.94-1.39)	1.71E-01	9.51E-01
220	M_LDL_PL_pct	0.041	0.020	1.04 (1-1.08)	3.41E-02	1.09E-01	-0.066	0.107	0.94 (0.76-1.15)	5.38E-01	9.51E-01
221	M_LDL_C_pct	-0.011	0.019	0.99 (0.95-1.03)	5.60E-01	7.12E-01	-0.097	0.101	0.91 (0.74-1.11)	3.36E-01	9.51E-01
222	M_LDL_CE_pct	-0.058	0.020	0.94 (0.91-0.98)	3.74E-03	3.33E-02	-0.017	0.104	0.98 (0.8-1.21)	8.72E-01	9.90E-01
223	M_LDL_FC_pct	0.051	0.020	1.05 (1.01-1.1)	1.16E-02	6.12E-02	-0.056	0.098	0.95 (0.78-1.15)	5.65E-01	9.51E-01
224	M_LDL_TG_pct	-0.014	0.020	0.99 (0.95-1.03)	4.85E-01	6.63E-01	0.143	0.099	1.15 (0.95-1.4)	1.51E-01	9.51E-01
225	S_LDL_PL_pct	0.070	0.021	1.07 (1.03-1.12)	6.97E-04	1.34E-02	0.102	0.111	1.11 (0.89-1.38)	3.55E-01	9.51E-01
226	S_LDL_C_pct	-0.018	0.020	0.98 (0.94-1.02)	3.66E-01	5.52E-01	-0.164	0.107	0.85 (0.69-1.05)	1.26E-01	9.51E-01
227	S_LDL_CE_pct	-0.063	0.020	0.94 (0.9-0.98)	1.92E-03	2.01E-02	-0.093	0.106	0.91 (0.74-1.12)	3.81E-01	9.51E-01
228	S_LDL_FC_pct	0.046	0.019	1.05 (1.01-1.09)	1.73E-02	7.62E-02	-0.054	0.095	0.95 (0.79-1.14)	5.70E-01	9.51E-01
22	S_LDL_TG_pct	-0.047	0.020	0.95 (0.92-0.99)	1.80E-02	7.72E-02	0.068	0.099	1.07 (0.88-1.3)	4.92E-01	9.51E-01

		The White ethnic group (n = 175,904)					The others (n = 8,660)				
ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR	Beta	SE	HR (95% CI)	p value	FDR
9											
230	XL_HDL_PL_pct	0.031	0.021	1.03 (0.99-1.07)	1.33E-01	2.70E-01	0.235	0.121	1.27 (1-1.6)	5.11E-02	9.51E-01
231	XL_HDL_C_pct	-0.005	0.020	1 (0.96-1.03)	8.03E-01	9.00E-01	-0.200	0.102	0.82 (0.67-1)	5.05E-02	9.51E-01
232	XL_HDL_CE_pct	-0.011	0.018	0.99 (0.95-1.03)	5.58E-01	7.12E-01	-0.077	0.095	0.93 (0.77-1.12)	4.18E-01	9.51E-01
233	XL_HDL_FC_pct	0.008	0.022	1.01 (0.97-1.05)	7.23E-01	8.43E-01	-0.248	0.117	0.78 (0.62-0.98)	3.43E-02	9.51E-01
234	XL_HDL_TG_pct	-0.049	0.021	0.95 (0.91-0.99)	1.99E-02	7.88E-02	-0.071	0.105	0.93 (0.76-1.14)	4.96E-01	9.51E-01
235	L_HDL_PL_pct	-0.030	0.020	0.97 (0.93-1.01)	1.45E-01	2.82E-01	-0.123	0.112	0.88 (0.71-1.1)	2.73E-01	9.51E-01
236	L_HDL_C_pct	0.025	0.020	1.03 (0.99-1.07)	2.02E-01	3.68E-01	0.114	0.114	1.12 (0.9-1.4)	3.16E-01	9.51E-01
237	L_HDL_CE_pct	0.014	0.019	1.01 (0.98-1.05)	4.78E-01	6.58E-01	0.073	0.111	1.08 (0.87-1.34)	5.08E-01	9.51E-01
238	L_HDL_FC_pct	0.068	0.021	1.07 (1.03-1.11)	9.60E-04	1.59E-02	0.195	0.107	1.22 (0.99-1.5)	6.87E-02	9.51E-01
239	L_HDL_TG_pct	-0.016	0.021	0.98 (0.95-1.03)	4.45E-01	6.35E-01	-0.057	0.092	0.95 (0.79-1.13)	5.39E-01	9.51E-01
240	M_HDL_PL_pct	0.005	0.021	1.01 (0.97-1.05)	8.11E-01	9.06E-01	-0.032	0.113	0.97 (0.78-1.21)	7.79E-01	9.82E-01
241	M_HDL_C_pct	0.003	0.020	1 (0.96-1.04)	8.93E-01	9.44E-01	0.004	0.107	1.01 (0.82-1.24)	9.66E-01	9.93E-01
242	M_HDL_CE_pct	0.001	0.020	1 (0.96-1.04)	9.41E-01	9.68E-01	-0.019	0.100	0.98 (0.81-1.19)	8.49E-01	9.87E-01
243	M_HDL_FC_pct	0.006	0.023	1.01 (0.96-1.05)	7.90E-01	8.90E-01	0.122	0.116	1.13 (0.9-1.42)	2.96E-01	9.51E-01
244	M_HDL_TG_pct	-0.007	0.021	0.99 (0.95-1.04)	7.39E-01	8.44E-01	-0.012	0.097	0.99 (0.82-1.2)	9.05E-01	9.90E-01
245	S_HDL_PL_pct	-0.024	0.021	0.98 (0.94-1.02)	2.43E-01	4.20E-01	0.120	0.106	1.13 (0.92-1.39)	2.55E-01	9.51E-01
246	S_HDL_C_pct	0.020	0.020	1.02 (0.98-1.06)	3.11E-01	4.93E-01	-0.093	0.097	0.91 (0.75-1.1)	3.38E-01	9.51E-01

		The White ethnic group (n = 175,904)					The others (n = 8,660)				
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**
247	S_HDL_CE_pct	0.001	0.019	1 (0.96-1.04)	9.45E-01	9.69E-01	-0.120	0.095	0.89 (0.74-1.07)	2.07E-01	9.51E-01
248	S_HDL_FC_pct	0.074	0.021	1.08 (1.03-1.12)	3.15E-04	9.10E-03	0.152	0.110	1.16 (0.94-1.45)	1.68E-01	9.51E-01
249	S_HDL_TG_pct	-0.008	0.021	0.99 (0.95-1.04)	7.11E-01	8.40E-01	0.025	0.107	1.03 (0.83-1.26)	8.12E-01	9.82E-01

SE, standard error. HR, hazard ratio. CI, confidence intervals. FDR, false discovery rate. Abbreviations are in line with **Supplementary Table 3.1**.

Supplementary Table 4.9 Association of individual metabolites with incident ischaemic stroke in model 7 with additional adjustment for blood pressure medication

ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR
1	Total_C	0.026	0.021	1.03 (0.99-1.07)	2.13E-01	3.60E-01
2	non_HDL_C	0.033	0.020	1.03 (0.99-1.08)	9.84E-02	2.03E-01
3	Remnant_C	0.036	0.020	1.04 (1-1.08)	7.30E-02	1.64E-01
4	VLDL_C	0.012	0.020	1.01 (0.97-1.05)	5.61E-01	6.93E-01
5	Clinical_LDL_C	0.039	0.020	1.04 (1-1.08)	5.34E-02	1.46E-01
6	LDL_C	0.029	0.020	1.03 (0.99-1.07)	1.45E-01	2.59E-01
7	HDL_C	-0.022	0.022	0.98 (0.94-1.02)	3.37E-01	4.88E-01
8	Total_TG	-0.037	0.020	0.96 (0.93-1)	6.43E-02	1.55E-01
9	VLDL_TG	-0.045	0.020	0.96 (0.92-1)	2.59E-02	8.17E-02
10	LDL_TG	0.011	0.020	1.01 (0.97-1.05)	5.65E-01	6.93E-01
11	HDL_TG	-0.031	0.020	0.97 (0.93-1.01)	1.21E-01	2.28E-01
12	Total_PL	-0.005	0.021	1 (0.96-1.04)	8.03E-01	8.77E-01
13	VLDL_PL	-0.010	0.020	0.99 (0.95-1.03)	6.09E-01	7.28E-01
14	LDL_PL	0.034	0.020	1.04 (1-1.08)	8.74E-02	1.86E-01
15	HDL_PL	-0.038	0.022	0.96 (0.92-1.01)	8.60E-02	1.86E-01
16	Total_CE	0.025	0.021	1.03 (0.99-1.07)	2.22E-01	3.71E-01
17	VLDL_CE	0.025	0.020	1.03 (0.99-1.07)	2.14E-01	3.60E-01
18	LDL_CE	0.024	0.020	1.02 (0.99-1.07)	2.36E-01	3.81E-01
19	HDL_CE	-0.026	0.022	0.97 (0.93-1.02)	2.40E-01	3.81E-01
20	Total_FC	0.026	0.020	1.03 (0.99-1.07)	1.99E-01	3.47E-01
21	VLDL_FC	-0.005	0.020	1 (0.96-1.03)	7.85E-01	8.70E-01
22	LDL_FC	0.043	0.020	1.04 (1-1.09)	3.13E-02	9.40E-02
23	HDL_FC	-0.003	0.023	1 (0.95-1.04)	9.02E-01	9.51E-01
24	Total_L	-0.001	0.020	1 (0.96-1.04)	9.41E-01	9.80E-01

ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR
25	VLDL_L	-0.024	0.020	0.98 (0.94-1.02)	2.34E-01	3.81E-01
26	LDL_L	0.030	0.020	1.03 (0.99-1.07)	1.34E-01	2.45E-01
27	HDL_L	-0.033	0.022	0.97 (0.93-1.01)	1.35E-01	2.46E-01
28	Total_P	-0.045	0.021	0.96 (0.92-1)	3.23E-02	9.57E-02
29	VLDL_P	0.006	0.020	1.01 (0.97-1.05)	7.71E-01	8.61E-01
30	LDL_P	0.036	0.020	1.04 (1-1.08)	7.00E-02	1.61E-01
31	HDL_P	-0.054	0.021	0.95 (0.91-0.99)	1.06E-02	4.12E-02
32	VLDL_size	-0.080	0.021	0.92 (0.89-0.96)	1.16E-04	3.21E-03
33	LDL_size	0.037	0.019	1.04 (1-1.08)	5.59E-02	1.49E-01
34	HDL_size	0.025	0.023	1.03 (0.98-1.07)	2.78E-01	4.25E-01
35	Phosphoglyc	-0.019	0.021	0.98 (0.94-1.02)	3.54E-01	5.06E-01
36	TG_by_PG	-0.036	0.021	0.97 (0.93-1.01)	9.09E-02	1.90E-01
37	Cholines	-0.014	0.021	0.99 (0.95-1.03)	4.98E-01	6.35E-01
38	Phosphatidylc	-0.016	0.021	0.98 (0.95-1.03)	4.55E-01	6.02E-01
39	Sphingomyelins	0.047	0.021	1.05 (1.01-1.09)	2.72E-02	8.47E-02
40	ApoB	0.038	0.020	1.04 (1-1.08)	5.93E-02	1.50E-01
41	ApoA1	-0.046	0.022	0.96 (0.92-1)	3.43E-02	1.01E-01
42	ApoB_by_ApoA1	0.056	0.020	1.06 (1.02-1.1)	5.70E-03	2.89E-02
43	Total_FA	-0.020	0.020	0.98 (0.94-1.02)	3.13E-01	4.64E-01
44	Unsaturation	-0.009	0.021	0.99 (0.95-1.03)	6.79E-01	7.90E-01
45	Omega_3	-0.037	0.021	0.96 (0.93-1)	7.24E-02	1.64E-01
46	Omega_6	-0.020	0.020	0.98 (0.94-1.02)	3.25E-01	4.73E-01
47	PUFA	-0.022	0.020	0.98 (0.94-1.02)	2.58E-01	4.07E-01
48	MUFA	-0.016	0.020	0.98 (0.95-1.02)	4.27E-01	5.74E-01
49	SFA	-0.018	0.020	0.98 (0.95-1.02)	3.72E-01	5.18E-01
50	LA	-0.020	0.020	0.98 (0.94-1.02)	3.02E-01	4.53E-01

ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR
51	DHA	-0.004	0.021	1 (0.96-1.04)	8.35E-01	9.01E-01
52	Omega_3_pct	-0.036	0.022	0.97 (0.93-1.01)	1.03E-01	2.08E-01
53	Omega_6_pct	0.010	0.019	1.01 (0.97-1.05)	6.11E-01	7.28E-01
54	PUFA_pct	0.004	0.020	1 (0.97-1.04)	8.36E-01	9.01E-01
55	MUFA_pct	-0.003	0.021	1 (0.96-1.04)	8.72E-01	9.28E-01
56	SFA_pct	-0.001	0.019	1 (0.96-1.04)	9.77E-01	9.89E-01
57	LA_pct	-0.001	0.020	1 (0.96-1.04)	9.60E-01	9.89E-01
58	DHA_pct	0.009	0.021	1.01 (0.97-1.05)	6.83E-01	7.92E-01
59	PUFA_by_MUFA	0.004	0.021	1 (0.96-1.05)	8.49E-01	9.11E-01
60	Omega_6_by_Omega_3	0.035	0.021	1.04 (0.99-1.08)	1.02E-01	2.08E-01
61	Ala	0.012	0.024	1.01 (0.97-1.06)	5.99E-01	7.20E-01
62	Gln	0.007	0.019	1.01 (0.97-1.05)	7.30E-01	8.27E-01
63	Gly	0.000	0.018	1 (0.97-1.04)	9.89E-01	9.93E-01
64	His	-0.040	0.019	0.96 (0.93-1)	3.00E-02	9.12E-02
65	Total_BCAA	-0.033	0.020	0.97 (0.93-1.01)	1.03E-01	2.08E-01
66	Ile	-0.014	0.020	0.99 (0.95-1.03)	4.85E-01	6.27E-01
67	Leu	-0.032	0.020	0.97 (0.93-1.01)	1.16E-01	2.23E-01
68	Val	-0.040	0.021	0.96 (0.92-1)	5.67E-02	1.49E-01
69	Phe	0.052	0.019	1.05 (1.01-1.09)	6.88E-03	3.11E-02
70	Tyr	0.028	0.020	1.03 (0.99-1.07)	1.51E-01	2.68E-01
71	Glucose	-0.018	0.019	0.98 (0.95-1.02)	3.58E-01	5.07E-01
72	Lactate	0.035	0.020	1.04 (1-1.08)	8.11E-02	1.79E-01
73	Pyruvate	0.017	0.020	1.02 (0.98-1.06)	3.90E-01	5.40E-01
74	Citrate	-0.011	0.021	0.99 (0.95-1.03)	5.81E-01	7.02E-01
75	bOHbutyrate	-0.003	0.021	1 (0.96-1.04)	8.98E-01	9.51E-01
76	Acetate	-0.011	0.017	0.99 (0.96-1.02)	5.08E-01	6.43E-01

ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR
77	Acetoacetate	-0.018	0.019	0.98 (0.95-1.02)	3.51E-01	5.05E-01
78	Acetone	0.000	0.019	1 (0.96-1.04)	9.88E-01	9.93E-01
79	Creatinine	0.034	0.022	1.03 (0.99-1.08)	1.31E-01	2.42E-01
80	Albumin	-0.139	0.019	0.87 (0.84-0.9)	9.51E-14	2.37E-11
81	GlycA	0.039	0.021	1.04 (1-1.08)	6.40E-02	1.55E-01
82	XXL_VLDL_P	-0.049	0.021	0.95 (0.91-0.99)	1.84E-02	6.53E-02
83	XXL_VLDL_L	-0.056	0.020	0.95 (0.91-0.98)	5.91E-03	2.89E-02
84	XXL_VLDL_PL	-0.050	0.021	0.95 (0.91-0.99)	1.77E-02	6.38E-02
85	XXL_VLDL_C	-0.048	0.021	0.95 (0.92-0.99)	2.25E-02	7.27E-02
86	XXL_VLDL_CE	-0.036	0.021	0.96 (0.93-1.01)	8.59E-02	1.86E-01
87	XXL_VLDL_FC	-0.061	0.021	0.94 (0.9-0.98)	2.99E-03	1.85E-02
88	XXL_VLDL_TG	-0.048	0.019	0.95 (0.92-0.99)	9.25E-03	3.78E-02
89	XL_VLDL_P	-0.040	0.021	0.96 (0.92-1)	5.95E-02	1.50E-01
90	XL_VLDL_L	-0.041	0.021	0.96 (0.92-1)	5.01E-02	1.43E-01
91	XL_VLDL_PL	-0.036	0.021	0.97 (0.93-1.01)	9.02E-02	1.90E-01
92	XL_VLDL_C	-0.023	0.021	0.98 (0.94-1.02)	2.71E-01	4.16E-01
93	XL_VLDL_CE	-0.012	0.021	0.99 (0.95-1.03)	5.68E-01	6.93E-01
94	XL_VLDL_FC	-0.033	0.021	0.97 (0.93-1.01)	1.08E-01	2.13E-01
95	XL_VLDL_TG	-0.049	0.021	0.95 (0.91-0.99)	1.95E-02	6.75E-02
96	L_VLDL_P	-0.036	0.021	0.96 (0.93-1)	7.72E-02	1.72E-01
97	L_VLDL_L	-0.038	0.021	0.96 (0.92-1)	6.12E-02	1.52E-01
98	L_VLDL_PL	-0.036	0.021	0.97 (0.93-1.01)	9.14E-02	1.90E-01
99	L_VLDL_C	-0.026	0.021	0.97 (0.94-1.02)	2.10E-01	3.59E-01
100	L_VLDL_CE	-0.017	0.021	0.98 (0.94-1.02)	4.02E-01	5.50E-01
101	L_VLDL_FC	-0.034	0.021	0.97 (0.93-1.01)	1.09E-01	2.13E-01
102	L_VLDL_TG	-0.045	0.020	0.96 (0.92-1)	2.59E-02	8.17E-02

ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR
103	M_VLDL_P	-0.001	0.020	1 (0.96-1.04)	9.66E-01	9.89E-01
104	M_VLDL_L	-0.011	0.020	0.99 (0.95-1.03)	5.76E-01	6.99E-01
105	M_VLDL_PL	0.006	0.020	1.01 (0.97-1.05)	7.60E-01	8.53E-01
106	M_VLDL_C	0.037	0.020	1.04 (1-1.08)	6.58E-02	1.56E-01
107	M_VLDL_CE	0.055	0.020	1.06 (1.02-1.1)	5.73E-03	2.89E-02
108	M_VLDL_FC	0.015	0.020	1.02 (0.98-1.06)	4.43E-01	5.93E-01
109	M_VLDL_TG	-0.039	0.020	0.96 (0.93-1)	5.12E-02	1.45E-01
110	S_VLDL_P	-0.008	0.020	0.99 (0.95-1.03)	7.02E-01	8.09E-01
111	S_VLDL_L	-0.005	0.020	1 (0.96-1.03)	7.86E-01	8.70E-01
112	S_VLDL_PL	0.012	0.020	1.01 (0.97-1.05)	5.62E-01	6.93E-01
113	S_VLDL_C	0.024	0.020	1.02 (0.98-1.07)	2.37E-01	3.81E-01
114	S_VLDL_CE	0.023	0.020	1.02 (0.98-1.06)	2.64E-01	4.08E-01
115	S_VLDL_FC	0.025	0.020	1.03 (0.99-1.07)	2.06E-01	3.53E-01
116	S_VLDL_TG	-0.034	0.020	0.97 (0.93-1.01)	8.65E-02	1.86E-01
117	XS_VLDL_P	0.054	0.020	1.06 (1.02-1.1)	6.45E-03	3.03E-02
118	XS_VLDL_L	0.052	0.020	1.05 (1.01-1.1)	8.91E-03	3.76E-02
119	XS_VLDL_PL	0.053	0.020	1.05 (1.01-1.1)	7.89E-03	3.39E-02
120	XS_VLDL_C	0.072	0.020	1.08 (1.03-1.12)	4.07E-04	6.17E-03
121	XS_VLDL_CE	0.077	0.021	1.08 (1.04-1.12)	1.91E-04	3.79E-03
122	XS_VLDL_FC	0.060	0.020	1.06 (1.02-1.1)	2.90E-03	1.85E-02
123	XS_VLDL_TG	-0.002	0.020	1 (0.96-1.04)	9.09E-01	9.51E-01
124	IDL_P	0.056	0.020	1.06 (1.02-1.1)	5.84E-03	2.89E-02
125	IDL_L	0.061	0.021	1.06 (1.02-1.11)	2.87E-03	1.85E-02
126	IDL_PL	0.068	0.021	1.07 (1.03-1.11)	1.03E-03	9.12E-03
127	IDL_C	0.059	0.021	1.06 (1.02-1.11)	3.90E-03	2.16E-02
128	IDL_CE	0.056	0.021	1.06 (1.02-1.1)	6.11E-03	2.93E-02

ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR
129	IDL_FC	0.065	0.020	1.07 (1.03-1.11)	1.42E-03	1.11E-02
130	IDL_TG	0.023	0.020	1.02 (0.99-1.06)	2.39E-01	3.81E-01
131	L_LDL_P	0.041	0.020	1.04 (1-1.08)	3.76E-02	1.09E-01
132	L_LDL_L	0.037	0.020	1.04 (1-1.08)	6.65E-02	1.56E-01
133	L_LDL_PL	0.038	0.020	1.04 (1-1.08)	5.89E-02	1.50E-01
134	L_LDL_C	0.036	0.020	1.04 (1-1.08)	7.20E-02	1.64E-01
135	L_LDL_CE	0.032	0.020	1.03 (0.99-1.07)	1.11E-01	2.16E-01
136	L_LDL_FC	0.047	0.020	1.05 (1.01-1.09)	2.02E-02	6.90E-02
137	L_LDL_TG	0.023	0.020	1.02 (0.99-1.06)	2.34E-01	3.81E-01
138	M_LDL_P	0.022	0.020	1.02 (0.98-1.06)	2.63E-01	4.08E-01
139	M_LDL_L	0.015	0.020	1.02 (0.98-1.06)	4.57E-01	6.02E-01
140	M_LDL_PL	0.020	0.020	1.02 (0.98-1.06)	3.19E-01	4.69E-01
141	M_LDL_C	0.014	0.020	1.01 (0.98-1.06)	4.86E-01	6.27E-01
142	M_LDL_CE	0.007	0.020	1.01 (0.97-1.05)	7.29E-01	8.27E-01
143	M_LDL_FC	0.032	0.020	1.03 (0.99-1.07)	1.06E-01	2.12E-01
144	M_LDL_TG	-0.001	0.020	1 (0.96-1.04)	9.66E-01	9.89E-01
145	S_LDL_P	0.029	0.020	1.03 (0.99-1.07)	1.43E-01	2.57E-01
146	S_LDL_L	0.027	0.020	1.03 (0.99-1.07)	1.73E-01	3.05E-01
147	S_LDL_PL	0.045	0.020	1.05 (1.01-1.09)	2.10E-02	7.06E-02
148	S_LDL_C	0.024	0.020	1.02 (0.99-1.07)	2.35E-01	3.81E-01
149	S_LDL_CE	0.014	0.020	1.01 (0.98-1.06)	4.75E-01	6.19E-01
150	S_LDL_FC	0.045	0.020	1.05 (1.01-1.09)	2.23E-02	7.27E-02
151	S_LDL_TG	-0.024	0.020	0.98 (0.94-1.02)	2.29E-01	3.80E-01
152	XL_HDL_P	0.066	0.022	1.07 (1.02-1.12)	3.24E-03	1.87E-02
153	XL_HDL_L	0.070	0.022	1.07 (1.03-1.12)	1.54E-03	1.16E-02
154	XL_HDL_PL	0.065	0.022	1.07 (1.02-1.11)	3.05E-03	1.85E-02

ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR
155	XL_HDL_C	0.076	0.022	1.08 (1.03-1.13)	5.58E-04	6.95E-03
156	XL_HDL_CE	0.068	0.022	1.07 (1.03-1.12)	2.19E-03	1.51E-02
157	XL_HDL_FC	0.097	0.021	1.1 (1.06-1.15)	5.50E-06	5.12E-04
158	XL_HDL_TG	0.000	0.019	1 (0.96-1.04)	9.99E-01	9.99E-01
159	L_HDL_P	0.011	0.023	1.01 (0.97-1.06)	6.33E-01	7.47E-01
160	L_HDL_L	0.010	0.023	1.01 (0.97-1.06)	6.65E-01	7.77E-01
161	L_HDL_PL	0.004	0.023	1 (0.96-1.05)	8.72E-01	9.28E-01
162	L_HDL_C	0.018	0.022	1.02 (0.98-1.06)	4.20E-01	5.68E-01
163	L_HDL_CE	0.015	0.022	1.02 (0.97-1.06)	5.00E-01	6.35E-01
164	L_HDL_FC	0.029	0.022	1.03 (0.99-1.08)	2.02E-01	3.49E-01
165	L_HDL_TG	-0.014	0.020	0.99 (0.95-1.03)	4.74E-01	6.19E-01
166	M_HDL_P	-0.056	0.021	0.95 (0.91-0.99)	9.13E-03	3.78E-02
167	M_HDL_L	-0.057	0.021	0.94 (0.91-0.98)	6.79E-03	3.11E-02
168	M_HDL_PL	-0.059	0.021	0.94 (0.9-0.98)	4.71E-03	2.55E-02
169	M_HDL_C	-0.050	0.021	0.95 (0.91-0.99)	1.93E-02	6.75E-02
170	M_HDL_CE	-0.052	0.021	0.95 (0.91-0.99)	1.47E-02	5.37E-02
171	M_HDL_FC	-0.040	0.022	0.96 (0.92-1)	6.34E-02	1.55E-01
172	M_HDL_TG	-0.039	0.020	0.96 (0.93-1)	5.24E-02	1.45E-01
173	S_HDL_P	-0.066	0.020	0.94 (0.9-0.97)	9.94E-04	9.12E-03
174	S_HDL_L	-0.073	0.020	0.93 (0.89-0.97)	3.07E-04	5.09E-03
175	S_HDL_PL	-0.075	0.020	0.93 (0.89-0.97)	2.13E-04	3.79E-03
176	S_HDL_C	-0.064	0.020	0.94 (0.9-0.98)	1.58E-03	1.16E-02
177	S_HDL_CE	-0.069	0.020	0.93 (0.9-0.97)	6.05E-04	7.11E-03
178	S_HDL_FC	-0.039	0.020	0.96 (0.92-1)	5.67E-02	1.49E-01
179	S_HDL_TG	-0.038	0.021	0.96 (0.93-1)	6.76E-02	1.57E-01
180	XXL_VLDL_PL_pct	-0.017	0.023	0.98 (0.94-1.03)	4.49E-01	5.98E-01

ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR
181	XXL_VLDL_C_pct	0.058	0.020	1.06 (1.02-1.1)	3.41E-03	1.93E-02
182	XXL_VLDL_CE_pct	0.072	0.020	1.08 (1.03-1.12)	4.32E-04	6.17E-03
183	XXL_VLDL_FC_pct	0.022	0.020	1.02 (0.98-1.06)	2.87E-01	4.36E-01
184	XXL_VLDL_TG_pct	-0.027	0.017	0.97 (0.94-1.01)	1.19E-01	2.27E-01
185	XL_VLDL_PL_pct	0.014	0.023	1.01 (0.97-1.06)	5.27E-01	6.59E-01
186	XL_VLDL_C_pct	0.070	0.021	1.07 (1.03-1.12)	7.51E-04	8.13E-03
187	XL_VLDL_CE_pct	0.071	0.021	1.07 (1.03-1.12)	6.28E-04	7.11E-03
188	XL_VLDL_FC_pct	0.057	0.021	1.06 (1.02-1.1)	7.00E-03	3.11E-02
189	XL_VLDL_TG_pct	-0.067	0.021	0.94 (0.9-0.97)	1.18E-03	9.76E-03
190	L_VLDL_PL_pct	-0.010	0.022	0.99 (0.95-1.03)	6.46E-01	7.59E-01
191	L_VLDL_C_pct	0.055	0.020	1.06 (1.02-1.1)	4.84E-03	2.56E-02
192	L_VLDL_CE_pct	0.061	0.020	1.06 (1.02-1.11)	2.28E-03	1.53E-02
193	L_VLDL_FC_pct	0.031	0.020	1.03 (0.99-1.07)	1.31E-01	2.42E-01
194	L_VLDL_TG_pct	-0.042	0.019	0.96 (0.92-1)	2.98E-02	9.12E-02
195	M_VLDL_PL_pct	0.083	0.020	1.09 (1.05-1.13)	3.78E-05	1.86E-03
196	M_VLDL_C_pct	0.081	0.020	1.08 (1.04-1.13)	6.21E-05	2.05E-03
197	M_VLDL_CE_pct	0.082	0.020	1.09 (1.04-1.13)	4.49E-05	1.86E-03
198	M_VLDL_FC_pct	0.081	0.020	1.08 (1.04-1.13)	6.59E-05	2.05E-03
199	M_VLDL_TG_pct	-0.076	0.021	0.93 (0.89-0.97)	2.06E-04	3.79E-03
200	S_VLDL_PL_pct	0.065	0.020	1.07 (1.03-1.11)	1.38E-03	1.11E-02
201	S_VLDL_C_pct	0.075	0.020	1.08 (1.04-1.12)	1.51E-04	3.61E-03
202	S_VLDL_CE_pct	0.074	0.019	1.08 (1.04-1.12)	1.59E-04	3.61E-03
203	S_VLDL_FC_pct	0.068	0.020	1.07 (1.03-1.11)	8.66E-04	8.79E-03
204	S_VLDL_TG_pct	-0.066	0.020	0.94 (0.9-0.97)	1.00E-03	9.12E-03
205	XS_VLDL_PL_pct	0.017	0.021	1.02 (0.98-1.06)	3.97E-01	5.46E-01
206	XS_VLDL_C_pct	0.068	0.020	1.07 (1.03-1.11)	8.83E-04	8.79E-03

ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR
207	XS_VLDL_CE_pct	0.067	0.020	1.07 (1.03-1.11)	1.08E-03	9.25E-03
208	XS_VLDL_FC_pct	0.056	0.019	1.06 (1.02-1.1)	3.16E-03	1.87E-02
209	XS_VLDL_TG_pct	-0.072	0.021	0.93 (0.89-0.97)	5.11E-04	6.70E-03
210	IDL_PL_pct	0.017	0.019	1.02 (0.98-1.06)	3.71E-01	5.18E-01
211	IDL_C_pct	0.019	0.019	1.02 (0.98-1.06)	3.13E-01	4.64E-01
212	IDL_CE_pct	0.005	0.019	1.01 (0.97-1.04)	7.90E-01	8.71E-01
213	IDL_FC_pct	0.035	0.018	1.04 (1-1.07)	5.62E-02	1.49E-01
214	IDL_TG_pct	-0.030	0.020	0.97 (0.93-1.01)	1.31E-01	2.42E-01
215	L_LDL_PL_pct	-0.001	0.018	1 (0.96-1.04)	9.69E-01	9.89E-01
216	L_LDL_C_pct	0.014	0.020	1.01 (0.98-1.05)	4.92E-01	6.31E-01
217	L_LDL_CE_pct	-0.029	0.019	0.97 (0.94-1.01)	1.13E-01	2.18E-01
218	L_LDL_FC_pct	0.050	0.020	1.05 (1.01-1.09)	1.09E-02	4.12E-02
219	L_LDL_TG_pct	-0.011	0.020	0.99 (0.95-1.03)	5.66E-01	6.93E-01
220	M_LDL_PL_pct	0.036	0.019	1.04 (1-1.08)	5.87E-02	1.50E-01
221	M_LDL_C_pct	-0.005	0.019	1 (0.96-1.03)	8.00E-01	8.77E-01
222	M_LDL_CE_pct	-0.053	0.020	0.95 (0.91-0.99)	7.17E-03	3.13E-02
223	M_LDL_FC_pct	0.051	0.020	1.05 (1.01-1.09)	9.99E-03	4.01E-02
224	M_LDL_TG_pct	-0.017	0.020	0.98 (0.95-1.02)	3.73E-01	5.18E-01
225	S_LDL_PL_pct	0.071	0.020	1.07 (1.03-1.12)	4.46E-04	6.17E-03
226	S_LDL_C_pct	-0.016	0.020	0.98 (0.95-1.02)	4.14E-01	5.63E-01
227	S_LDL_CE_pct	-0.062	0.020	0.94 (0.9-0.98)	1.79E-03	1.27E-02
228	S_LDL_FC_pct	0.047	0.019	1.05 (1.01-1.09)	1.39E-02	5.16E-02
229	S_LDL_TG_pct	-0.050	0.020	0.95 (0.92-0.99)	1.02E-02	4.05E-02
230	XL_HDL_PL_pct	0.039	0.020	1.04 (1-1.08)	5.17E-02	1.45E-01
231	XL_HDL_C_pct	-0.013	0.019	0.99 (0.95-1.03)	5.18E-01	6.51E-01
232	XL_HDL_CE_pct	-0.009	0.018	0.99 (0.96-1.03)	6.16E-01	7.31E-01

ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR
233	XL_HDL_FC_pct	-0.007	0.022	0.99 (0.95-1.04)	7.60E-01	8.53E-01
234	XL_HDL_TG_pct	-0.052	0.020	0.95 (0.91-0.99)	1.09E-02	4.12E-02
235	L_HDL_PL_pct	-0.046	0.020	0.96 (0.92-0.99)	2.13E-02	7.07E-02
236	L_HDL_C_pct	0.036	0.019	1.04 (1-1.08)	6.56E-02	1.56E-01
237	L_HDL_CE_pct	0.021	0.019	1.02 (0.98-1.06)	2.64E-01	4.08E-01
238	L_HDL_FC_pct	0.087	0.020	1.09 (1.05-1.14)	2.01E-05	1.25E-03
239	L_HDL_TG_pct	-0.020	0.020	0.98 (0.94-1.02)	3.25E-01	4.73E-01
240	M_HDL_PL_pct	-0.001	0.021	1 (0.96-1.04)	9.52E-01	9.88E-01
241	M_HDL_C_pct	0.005	0.020	1.01 (0.97-1.05)	8.18E-01	8.89E-01
242	M_HDL_CE_pct	0.001	0.019	1 (0.96-1.04)	9.77E-01	9.89E-01
243	M_HDL_FC_pct	0.021	0.022	1.02 (0.98-1.07)	3.55E-01	5.06E-01
244	M_HDL_TG_pct	-0.008	0.021	0.99 (0.95-1.03)	7.08E-01	8.12E-01
245	S_HDL_PL_pct	-0.026	0.020	0.97 (0.94-1.01)	1.96E-01	3.43E-01
246	S_HDL_C_pct	0.020	0.019	1.02 (0.98-1.06)	2.94E-01	4.44E-01
247	S_HDL_CE_pct	-0.002	0.019	1 (0.96-1.04)	9.06E-01	9.51E-01
248	S_HDL_FC_pct	0.092	0.020	1.1 (1.05-1.14)	6.17E-06	5.12E-04
249	S_HDL_TG_pct	-0.007	0.021	0.99 (0.95-1.03)	7.29E-01	8.27E-01

SE, standard error. HR, hazard ratio. CI, confidence intervals. FDR, false discovery rate.

Supplementary Table 4.10 Association of individual metabolites with incident ischaemic stroke in Cox regression in UK Biobank when adjusting for the same confounding factors as other cohorts (n = 184,564)

		Adjusting for the same confounding factors as China Kadoorie Biobank					Adjusting for the same confounding factors as a consortium of 7 studies on metabolomics and stroke				
ID	Metabolite	Beta	SE	HR (95% CI)	p value	FDR	Beta	SE	HR (95% CI)	p value	FDR
1	Total_C	-0.023	0.018	0.98 (0.94-1.01)	2.00E-01	4.11E-01	0.037	0.020	1.04 (1-1.08)	6.47E-02	1.50E-01
2	non_HDL_C	-0.011	0.018	0.99 (0.96-1.03)	5.47E-01	7.09E-01	0.048	0.020	1.05 (1.01-1.09)	1.53E-02	6.13E-02
3	Remnant_C	-0.008	0.018	0.99 (0.96-1.03)	6.46E-01	7.85E-01	0.052	0.020	1.05 (1.01-1.1)	8.35E-03	4.00E-02
4	VLDL_C	-0.015	0.019	0.99 (0.95-1.02)	4.19E-01	6.10E-01	0.033	0.020	1.03 (0.99-1.07)	9.89E-02	2.00E-01
5	Clinical_LDL_C	-0.008	0.018	0.99 (0.96-1.03)	6.64E-01	7.88E-01	0.050	0.020	1.05 (1.01-1.09)	1.01E-02	4.58E-02
6	LDL_C	-0.013	0.018	0.99 (0.95-1.02)	4.72E-01	6.53E-01	0.041	0.019	1.04 (1-1.08)	3.30E-02	9.35E-02

		Adjusting for the same confounding factors as China Kadoorie Biobank					Adjusting for the same confounding factors as a consortium of 7 studies on metabolomics and stroke				
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**
7	HDL_C	-0.055	0.021	0.95 (0.91-0.99)	9.49E-03	7.02E-02	-0.026	0.021	0.98 (0.94-1.02)	2.25E-01	3.70E-01
8	Total_TG	-0.034	0.020	0.97 (0.93-1.01)	8.63E-02	2.79E-01	-0.015	0.020	0.99 (0.95-1.02)	4.45E-01	6.30E-01
9	VLDL_TG	-0.040	0.020	0.96 (0.92-1)	4.24E-02	1.89E-01	-0.024	0.020	0.98 (0.94-1.02)	2.37E-01	3.86E-01
10	LDL_TG	0.006	0.019	1.01 (0.97-1.04)	7.68E-01	8.74E-01	0.032	0.019	1.03 (1-1.07)	9.33E-02	1.95E-01
11	HDL_TG	-0.027	0.019	0.97 (0.94-1.01)	1.58E-01	3.72E-01	-0.010	0.019	0.99 (0.95-1.03)	5.96E-01	7.35E-01
12	Total_PL	-0.044	0.019	0.96 (0.92-0.99)	2.22E-02	1.20E-01	0.011	0.020	1.01 (0.97-1.05)	5.66E-01	7.20E-01
13	VLDL_PL	-0.023	0.019	0.98 (0.94-1.01)	2.25E-01	4.29E-01	0.012	0.020	1.01 (0.97-1.05)	5.31E-01	6.99E-01
14	LDL_PL	-0.008	0.018	0.99 (0.96-1.03)	6.75E-01	7.96E-01	0.048	0.019	1.05 (1.01-1.09)	1.41E-02	5.77E-02
15	HDL_PL	-0.061	0.021	0.94 (0.9-0.98)	3.31E-03	3.74E-02	-0.032	0.021	0.97 (0.93-1.01)	1.22E-01	2.35E-01
16	Total_CE	-0.024	0.018	0.98 (0.94-1.01)	1.85E-01	3.99E-01	0.035	0.020	1.04 (1-1.08)	7.38E-02	1.65E-01
17	VLDL_CE	-0.009	0.018	0.99 (0.96-1.03)	6.42E-01	7.85E-01	0.045	0.020	1.05 (1.01-1.09)	2.46E-02	8.61E-02
18	LDL_CE	-0.016	0.018	0.98 (0.95-1.02)	3.81E-01	5.68E-01	0.037	0.019	1.04 (1-1.08)	5.30E-02	1.31E-01
19	HDL_CE	-0.058	0.021	0.94 (0.91-0.98)	6.24E-03	5.36E-02	-0.030	0.021	0.97 (0.93-1.01)	1.51E-01	2.74E-01
20	Total_FC	-0.021	0.018	0.98 (0.95-1.02)	2.51E-01	4.59E-01	0.039	0.020	1.04 (1-1.08)	5.00E-02	1.28E-01
21	VLDL_FC	-0.024	0.019	0.98 (0.94-1.01)	2.17E-01	4.20E-01	0.017	0.020	1.02 (0.98-1.06)	3.99E-01	5.98E-01
22	LDL_FC	-0.005	0.018	1 (0.96-1.03)	7.89E-01	8.83E-01	0.050	0.019	1.05 (1.01-1.09)	9.55E-03	4.45E-02
23	HDL_FC	-0.042	0.021	0.96 (0.92-1)	5.20E-02	2.16E-01	-0.008	0.021	0.99 (0.95-1.04)	7.24E-01	8.23E-01
24	Total_L	-0.036	0.018	0.97 (0.93-1)	5.20E-02	2.16E-01	0.017	0.019	1.02 (0.98-1.06)	3.82E-01	5.80E-01
25	VLDL_L	-0.031	0.019	0.97 (0.93-1.01)	1.06E-01	2.98E-01	-0.001	0.020	1 (0.96-1.04)	9.52E-01	9.76E-01
26	LDL_L	-0.011	0.018	0.99 (0.96-1.03)	5.47E-01	7.09E-01	0.043	0.019	1.04 (1.01-1.09)	2.51E-02	8.62E-02
27	HDL_L	-0.061	0.021	0.94 (0.9-0.98)	3.78E-03	4.03E-02	-0.030	0.021	0.97 (0.93-1.01)	1.51E-01	2.74E-01
28	Total_P	-0.072	0.019	0.93 (0.9-0.97)	2.14E-04	5.80E-03	-0.027	0.019	0.97 (0.94-1.01)	1.73E-01	3.05E-01
29	VLDL_P	-0.014	0.019	0.99 (0.95-1.02)	4.46E-01	6.31E-01	0.026	0.020	1.03 (0.99-1.07)	1.89E-01	3.32E-01
30	LDL_P	-0.004	0.018	1 (0.96-1.03)	8.44E-01	9.18E-01	0.049	0.020	1.05 (1.01-1.09)	1.13E-02	4.92E-02
31	HDL_P	-0.076	0.020	0.93 (0.89-0.96)	1.15E-04	4.10E-03	-0.036	0.020	0.96 (0.93-1)	6.38E-02	1.50E-01
32	VLDL_size	-0.062	0.020	0.94 (0.9-0.98)	2.28E-03	2.84E-02	-0.057	0.020	0.95 (0.91-0.98)	4.98E-03	3.10E-02
33	LDL_size	0.014	0.019	1.01 (0.98-1.05)	4.49E-01	6.32E-01	0.028	0.019	1.03 (0.99-1.07)	1.40E-01	2.62E-01
34	HDL_size	-0.004	0.023	1 (0.95-1.04)	8.72E-01	9.35E-01	-0.001	0.023	1 (0.96-1.04)	9.48E-01	9.75E-01
35	Phosphoglyc	-0.051	0.019	0.95 (0.91-0.99)	8.13E-03	6.53E-02	-0.003	0.020	1 (0.96-1.04)	8.92E-01	9.47E-01
36	TG_by_PG	-0.014	0.021	0.99 (0.95-1.03)	4.95E-01	6.67E-01	-0.017	0.021	0.98 (0.94-1.02)	4.22E-01	6.16E-01
37	Cholines	-0.050	0.019	0.95 (0.92-0.99)	9.44E-03	7.02E-02	0.001	0.020	1 (0.96-1.04)	9.45E-01	9.75E-01
38	Phosphatidylc	-0.051	0.020	0.95 (0.92-0.99)	9.58E-03	7.02E-02	-0.004	0.020	1 (0.96-1.04)	8.39E-01	9.08E-01
39	Sphingomyelins	-0.005	0.019	1 (0.96-1.03)	7.88E-01	8.83E-01	0.055	0.020	1.06 (1.02-1.1)	6.59E-03	3.57E-02
40	ApoB	-0.003	0.018	1 (0.96-1.03)	8.75E-01	9.35E-01	0.052	0.020	1.05 (1.01-1.1)	7.83E-03	3.93E-02
41	ApoA1	-0.069	0.020	0.93 (0.9-0.97)	7.41E-04	1.15E-02	-0.035	0.020	0.97 (0.93-1.01)	8.39E-02	1.82E-01
42	ApoB_by_ApoA1	0.032	0.019	1.03 (1-1.07)	8.86E-02	2.82E-01	0.065	0.020	1.07 (1.03-1.11)	1.14E-03	1.05E-02
43	Total_FA	-0.038	0.019	0.96 (0.93-1)	4.36E-02	1.90E-01	0.004	0.019	1 (0.97-1.04)	8.25E-01	8.98E-01
44	Unsaturation	-0.037	0.019	0.96 (0.93-1)	5.46E-02	2.23E-01	-0.046	0.019	0.96 (0.92-0.99)	1.83E-02	7.24E-02
45	Omega_3	-0.062	0.019	0.94 (0.91-0.98)	1.08E-03	1.58E-02	-0.056	0.019	0.95 (0.91-0.98)	3.55E-03	2.39E-02
46	Omega_6	-0.048	0.019	0.95 (0.92-0.99)	1.09E-02	7.47E-02	-0.002	0.019	1 (0.96-1.04)	9.31E-01	9.75E-01
47	PUFA	-0.054	0.019	0.95 (0.91-0.98)	3.88E-03	4.03E-02	-0.013	0.019	0.99 (0.95-1.03)	4.90E-01	6.71E-01
48	MUFA	-0.019	0.019	0.98 (0.94-1.02)	3.15E-01	5.06E-01	0.015	0.019	1.02 (0.98-1.06)	4.28E-01	6.18E-01
49	SFA	-0.034	0.019	0.97 (0.93-1)	7.43E-02	2.53E-01	0.010	0.019	1.01 (0.97-1.05)	6.03E-01	7.36E-01

		Adjusting for the same confounding factors as China Kadoorie Biobank					Adjusting for the same confounding factors as a consortium of 7 studies on metabolomics and stroke				
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**
50	LA	-0.045	0.018	0.96 (0.92-0.99)	1.35E-02	8.20E-02	-0.005	0.019	1 (0.96-1.03)	7.91E-01	8.79E-01
51	DHA	-0.041	0.019	0.96 (0.93-1)	2.96E-02	1.44E-01	-0.037	0.019	0.96 (0.93-1)	4.67E-02	1.22E-01
52	Omega_3_pct	-0.052	0.019	0.95 (0.91-0.99)	5.90E-03	5.36E-02	-0.070	0.019	0.93 (0.9-0.97)	2.70E-04	5.68E-03
53	Omega_6_pct	0.001	0.019	1 (0.96-1.04)	9.78E-01	9.89E-01	-0.011	0.019	0.99 (0.95-1.03)	5.60E-01	7.15E-01
54	PUFA_pct	-0.015	0.019	0.99 (0.95-1.02)	4.40E-01	6.29E-01	-0.033	0.019	0.97 (0.93-1.01)	8.57E-02	1.84E-01
55	MUFA_pct	0.025	0.020	1.03 (0.99-1.07)	2.12E-01	4.11E-01	0.034	0.020	1.03 (0.99-1.08)	9.69E-02	1.98E-01
56	SFA_pct	-0.002	0.019	1 (0.96-1.04)	9.32E-01	9.61E-01	0.022	0.018	1.02 (0.99-1.06)	2.23E-01	3.70E-01
57	LA_pct	-0.017	0.019	0.98 (0.95-1.02)	3.52E-01	5.41E-01	-0.016	0.019	0.98 (0.95-1.02)	4.14E-01	6.12E-01
58	DHA_pct	-0.020	0.019	0.98 (0.95-1.02)	2.91E-01	4.99E-01	-0.041	0.019	0.96 (0.93-1)	3.10E-02	9.35E-02
59	PUFA_by_MUFA	-0.021	0.020	0.98 (0.94-1.02)	2.82E-01	4.88E-01	-0.035	0.020	0.97 (0.93-1)	7.73E-02	1.70E-01
60	Omega_6_by_Omega_3	0.048	0.019	1.05 (1.01-1.09)	1.03E-02	7.36E-02	0.062	0.019	1.06 (1.03-1.1)	1.21E-03	1.07E-02
61	Ala	0.011	0.019	1.01 (0.98-1.05)	5.49E-01	7.09E-01	-0.012	0.019	0.99 (0.95-1.03)	5.15E-01	6.87E-01
62	Gln	0.000	0.019	1 (0.96-1.04)	9.82E-01	9.89E-01	0.008	0.019	1.01 (0.97-1.05)	6.73E-01	7.83E-01
63	Gly	-0.001	0.018	1 (0.97-1.03)	9.36E-01	9.61E-01	-0.002	0.018	1 (0.96-1.03)	8.97E-01	9.47E-01
64	His	-0.046	0.018	0.96 (0.92-0.99)	1.18E-02	7.52E-02	-0.046	0.018	0.96 (0.92-0.99)	1.20E-02	5.15E-02
65	Total_BCAA	-0.027	0.020	0.97 (0.94-1.01)	1.73E-01	3.88E-01	-0.045	0.020	0.96 (0.92-0.99)	2.22E-02	8.37E-02
66	Ile	-0.007	0.019	0.99 (0.96-1.03)	6.99E-01	8.21E-01	-0.022	0.019	0.98 (0.94-1.02)	2.62E-01	4.18E-01
67	Leu	-0.026	0.020	0.97 (0.94-1.01)	1.83E-01	3.99E-01	-0.042	0.020	0.96 (0.92-1)	3.33E-02	9.35E-02
68	Val	-0.033	0.020	0.97 (0.93-1.01)	9.52E-02	2.93E-01	-0.053	0.020	0.95 (0.91-0.99)	7.89E-03	3.93E-02
69	Phe	0.053	0.019	1.06 (1.02-1.1)	5.01E-03	4.99E-02	0.040	0.019	1.04 (1-1.08)	3.32E-02	9.35E-02
70	Tyr	0.021	0.019	1.02 (0.98-1.06)	2.73E-01	4.82E-01	0.010	0.019	1.01 (0.97-1.05)	6.02E-01	7.36E-01
71	Glucose	0.013	0.019	1.01 (0.98-1.05)	4.81E-01	6.58E-01	0.001	0.018	1 (0.97-1.04)	9.73E-01	9.82E-01
72	Lactate	0.052	0.019	1.05 (1.02-1.09)	6.79E-03	5.64E-02	0.040	0.019	1.04 (1-1.08)	3.23E-02	9.35E-02
73	Pyruvate	0.024	0.020	1.02 (0.99-1.06)	2.26E-01	4.29E-01	0.028	0.019	1.03 (0.99-1.07)	1.42E-01	2.64E-01
74	Citrate	0.005	0.019	1.01 (0.97-1.04)	8.07E-01	8.93E-01	0.001	0.019	1 (0.96-1.04)	9.71E-01	9.82E-01
75	bOHbutyrate	-0.001	0.021	1 (0.96-1.04)	9.65E-01	9.81E-01	0.014	0.021	1.01 (0.97-1.06)	5.19E-01	6.88E-01
76	Acetate	-0.018	0.017	0.98 (0.95-1.02)	2.75E-01	4.82E-01	-0.018	0.016	0.98 (0.95-1.01)	2.67E-01	4.23E-01
77	Acetoacetate	-0.020	0.019	0.98 (0.95-1.02)	3.05E-01	5.06E-01	-0.013	0.019	0.99 (0.95-1.03)	4.99E-01	6.79E-01
78	Acetone	-0.005	0.019	1 (0.96-1.03)	7.91E-01	8.83E-01	-0.005	0.018	1 (0.96-1.03)	7.99E-01	8.84E-01
79	Creatinine	0.039	0.022	1.04 (1-1.09)	7.18E-02	2.53E-01	0.035	0.022	1.04 (0.99-1.08)	1.05E-01	2.10E-01
80	Albumin	-0.136	0.018	0.87 (0.84-0.9)	6.79E-14	1.69E-11	-0.127	0.018	0.88 (0.85-0.91)	2.63E-12	6.56E-10
81	GlycA	0.050	0.020	1.05 (1.01-1.09)	1.22E-02	7.62E-02	0.067	0.020	1.07 (1.03-1.11)	7.91E-04	8.21E-03
82	XXL_VLDL_P	-0.038	0.021	0.96 (0.92-1)	6.37E-02	2.37E-01	-0.026	0.021	0.98 (0.94-1.02)	2.16E-01	3.61E-01
83	XXL_VLDL_L	-0.045	0.020	0.96 (0.92-0.99)	2.40E-02	1.27E-01	-0.032	0.020	0.97 (0.93-1.01)	1.14E-01	2.21E-01
84	XXL_VLDL_PL	-0.039	0.021	0.96 (0.92-1)	6.60E-02	2.42E-01	-0.027	0.021	0.97 (0.93-1.02)	2.03E-01	3.46E-01
85	XXL_VLDL_C	-0.040	0.020	0.96 (0.92-1)	4.92E-02	2.11E-01	-0.019	0.021	0.98 (0.94-1.02)	3.56E-01	5.47E-01
86	XXL_VLDL_CE	-0.031	0.021	0.97 (0.93-1.01)	1.30E-01	3.38E-01	-0.006	0.021	0.99 (0.95-1.04)	7.67E-01	8.61E-01
87	XXL_VLDL_FC	-0.051	0.020	0.95 (0.91-0.99)	1.14E-02	7.47E-02	-0.036	0.020	0.97 (0.93-1)	8.12E-02	1.77E-01
88	XXL_VLDL_TG	-0.040	0.018	0.96 (0.93-1)	2.92E-02	1.44E-01	-0.033	0.019	0.97 (0.93-1)	7.37E-02	1.65E-01
89	XL_VLDL_P	-0.034	0.021	0.97 (0.93-1.01)	9.93E-02	2.94E-01	-0.015	0.021	0.99 (0.95-1.03)	4.77E-01	6.57E-01
90	XL_VLDL_L	-0.036	0.021	0.97 (0.93-1)	7.88E-02	2.62E-01	-0.017	0.021	0.98 (0.94-1.02)	4.23E-01	6.16E-01
91	XL_VLDL_PL	-0.033	0.021	0.97 (0.93-1.01)	1.11E-01	3.06E-01	-0.011	0.021	0.99 (0.95-1.03)	5.84E-01	7.30E-01
92	XL_VLDL_C	-0.029	0.020	0.97 (0.93-1.01)	1.52E-01	3.72E-01	0.003	0.021	1 (0.96-1.04)	8.88E-01	9.47E-01

		Adjusting for the same confounding factors as China Kadoorie Biobank					Adjusting for the same confounding factors as a consortium of 7 studies on metabolomics and stroke				
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**
93	XL_VLDL_CE	-0.024	0.020	0.98 (0.94-1.02)	2.33E-01	4.39E-01	0.015	0.021	1.02 (0.98-1.06)	4.77E-01	6.57E-01
94	XL_VLDL_FC	-0.033	0.020	0.97 (0.93-1.01)	1.01E-01	2.94E-01	-0.009	0.021	0.99 (0.95-1.03)	6.48E-01	7.68E-01
95	XL_VLDL_TG	-0.039	0.021	0.96 (0.92-1)	5.59E-02	2.24E-01	-0.026	0.021	0.97 (0.94-1.02)	2.11E-01	3.58E-01
96	L_VLDL_P	-0.035	0.020	0.97 (0.93-1)	8.16E-02	2.67E-01	-0.014	0.020	0.99 (0.95-1.03)	5.07E-01	6.86E-01
97	L_VLDL_L	-0.038	0.020	0.96 (0.93-1)	5.84E-02	2.31E-01	-0.017	0.020	0.98 (0.95-1.02)	4.16E-01	6.12E-01
98	L_VLDL_PL	-0.033	0.021	0.97 (0.93-1.01)	1.13E-01	3.06E-01	-0.012	0.021	0.99 (0.95-1.03)	5.81E-01	7.30E-01
99	L_VLDL_C	-0.032	0.020	0.97 (0.93-1.01)	1.12E-01	3.06E-01	-0.003	0.020	1 (0.96-1.04)	8.98E-01	9.47E-01
100	L_VLDL_CE	-0.029	0.020	0.97 (0.94-1.01)	1.37E-01	3.49E-01	0.004	0.020	1 (0.97-1.05)	8.25E-01	8.98E-01
101	L_VLDL_FC	-0.034	0.020	0.97 (0.93-1.01)	9.92E-02	2.94E-01	-0.009	0.021	0.99 (0.95-1.03)	6.60E-01	7.79E-01
102	L_VLDL_TG	-0.042	0.020	0.96 (0.92-1)	3.29E-02	1.57E-01	-0.026	0.020	0.98 (0.94-1.01)	1.98E-01	3.42E-01
103	M_VLDL_P	-0.026	0.019	0.97 (0.94-1.01)	1.62E-01	3.72E-01	0.018	0.020	1.02 (0.98-1.06)	3.51E-01	5.43E-01
104	M_VLDL_L	-0.031	0.019	0.97 (0.94-1.01)	1.04E-01	2.98E-01	0.008	0.020	1.01 (0.97-1.05)	6.81E-01	7.85E-01
105	M_VLDL_PL	-0.022	0.019	0.98 (0.94-1.02)	2.42E-01	4.50E-01	0.025	0.020	1.03 (0.99-1.07)	2.13E-01	3.58E-01
106	M_VLDL_C	-0.008	0.018	0.99 (0.96-1.03)	6.63E-01	7.88E-01	0.049	0.020	1.05 (1.01-1.09)	1.27E-02	5.36E-02
107	M_VLDL_CE	0.003	0.017	1 (0.97-1.04)	8.69E-01	9.35E-01	0.063	0.020	1.07 (1.03-1.11)	1.36E-03	1.17E-02
108	M_VLDL_FC	-0.018	0.018	0.98 (0.95-1.02)	3.36E-01	5.23E-01	0.032	0.020	1.03 (0.99-1.07)	1.06E-01	2.12E-01
109	M_VLDL_TG	-0.042	0.020	0.96 (0.92-1)	3.42E-02	1.60E-01	-0.019	0.020	0.98 (0.94-1.02)	3.28E-01	5.10E-01
110	S_VLDL_P	-0.019	0.019	0.98 (0.94-1.02)	3.19E-01	5.06E-01	0.012	0.020	1.01 (0.97-1.05)	5.51E-01	7.11E-01
111	S_VLDL_L	-0.020	0.019	0.98 (0.94-1.02)	3.03E-01	5.06E-01	0.015	0.020	1.02 (0.98-1.06)	4.58E-01	6.40E-01
112	S_VLDL_PL	-0.014	0.019	0.99 (0.95-1.02)	4.52E-01	6.33E-01	0.030	0.020	1.03 (0.99-1.07)	1.26E-01	2.39E-01
113	S_VLDL_C	-0.005	0.019	1 (0.96-1.03)	7.84E-01	8.83E-01	0.043	0.020	1.04 (1-1.09)	3.08E-02	9.35E-02
114	S_VLDL_CE	-0.003	0.019	1 (0.96-1.04)	8.84E-01	9.37E-01	0.043	0.020	1.04 (1-1.09)	3.12E-02	9.35E-02
115	S_VLDL_FC	-0.009	0.018	0.99 (0.96-1.03)	6.25E-01	7.78E-01	0.041	0.020	1.04 (1-1.08)	3.45E-02	9.44E-02
116	S_VLDL_TG	-0.031	0.020	0.97 (0.93-1.01)	1.22E-01	3.24E-01	-0.017	0.020	0.98 (0.95-1.02)	3.90E-01	5.89E-01
117	XS_VLDL_P	0.017	0.019	1.02 (0.98-1.06)	3.65E-01	5.48E-01	0.069	0.020	1.07 (1.03-1.11)	4.87E-04	7.51E-03
118	XS_VLDL_L	0.016	0.019	1.02 (0.98-1.05)	4.04E-01	5.92E-01	0.067	0.020	1.07 (1.03-1.11)	7.81E-04	8.21E-03
119	XS_VLDL_PL	0.021	0.019	1.02 (0.98-1.06)	2.69E-01	4.79E-01	0.068	0.020	1.07 (1.03-1.11)	6.01E-04	7.51E-03
120	XS_VLDL_C	0.019	0.018	1.02 (0.98-1.06)	3.08E-01	5.06E-01	0.082	0.020	1.09 (1.04-1.13)	4.97E-05	2.62E-03
121	XS_VLDL_CE	0.019	0.018	1.02 (0.98-1.06)	3.11E-01	5.06E-01	0.084	0.020	1.09 (1.05-1.13)	3.23E-05	2.62E-03
122	XS_VLDL_FC	0.019	0.019	1.02 (0.98-1.06)	3.14E-01	5.06E-01	0.073	0.020	1.08 (1.04-1.12)	2.20E-04	5.47E-03
123	XS_VLDL_TG	-0.003	0.020	1 (0.96-1.04)	8.84E-01	9.37E-01	0.015	0.020	1.02 (0.98-1.06)	4.32E-01	6.18E-01
124	IDL_P	0.006	0.018	1.01 (0.97-1.04)	7.44E-01	8.52E-01	0.068	0.020	1.07 (1.03-1.11)	6.28E-04	7.51E-03
125	IDL_L	0.004	0.018	1 (0.97-1.04)	8.42E-01	9.18E-01	0.070	0.020	1.07 (1.03-1.12)	5.33E-04	7.51E-03
126	IDL_PL	0.010	0.018	1.01 (0.97-1.05)	5.93E-01	7.45E-01	0.076	0.020	1.08 (1.04-1.12)	1.77E-04	5.47E-03
127	IDL_C	0.000	0.018	1 (0.97-1.04)	9.89E-01	9.89E-01	0.066	0.020	1.07 (1.03-1.11)	1.07E-03	1.04E-02
128	IDL_CE	-0.002	0.018	1 (0.96-1.03)	9.09E-01	9.50E-01	0.063	0.020	1.07 (1.02-1.11)	1.75E-03	1.46E-02
129	IDL_FC	0.007	0.018	1.01 (0.97-1.04)	7.16E-01	8.29E-01	0.071	0.020	1.07 (1.03-1.12)	3.71E-04	6.60E-03
130	IDL_TG	0.015	0.019	1.02 (0.98-1.05)	4.39E-01	6.29E-01	0.041	0.019	1.04 (1-1.08)	3.35E-02	9.35E-02
131	L_LDL_P	0.000	0.018	1 (0.97-1.04)	9.88E-01	9.89E-01	0.052	0.019	1.05 (1.01-1.09)	7.17E-03	3.73E-02
132	L_LDL_L	-0.008	0.018	0.99 (0.96-1.03)	6.49E-01	7.85E-01	0.048	0.019	1.05 (1.01-1.09)	1.39E-02	5.77E-02
133	L_LDL_PL	-0.008	0.018	0.99 (0.96-1.03)	6.48E-01	7.85E-01	0.051	0.020	1.05 (1.01-1.09)	9.65E-03	4.45E-02
134	L_LDL_C	-0.010	0.018	0.99 (0.96-1.03)	5.86E-01	7.40E-01	0.045	0.019	1.05 (1.01-1.09)	1.96E-02	7.64E-02
135	L_LDL_CE	-0.012	0.018	0.99 (0.95-1.02)	5.18E-01	6.86E-01	0.042	0.019	1.04 (1-1.08)	3.04E-02	9.35E-02

		Adjusting for the same confounding factors as China Kadoorie Biobank					Adjusting for the same confounding factors as a consortium of 7 studies on metabolomics and stroke				
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**
136	L_LDL_FC	-0.005	0.018	1 (0.96-1.03)	8.02E-01	8.92E-01	0.054	0.020	1.06 (1.02-1.1)	6.23E-03	3.57E-02
137	L_LDL_TG	0.015	0.019	1.02 (0.98-1.05)	4.31E-01	6.24E-01	0.044	0.019	1.05 (1.01-1.09)	2.32E-02	8.43E-02
138	M_LDL_P	-0.011	0.018	0.99 (0.95-1.03)	5.32E-01	7.01E-01	0.039	0.020	1.04 (1-1.08)	4.69E-02	1.22E-01
139	M_LDL_L	-0.018	0.018	0.98 (0.95-1.02)	3.30E-01	5.17E-01	0.033	0.019	1.03 (1-1.07)	9.14E-02	1.93E-01
140	M_LDL_PL	-0.013	0.018	0.99 (0.95-1.02)	4.79E-01	6.58E-01	0.037	0.019	1.04 (1-1.08)	5.90E-02	1.43E-01
141	M_LDL_C	-0.020	0.018	0.98 (0.95-1.02)	2.70E-01	4.79E-01	0.031	0.019	1.03 (0.99-1.07)	1.07E-01	2.12E-01
142	M_LDL_CE	-0.024	0.018	0.98 (0.94-1.01)	1.92E-01	3.99E-01	0.027	0.020	1.03 (0.99-1.07)	1.70E-01	3.03E-01
143	M_LDL_FC	-0.009	0.018	0.99 (0.96-1.03)	6.23E-01	7.78E-01	0.042	0.019	1.04 (1-1.08)	3.07E-02	9.35E-02
144	M_LDL_TG	-0.004	0.019	1 (0.96-1.04)	8.55E-01	9.26E-01	0.021	0.019	1.02 (0.98-1.06)	2.83E-01	4.46E-01
145	S_LDL_P	-0.004	0.018	1 (0.96-1.03)	8.32E-01	9.16E-01	0.044	0.020	1.05 (1.01-1.09)	2.36E-02	8.43E-02
146	S_LDL_L	-0.008	0.018	0.99 (0.96-1.03)	6.65E-01	7.88E-01	0.042	0.019	1.04 (1-1.08)	3.18E-02	9.35E-02
147	S_LDL_PL	0.008	0.018	1.01 (0.97-1.04)	6.64E-01	7.88E-01	0.054	0.019	1.06 (1.02-1.1)	5.29E-03	3.14E-02
148	S_LDL_C	-0.013	0.018	0.99 (0.95-1.02)	4.90E-01	6.67E-01	0.039	0.019	1.04 (1-1.08)	4.49E-02	1.20E-01
149	S_LDL_CE	-0.018	0.018	0.98 (0.95-1.02)	3.15E-01	5.06E-01	0.033	0.020	1.03 (1-1.07)	8.76E-02	1.87E-01
150	S_LDL_FC	0.004	0.018	1 (0.97-1.04)	8.38E-01	9.18E-01	0.049	0.019	1.05 (1.01-1.09)	1.06E-02	4.72E-02
151	S_LDL_TG	-0.022	0.019	0.98 (0.94-1.02)	2.47E-01	4.56E-01	-0.003	0.019	1 (0.96-1.04)	8.89E-01	9.47E-01
152	XL_HDL_P	0.019	0.021	1.02 (0.98-1.06)	3.66E-01	5.48E-01	0.043	0.022	1.04 (1-1.09)	5.21E-02	1.30E-01
153	XL_HDL_L	0.028	0.021	1.03 (0.99-1.07)	1.92E-01	3.99E-01	0.043	0.022	1.04 (1-1.09)	4.65E-02	1.22E-01
154	XL_HDL_PL	0.028	0.021	1.03 (0.99-1.07)	1.91E-01	3.99E-01	0.038	0.021	1.04 (1-1.08)	7.59E-02	1.69E-01
155	XL_HDL_C	0.030	0.021	1.03 (0.99-1.07)	1.63E-01	3.72E-01	0.049	0.022	1.05 (1.01-1.1)	2.37E-02	8.43E-02
156	XL_HDL_CE	0.020	0.021	1.02 (0.98-1.06)	3.38E-01	5.23E-01	0.041	0.022	1.04 (1-1.09)	6.08E-02	1.46E-01
157	XL_HDL_FC	0.055	0.020	1.06 (1.02-1.1)	6.16E-03	5.36E-02	0.072	0.021	1.08 (1.03-1.12)	6.03E-04	7.51E-03
158	XL_HDL_TG	-0.009	0.019	0.99 (0.96-1.03)	6.31E-01	7.82E-01	0.008	0.019	1.01 (0.97-1.05)	6.73E-01	7.83E-01
159	L_HDL_P	-0.029	0.022	0.97 (0.93-1.01)	1.88E-01	3.99E-01	-0.009	0.022	0.99 (0.95-1.04)	6.81E-01	7.85E-01
160	L_HDL_L	-0.026	0.022	0.97 (0.93-1.02)	2.38E-01	4.46E-01	-0.010	0.022	0.99 (0.95-1.03)	6.43E-01	7.66E-01
161	L_HDL_PL	-0.028	0.022	0.97 (0.93-1.02)	2.03E-01	4.11E-01	-0.014	0.022	0.99 (0.95-1.03)	5.16E-01	6.87E-01
162	L_HDL_C	-0.021	0.022	0.98 (0.94-1.02)	3.28E-01	5.16E-01	-0.005	0.022	1 (0.95-1.04)	8.24E-01	8.98E-01
163	L_HDL_CE	-0.023	0.022	0.98 (0.94-1.02)	2.80E-01	4.88E-01	-0.008	0.022	0.99 (0.95-1.04)	7.11E-01	8.12E-01
164	L_HDL_FC	-0.012	0.022	0.99 (0.95-1.03)	5.65E-01	7.22E-01	0.007	0.022	1.01 (0.97-1.05)	7.46E-01	8.45E-01
165	L_HDL_TG	-0.022	0.020	0.98 (0.94-1.02)	2.62E-01	4.76E-01	-0.008	0.020	0.99 (0.95-1.03)	6.92E-01	7.95E-01
166	M_HDL_P	-0.073	0.020	0.93 (0.89-0.97)	2.56E-04	5.80E-03	-0.044	0.020	0.96 (0.92-1)	2.60E-02	8.64E-02
167	M_HDL_L	-0.072	0.020	0.93 (0.9-0.97)	3.01E-04	5.83E-03	-0.044	0.020	0.96 (0.92-1)	2.56E-02	8.62E-02
168	M_HDL_PL	-0.071	0.020	0.93 (0.9-0.97)	3.04E-04	5.83E-03	-0.045	0.020	0.96 (0.92-0.99)	2.36E-02	8.43E-02
169	M_HDL_C	-0.069	0.020	0.93 (0.9-0.97)	6.21E-04	1.03E-02	-0.043	0.020	0.96 (0.92-1)	3.28E-02	9.35E-02
170	M_HDL_CE	-0.069	0.020	0.93 (0.9-0.97)	5.43E-04	9.65E-03	-0.045	0.020	0.96 (0.92-1)	2.54E-02	8.62E-02
171	M_HDL_FC	-0.064	0.020	0.94 (0.9-0.98)	1.49E-03	1.95E-02	-0.034	0.020	0.97 (0.93-1.01)	9.67E-02	1.98E-01
172	M_HDL_TG	-0.035	0.020	0.97 (0.93-1)	7.63E-02	2.57E-01	-0.016	0.020	0.98 (0.95-1.02)	4.06E-01	6.05E-01
173	S_HDL_P	-0.076	0.019	0.93 (0.89-0.96)	5.05E-05	2.52E-03	-0.034	0.019	0.97 (0.93-1)	6.55E-02	1.50E-01
174	S_HDL_L	-0.078	0.019	0.93 (0.89-0.96)	3.72E-05	2.52E-03	-0.038	0.019	0.96 (0.93-1)	4.11E-02	1.11E-01
175	S_HDL_PL	-0.079	0.019	0.92 (0.89-0.96)	3.30E-05	2.52E-03	-0.041	0.019	0.96 (0.93-1)	3.01E-02	9.35E-02
176	S_HDL_C	-0.074	0.019	0.93 (0.9-0.96)	7.58E-05	3.15E-03	-0.034	0.019	0.97 (0.93-1)	6.58E-02	1.50E-01
177	S_HDL_CE	-0.076	0.019	0.93 (0.89-0.96)	4.16E-05	2.52E-03	-0.039	0.018	0.96 (0.93-1)	3.38E-02	9.35E-02
178	S_HDL_FC	-0.057	0.019	0.94 (0.91-0.98)	2.39E-03	2.84E-02	-0.015	0.019	0.99 (0.95-1.02)	4.31E-01	6.18E-01

		Adjusting for the same confounding factors as China Kadoorie Biobank					Adjusting for the same confounding factors as a consortium of 7 studies on metabolomics and stroke				
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**
179	S_HDL_TG	-0.026	0.020	0.98 (0.94-1.01)	2.08E-01	4.11E-01	-0.010	0.020	0.99 (0.95-1.03)	6.15E-01	7.45E-01
180	XXL_VLDL_PL_pct	-0.009	0.023	0.99 (0.95-1.04)	7.04E-01	8.23E-01	-0.006	0.023	0.99 (0.95-1.04)	8.06E-01	8.88E-01
181	XXL_VLDL_C_pct	0.041	0.019	1.04 (1-1.08)	3.46E-02	1.60E-01	0.051	0.019	1.05 (1.01-1.09)	8.13E-03	3.97E-02
182	XXL_VLDL_CE_pct	0.049	0.019	1.05 (1.01-1.09)	1.12E-02	7.47E-02	0.066	0.019	1.07 (1.03-1.11)	6.33E-04	7.51E-03
183	XXL_VLDL_FC_pct	0.012	0.020	1.01 (0.97-1.05)	5.54E-01	7.11E-01	0.010	0.020	1.01 (0.97-1.05)	6.22E-01	7.48E-01
184	XXL_VLDL_TG_pct	-0.024	0.017	0.98 (0.94-1.01)	1.70E-01	3.84E-01	-0.025	0.017	0.98 (0.94-1.01)	1.52E-01	2.74E-01
185	XL_VLDL_PL_pct	0.002	0.021	1 (0.96-1.05)	9.26E-01	9.61E-01	0.026	0.022	1.03 (0.98-1.07)	2.47E-01	3.96E-01
186	XL_VLDL_C_pct	0.036	0.020	1.04 (1-1.08)	7.20E-02	2.53E-01	0.059	0.020	1.06 (1.02-1.11)	3.83E-03	2.51E-02
187	XL_VLDL_CE_pct	0.037	0.020	1.04 (1-1.08)	6.31E-02	2.37E-01	0.060	0.020	1.06 (1.02-1.11)	2.97E-03	2.22E-02
188	XL_VLDL_FC_pct	0.026	0.020	1.03 (0.99-1.07)	2.07E-01	4.11E-01	0.045	0.021	1.05 (1.01-1.09)	2.75E-02	9.01E-02
189	XL_VLDL_TG_pct	-0.037	0.020	0.96 (0.93-1)	6.34E-02	2.37E-01	-0.059	0.020	0.94 (0.91-0.98)	3.46E-03	2.39E-02
190	L_VLDL_PL_pct	0.002	0.022	1 (0.96-1.05)	9.38E-01	9.61E-01	0.013	0.022	1.01 (0.97-1.06)	5.48E-01	7.10E-01
191	L_VLDL_C_pct	0.030	0.019	1.03 (0.99-1.07)	1.20E-01	3.22E-01	0.054	0.019	1.06 (1.02-1.1)	5.25E-03	3.14E-02
192	L_VLDL_CE_pct	0.029	0.019	1.03 (0.99-1.07)	1.40E-01	3.51E-01	0.054	0.020	1.06 (1.02-1.1)	6.46E-03	3.57E-02
193	L_VLDL_FC_pct	0.027	0.020	1.03 (0.99-1.07)	1.81E-01	3.99E-01	0.047	0.020	1.05 (1.01-1.09)	2.08E-02	7.96E-02
194	L_VLDL_TG_pct	-0.028	0.019	0.97 (0.94-1.01)	1.34E-01	3.45E-01	-0.051	0.019	0.95 (0.92-0.99)	7.19E-03	3.73E-02
195	M_VLDL_PL_pct	0.031	0.018	1.03 (1-1.07)	8.94E-02	2.82E-01	0.084	0.020	1.09 (1.05-1.13)	2.67E-05	2.62E-03
196	M_VLDL_C_pct	0.032	0.019	1.03 (1-1.07)	9.10E-02	2.83E-01	0.073	0.020	1.08 (1.03-1.12)	2.74E-04	5.68E-03
197	M_VLDL_CE_pct	0.034	0.019	1.03 (1-1.07)	7.04E-02	2.53E-01	0.073	0.020	1.08 (1.04-1.12)	2.16E-04	5.47E-03
198	M_VLDL_FC_pct	0.031	0.019	1.03 (0.99-1.07)	1.01E-01	2.94E-01	0.076	0.020	1.08 (1.04-1.12)	1.56E-04	5.47E-03
199	M_VLDL_TG_pct	-0.031	0.019	0.97 (0.93-1.01)	1.06E-01	2.98E-01	-0.069	0.020	0.93 (0.9-0.97)	6.71E-04	7.59E-03
200	S_VLDL_PL_pct	0.019	0.019	1.02 (0.98-1.06)	3.12E-01	5.06E-01	0.059	0.020	1.06 (1.02-1.1)	3.46E-03	2.39E-02
201	S_VLDL_C_pct	0.033	0.018	1.03 (1-1.07)	7.38E-02	2.53E-01	0.075	0.020	1.08 (1.04-1.12)	1.27E-04	5.28E-03
202	S_VLDL_CE_pct	0.040	0.018	1.04 (1-1.08)	2.86E-02	1.44E-01	0.078	0.019	1.08 (1.04-1.12)	5.25E-05	2.62E-03
203	S_VLDL_FC_pct	0.019	0.019	1.02 (0.98-1.06)	3.02E-01	5.06E-01	0.062	0.020	1.06 (1.02-1.11)	2.24E-03	1.80E-02
204	S_VLDL_TG_pct	-0.027	0.019	0.97 (0.94-1.01)	1.57E-01	3.72E-01	-0.065	0.020	0.94 (0.9-0.97)	1.08E-03	1.04E-02
205	XS_VLDL_PL_pct	0.045	0.019	1.05 (1.01-1.09)	1.74E-02	9.65E-02	0.022	0.019	1.02 (0.99-1.06)	2.42E-01	3.92E-01
206	XS_VLDL_C_pct	0.016	0.019	1.02 (0.98-1.05)	4.00E-01	5.89E-01	0.057	0.020	1.06 (1.02-1.1)	4.51E-03	2.88E-02
207	XS_VLDL_CE_pct	0.014	0.019	1.01 (0.98-1.05)	4.63E-01	6.44E-01	0.055	0.020	1.06 (1.02-1.1)	6.39E-03	3.57E-02
208	XS_VLDL_FC_pct	0.025	0.018	1.03 (0.99-1.06)	1.59E-01	3.72E-01	0.056	0.019	1.06 (1.02-1.1)	3.03E-03	2.22E-02
209	XS_VLDL_TG_pct	-0.024	0.019	0.98 (0.94-1.01)	2.10E-01	4.11E-01	-0.062	0.021	0.94 (0.9-0.98)	2.43E-03	1.89E-02
210	IDL_PL_pct	0.034	0.018	1.03 (1-1.07)	6.18E-02	2.37E-01	0.011	0.018	1.01 (0.98-1.05)	5.59E-01	7.15E-01
211	IDL_C_pct	-0.019	0.018	0.98 (0.95-1.02)	2.94E-01	5.02E-01	0.013	0.019	1.01 (0.98-1.05)	4.76E-01	6.57E-01
212	IDL_CE_pct	-0.029	0.018	0.97 (0.94-1.01)	1.01E-01	2.94E-01	0.001	0.019	1 (0.97-1.04)	9.78E-01	9.82E-01
213	IDL_FC_pct	0.014	0.018	1.01 (0.98-1.05)	4.42E-01	6.29E-01	0.030	0.018	1.03 (1-1.07)	9.72E-02	1.98E-01
214	IDL_TG_pct	0.011	0.019	1.01 (0.97-1.05)	5.76E-01	7.32E-01	-0.021	0.020	0.98 (0.94-1.02)	3.02E-01	4.73E-01
215	L_LDL_PL_pct	0.001	0.018	1 (0.97-1.04)	9.46E-01	9.66E-01	0.005	0.018	1.01 (0.97-1.04)	7.75E-01	8.65E-01
216	L_LDL_C_pct	-0.026	0.019	0.97 (0.94-1.01)	1.56E-01	3.72E-01	0.000	0.019	1 (0.96-1.04)	9.81E-01	9.82E-01
217	L_LDL_CE_pct	-0.044	0.018	0.96 (0.92-0.99)	1.59E-02	9.00E-02	-0.033	0.018	0.97 (0.93-1)	6.51E-02	1.50E-01
218	L_LDL_FC_pct	0.013	0.019	1.01 (0.98-1.05)	5.02E-01	6.69E-01	0.038	0.020	1.04 (1-1.08)	5.14E-02	1.30E-01
219	L_LDL_TG_pct	0.025	0.019	1.03 (0.99-1.06)	1.84E-01	3.99E-01	-0.001	0.020	1 (0.96-1.04)	9.40E-01	9.75E-01
220	M_LDL_PL_pct	0.046	0.019	1.05 (1.01-1.09)	1.44E-02	8.35E-02	0.024	0.019	1.03 (0.99-1.06)	1.97E-01	3.42E-01
221	M_LDL_C_pct	-0.039	0.017	0.96 (0.93-1)	2.56E-02	1.33E-01	-0.004	0.019	1 (0.96-1.03)	8.50E-01	9.16E-01

		Adjusting for the same confounding factors as China Kadoorie Biobank					Adjusting for the same confounding factors as a consortium of 7 studies on metabolomics and stroke				
ID	**Metabolite**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**	**Beta**	**SE**	**HR (95% CI)**	**p value**	**FDR**
222	M_LDL_CE_pct	-0.062	0.019	0.94 (0.91-0.98)	1.26E-03	1.74E-02	-0.031	0.019	0.97 (0.93-1.01)	1.09E-01	2.14E-01
223	M_LDL_FC_pct	0.027	0.019	1.03 (0.99-1.07)	1.61E-01	3.72E-01	0.030	0.020	1.03 (0.99-1.07)	1.27E-01	2.40E-01
224	M_LDL_TG_pct	0.017	0.019	1.02 (0.98-1.06)	3.63E-01	5.48E-01	-0.012	0.019	0.99 (0.95-1.03)	5.44E-01	7.10E-01
225	S_LDL_PL_pct	0.071	0.019	1.07 (1.03-1.11)	1.88E-04	5.80E-03	0.036	0.019	1.04 (1-1.08)	5.66E-02	1.38E-01
226	S_LDL_C_pct	-0.044	0.018	0.96 (0.92-0.99)	1.41E-02	8.35E-02	0.000	0.019	1 (0.96-1.04)	9.82E-01	9.82E-01
227	S_LDL_CE_pct	-0.069	0.019	0.93 (0.9-0.97)	2.51E-04	5.80E-03	-0.028	0.019	0.97 (0.94-1.01)	1.46E-01	2.69E-01
228	S_LDL_FC_pct	0.026	0.019	1.03 (0.99-1.07)	1.58E-01	3.72E-01	0.029	0.019	1.03 (0.99-1.07)	1.26E-01	2.39E-01
229	S_LDL_TG_pct	-0.019	0.019	0.98 (0.95-1.02)	3.18E-01	5.06E-01	-0.038	0.019	0.96 (0.93-1)	5.22E-02	1.30E-01
230	XL_HDL_PL_pct	0.022	0.020	1.02 (0.98-1.06)	2.67E-01	4.79E-01	0.018	0.020	1.02 (0.98-1.06)	3.64E-01	5.56E-01
231	XL_HDL_C_pct	-0.007	0.019	0.99 (0.96-1.03)	7.31E-01	8.43E-01	-0.001	0.019	1 (0.96-1.04)	9.58E-01	9.78E-01
232	XL_HDL_CE_pct	-0.027	0.018	0.97 (0.94-1.01)	1.27E-01	3.32E-01	-0.009	0.018	0.99 (0.96-1.03)	5.90E-01	7.31E-01
233	XL_HDL_FC_pct	0.027	0.021	1.03 (0.99-1.07)	2.11E-01	4.11E-01	0.014	0.021	1.01 (0.97-1.06)	5.15E-01	6.87E-01
234	XL_HDL_TG_pct	-0.029	0.020	0.97 (0.93-1.01)	1.48E-01	3.68E-01	-0.026	0.020	0.97 (0.94-1.01)	2.00E-01	3.43E-01
235	L_HDL_PL_pct	-0.007	0.019	0.99 (0.96-1.03)	7.12E-01	8.28E-01	-0.028	0.020	0.97 (0.94-1.01)	1.67E-01	3.00E-01
236	L_HDL_C_pct	0.002	0.019	1 (0.97-1.04)	9.12E-01	9.50E-01	0.014	0.019	1.01 (0.98-1.05)	4.54E-01	6.39E-01
237	L_HDL_CE_pct	-0.006	0.019	0.99 (0.96-1.03)	7.46E-01	8.52E-01	0.001	0.019	1 (0.97-1.04)	9.42E-01	9.75E-01
238	L_HDL_FC_pct	0.039	0.019	1.04 (1-1.08)	4.13E-02	1.87E-01	0.070	0.020	1.07 (1.03-1.12)	4.84E-04	7.51E-03
239	L_HDL_TG_pct	0.002	0.020	1 (0.96-1.04)	9.10E-01	9.50E-01	0.002	0.020	1 (0.96-1.04)	9.35E-01	9.75E-01
240	M_HDL_PL_pct	0.026	0.020	1.03 (0.99-1.07)	1.91E-01	3.99E-01	0.011	0.020	1.01 (0.97-1.05)	5.87E-01	7.30E-01
241	M_HDL_C_pct	-0.017	0.019	0.98 (0.95-1.02)	3.85E-01	5.71E-01	-0.012	0.019	0.99 (0.95-1.03)	5.40E-01	7.07E-01
242	M_HDL_CE_pct	-0.013	0.019	0.99 (0.95-1.02)	4.94E-01	6.67E-01	-0.014	0.019	0.99 (0.95-1.02)	4.45E-01	6.30E-01
243	M_HDL_FC_pct	-0.026	0.021	0.97 (0.94-1.02)	2.08E-01	4.11E-01	0.006	0.021	1.01 (0.97-1.05)	7.62E-01	8.59E-01
244	M_HDL_TG_pct	0.009	0.020	1.01 (0.97-1.05)	6.47E-01	7.85E-01	0.011	0.020	1.01 (0.97-1.05)	5.78E-01	7.30E-01
245	S_HDL_PL_pct	-0.018	0.019	0.98 (0.95-1.02)	3.64E-01	5.48E-01	-0.023	0.019	0.98 (0.94-1.02)	2.36E-01	3.86E-01
246	S_HDL_C_pct	0.003	0.019	1 (0.97-1.04)	8.88E-01	9.37E-01	0.008	0.019	1.01 (0.97-1.05)	6.68E-01	7.83E-01
247	S_HDL_CE_pct	-0.011	0.019	0.99 (0.95-1.03)	5.50E-01	7.09E-01	-0.009	0.019	0.99 (0.96-1.03)	6.17E-01	7.45E-01
248	S_HDL_FC_pct	0.054	0.020	1.06 (1.02-1.1)	5.88E-03	5.36E-02	0.072	0.020	1.08 (1.03-1.12)	3.17E-04	6.06E-03
249	S_HDL_TG_pct	0.014	0.021	1.01 (0.97-1.06)	5.03E-01	6.69E-01	0.010	0.021	1.01 (0.97-1.05)	6.29E-01	7.53E-01

SE, standard error. HR, hazard ratio. CI, confidence intervals. FDR, false discovery rate. Abbreviations are in line with **Supplementary Table 3.1**.

Supplementary Table 5.1 Baseline characteristics of participants in the training and testing dataset for machine learning.

	All subjects	Training dataset	Testing dataset	p value*
Sample size	**(n = 184,564)**	**(n = 147,652)**	**(n = 36,912)**	
Incident ischaemic stroke, n (%)	3002 (1.6)	2418 (1.6)	584 (1.6)	0.465
Survival time (ischaemic stroke), median (Q1-Q3)	13.7 (13-14.4)	13.7 (13-14.4)	13.7 (13-14.4)	0.657
All-cause mortality, n (%)	12593 (6.8)	10036 (6.8)	2557 (6.9)	0.381
Survival time (all-cause mortality), median (Q1-Q3)	13.8 (13.1-14.5)	13.8 (13.1-14.5)	13.8 (13.1-14.5)	0.746
Age, median (Q1-Q3)	57 (49-63)	57 (49-63)	57 (49-63)	0.206
Male, n (%)	83236 (45.1)	66542 (45.1)	16694 (45.2)	0.585
Ethnicity, n (%)				0.803
White	175904 (95.3)	140710 (95.3)	35194 (95.3)	
Asian or Asian British	3214 (1.7)	2587 (1.8)	627 (1.7)	
Black or Black British	2319 (1.3)	1872 (1.3)	447 (1.2)	
Mixed	1022 (0.6)	817 (0.6)	205 (0.6)	
Chinese	564 (0.3)	449 (0.3)	115 (0.3)	
Other ethnic group	1541 (0.8)	1217 (0.8)	324 (0.9)	
BMI, kg/m2, median (Q1-Q3)	26.6 (24.1-29.7)	26.6 (24.1-29.7)	26.6 (24.1-29.7)	0.669
Alcohol intake, n (%)				0.74
Daily or almost daily	35711 (19.3)	28641 (19.4)	7070 (19.2)	
Three or four times a week	43999 (23.8)	35161 (23.8)	8838 (23.9)	
Once or twice a week	49710 (26.9)	39719 (26.9)	9991 (27.1)	
One to three times a month	21146 (11.5)	16960 (11.5)	4186 (11.3)	
Special occasions only	20494 (11.1)	16350 (11.1)	4144 (11.2)	
Non-drinker	13504 (7.3)	10821 (7.3)	2683 (7.3)	
Smoking status, n (%)				0.628
Never	104634 (56.7)	83787 (56.7)	20847 (56.5)	
Previous	61262 (33.2)	48937 (33.1)	12325 (33.4)	
Current	18668 (10.1)	14928 (10.1)	3740 (10.1)	
Physical activity based on International Physical Activity Questionnaire, n (%)				0.639
Low	34974 (18.9)	28038 (19)	6936 (18.8)	
Moderate	77597 (42)	62076 (42)	15521 (42)	
High	71993 (39)	57538 (39)	14455 (39.2)	
Systolic blood pressure, mmHg, median (Q1-Q3)	136 (124.5-149)	136 (124.5-149)	136 (124.5-149)	0.817
Diastolic blood pressure, mmHg, median (Q1-Q3)	82 (75.5-89)	82 (75.5-89)	82 (75.5-89)	0.532
Baseline hypertension, n (%)	43432 (23.5)	34785 (23.6)	8647 (23.4)	0.595
Blood pressure medication, n (%)	30034 (16.3)	24027 (16.3)	6007 (16.3)	1
Baseline diabetes mellitus, n (%)	7058 (3.8)	5643 (3.8)	1415 (3.8)	0.929

	All subjects	Training dataset	Testing dataset	p value*
Triglycerides, mmol/l, median (Q1-Q3)	1.5 (1.1-2.1)	1.5 (1.1-2.1)	1.5 (1.1-2.1)	0.886
LDL cholesterol, mmol/l, median (Q1-Q3)	3.6 (3.1-4.2)	3.6 (3.1-4.2)	3.6 (3.1-4.2)	0.214
HDL cholesterol, mmol/l, median (Q1-Q3)	1.4 (1.2-1.7)	1.4 (1.2-1.7)	1.4 (1.2-1.7)	0.211
Fasting time, hours, median (Q1-Q3)	3 (2-4)	3 (2-4)	3 (2-4)	0.836
Lipid modifying medication, n (%)	26135 (14.2)	20905 (14.2)	5230 (14.2)	0.965
Dietary supplements, n (%)	92574 (50.2)	74082 (50.2)	18492 (50.1)	0.799
Fresh fruit, n (%)				0.076
< 1 servings/day	16779 (9.1)	13302 (9)	3477 (9.4)	
≥ 1 and < 2 servings/day	48926 (26.5)	39121 (26.5)	9805 (26.6)	
≥ 2 and < 3 servings/day	52697 (28.6)	42257 (28.6)	10440 (28.3)	
≥ 3 servings/day	66162 (35.8)	52972 (35.9)	13190 (35.7)	
Whole grain foods, n (%)				0.306
0 time/day	51732 (28)	41448 (28.1)	10284 (27.9)	
> 0 and < 1 time/day	72328 (39.2)	57931 (39.2)	14397 (39)	
≥ 1 and < 3 times/day	60335 (32.7)	48133 (32.6)	12202 (33.1)	
≥ 3 times/day	169 (0.1)	140 (0.1)	29 (0.1)	
Refined grain foods, n (%)				0.309
0 time/day	3021 (1.6)	2425 (1.6)	596 (1.6)	
> 0 and < 1 time/day	41022 (22.2)	32686 (22.1)	8336 (22.6)	
≥ 1 and < 3 times/day	109868 (59.5)	88014 (59.6)	21854 (59.2)	
≥ 3 times/day	30653 (16.6)	24527 (16.6)	6126 (16.6)	
Caffeinated coffee, n (%)				0.261
0 cup/day	67830 (36.8)	54196 (36.7)	13634 (36.9)	
> 0 and < 2 cups/day	41691 (22.6)	33489 (22.7)	8202 (22.2)	
≥ 2 and < 4 cups/day	46872 (25.4)	37421 (25.3)	9451 (25.6)	
≥ 4 cups/day	28171 (15.3)	22546 (15.3)	5625 (15.2)	
Decaffeinated coffee, n (%)				0.948
0 cup/day	156381 (84.7)	125125 (84.7)	31256 (84.7)	
> 0 and < 3 cups/day	17079 (9.3)	13654 (9.2)	3425 (9.3)	
≥ 3 cups/day	11104 (6)	8873 (6)	2231 (6)	
Tea, n (%)				0.101
0 cup/day	26556 (14.4)	21138 (14.3)	5418 (14.7)	
> 0 and < 3 cups/day	46926 (25.4)	37487 (25.4)	9439 (25.6)	
≥ 3 and < 5 cups/day	54274 (29.4)	43585 (29.5)	10689 (29)	
≥ 5 cups/day	56808 (30.8)	45442 (30.8)	11366 (30.8)	
Vegetables servings, n (%)				0.692

	All subjects	Training dataset	Testing dataset	p value*
< 1 servings/day	11176 (6.1)	8941 (6.1)	2235 (6.1)	
≥ 1 and < 2 servings/day	53858 (29.2)	43136 (29.2)	10722 (29)	
≥ 2 and < 3 servings/day	62710 (34)	50213 (34)	12497 (33.9)	
≥ 3 servings/day	56820 (30.8)	45362 (30.7)	11458 (31)	
Cheese, n (%)				0.181
< 1 time/week	36124 (19.6)	28972 (19.6)	7152 (19.4)	
≥ 1 and < 2 times/week	39941 (21.6)	32016 (21.7)	7925 (21.5)	
≥ 2 and < 5 times/week	84305 (45.7)	67256 (45.6)	17049 (46.2)	
≥ 5 times/week	24194 (13.1)	19408 (13.1)	4786 (13)	
Unprocessed red meat, n (%)				0.902
0 time/week	12267 (6.6)	9795 (6.6)	2472 (6.7)	
< 1 time/week	6206 (3.4)	4970 (3.4)	1236 (3.3)	
≥ 1 and < 2 times/week	15075 (8.2)	12033 (8.1)	3042 (8.2)	
≥ 2 times/week	151016 (81.8)	120854 (81.9)	30162 (81.7)	
Processed meat, n (%)				0.942
0 time/week	16508 (8.9)	13220 (9)	3288 (8.9)	
< 1 time/week	55760 (30.2)	44582 (30.2)	11178 (30.3)	
≥ 1 and < 2 times/week	54535 (29.5)	43602 (29.5)	10933 (29.6)	
≥ 2 times/week	57761 (31.3)	46248 (31.3)	11513 (31.2)	
Oily fish, n (%)				0.487
0 time/week	20250 (11)	16228 (11)	4022 (10.9)	
< 1 time/week	62929 (34.1)	50221 (34)	12708 (34.4)	
≥ 1 and < 2 times/week	70213 (38)	56213 (38.1)	14000 (37.9)	
≥ 2 times/week	31172 (16.9)	24990 (16.9)	6182 (16.7)	
Non-oily fish, n (%)				0.619
0 time/week	8372 (4.5)	6676 (4.5)	1696 (4.6)	
< 1 time/week	54055 (29.3)	43219 (29.3)	10836 (29.4)	
≥ 1 and < 2 times/week	92025 (49.9)	73723 (49.9)	18302 (49.6)	
≥ 2 times/week	30112 (16.3)	24034 (16.3)	6078 (16.5)	
Spread type, n (%)				0.156
Never/rarely use spread	19694 (10.7)	15847 (10.7)	3847 (10.4)	
Butter/spreadable butter	67811 (36.7)	54143 (36.7)	13668 (37)	
Other type of spread/margarine	97059 (52.6)	77662 (52.6)	19397 (52.5)	

Body mass index; LDL, low-density lipoprotein; HDL, high-density lipoprotein. *Chi-squared test was used for categorical variables. Student's t-test or Wilcoxon rank-sum (Mann-Whitney) test were used for normally or non-normally distributed continuous variables.

Supplementary Table 5.2 Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) checklist

Section/Topic	Item	Checklist Item	Chapter
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	NA
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	Abstract
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	1.4, 2.3
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	1.6
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	3.1
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	3.1
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	3.1
	5b	Describe eligibility criteria for participants.	3.1
	5c	Give details of treatments received, if relevant.	NA
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	3.4
	6b	Report any actions to blind assessment of the outcome to be predicted.	NA
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	3.2, 3.5
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	NA
Sample size	8	Explain how the study size was arrived at.	3.1
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	3.6
Statistical analysis methods	10a	Describe how predictors were handled in the analyses.	6.2
	10b	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	6.2
	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	6.2
Risk groups	11	Provide details on how risk groups were created, if done.	NA
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	3.1
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	5.3
Model development	14a	Specify the number of participants and outcome events in each analysis.	6.2, 6.3
	14b	If done, report the unadjusted association between each candidate predictor and outcome.	4.4.3
Model specification	15a	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	6.3
	15b	Explain how to the use the prediction model.	6.3
Model performance	16	Report performance measures (with CIs) for the prediction model.	6.3
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	6.4
Interpretation	19b	Give an overall interpretation of the results, considering objectives, limitations, and results from similar studies, and other relevant evidence.	6.4
Implications	20	Discuss the potential clinical use of the model and implications for future research.	6.4
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	NA
Funding	22	Give the source of funding and the role of the funders for the present study.	NA

Supplementary Table 5.3 Details of hyperparameter optimization and model performance in the training dataset in XGBoost model.

Features	General parameters	Booster parameters						Learning task parameters	Evaluation metrics
	Booster	Maximum tree depth (3-10)	Boosting learning rate, eta (0.01-0.3)	Subsample ratio of the training instance (0.5-1)	Percentage of features will be used for building each tree (0.5-1)	Minimum sum of instance weight, hessian, needed in a child (0-10)	Number of boosted trees to fit (maximum 5000)	The learning task and the corresponding learning objective	Negative log-likelihood function
Age, sex, and 249 metabolites	gbtree	3	0.033222672	0.864451263	0.807033265	9	187	survival:cox	9.83282671
Age and sex		3	0.228630936	0.531546871	0.721240967	2	29		9.883217114
Age, sex, and metabolites* selected by Boruta algorithm		3	0.020781669	0.637571755	0.821034975	10	243		9.853438815
Age, sex, and significant metabolites** in the training dataset		3	0.166219098	0.96405003	0.538083873	6	36		9.864294512
FSRP#		3	0.102111306	0.892672307	0.565764778	9	63		9.822098078
FSRP# and metabolites* selected by Boruta algorithm		3	0.015039388	0.834174765	0.533487982	5	354		9.809502733
FSRP# and significant metabolites** in the training dataset		3	0.015039388	0.834174765	0.533487982	5	424		9.807688172

Framingham Stroke Risk Profile: age, sex, smoking status, prevalent diabetes mellitus, blood pressure medication, and systolic blood pressure.

* Age, sex, albumin, creatinine, glucose, ratio of polyunsaturated fatty acids to total fatty acids, ratio of linoleic acid to total fatty acids, omega-3 fatty acids, glycoprotein acetyls, and free cholesterol to total lipids ratio in very large HDL.

** In fully adjusted Cox regression model in the training dataset, there were 15 metabolites significantly related to incident ischaemic stroke, including albumin, free cholesterol to total lipids ratio in large HDL, free cholesterol in very large HDL, phospholipids to total lipids ratio in small LDL, average diameter for VLDL particles, triglycerides to total lipids ratio in very small VLDL, free cholesterol to total lipids ratio in small HDL, cholesteryl esters to total lipids ratio in small LDL, total lipids in small HDL, phospholipids in small HDL, phospholipids to total lipids ratio in medium VLDL, cholesterol to total lipids ratio in medium VLDL, cholesteryl esters to total lipids ratio in medium VLDL, free cholesterol to total lipids ratio in medium VLDL, and triglycerides to total lipids ratio in medium VLDL.

Supplementary Table 5.4 Details of hyperparameter optimization and evaluation metrics in the older group (age ≥ 50 years) in XGBoost model

Features	General parameters	Booster parameters						Learning task parameters	Evaluation metrics
	Booster	Maximum tree depth (3-10)	Boosting learning rate, eta (0.01-0.3)	Subsample ratio of the training instance (0.5-1)	Percentage of features will be used for building each tree (0.5-1)	Minimum sum of instance weight, hessian, needed in a child (0-10)	Number of boosted trees to fit (maximum 5000)	The learning task and the corresponding learning objective	Negative log-likelihood function
Age, sex, and 249 metabolites	gbtree	4	0.01727986	0.873046799	0.514813	1	291	survival:cox	9.620967593
Age and sex		3	0.129534318	0.587455852	0.50539	5	54		9.66432003
Age, sex, and metabolites* selected by Boruta algorithm		3	0.102111306	0.892672307	0.565765	9	51		9.853438815
FSRP#		3	0.026807983	0.697946344	0.532464	3	215		9.593841376
FSRP# and metabolites* selected by Boruta algorithm		3	0.015039388	0.834174765	0.533488	5	342		9.580451043

* Docosahexaenoic acid, ratio of polyunsaturated fatty acids to total fatty acids, ratio of linoleic acid to total fatty acids, glucose, lactate, creatinine, albumin, and glycoprotein acetyls.
# Framingham Stroke Risk Profile: age, sex, smoking status, prevalent diabetes mellitus, blood pressure medication, and systolic blood pressure.

Supplementary Table 5.5 Details of hyperparameter optimization and evaluation metrics in the male group in XGBoost model

Features	General parameters	Booster parameters						Learning task parameters	Evaluation metrics
	Booster	Maximum tree depth (3-10)	Boosting learning rate, eta (0.01-0.3)	Subsample ratio of the training instance (0.5-1)	Percentage of features will be used for building each tree (0.5-1)	Minimum sum of instance weight, hessian, needed in a child (0-10)	Number of boosted trees to fit (maximum 5000)	The learning task and the corresponding learning objective	Negative log-likelihood function
Age and 249 metabolites	gbtree	3	0.020781669	0.637571755	0.821034975	10	342	survival:cox	9.083708497
Age		3	0.088702415	0.944471475	0.773946647	8	64		9.116572027
Age and metabolites* selected by Boruta algorithm		3	0.020113693	0.669195642	0.957531881	3	164		9.099980232
FSRP#		3	0.015039388	0.834174765	0.533487982	5	336		9.054968583
FSRP# and metabolites* selected by Boruta algorithm		3	0.015039388	0.834174765	0.533487982	5	288		9.050676819

* Ratio of linoleic acid to total fatty acids, glucose, 3-Hydroxybutyrate, creatinine, albumin, and glycoprotein acetyls.
# Framingham Stroke Risk Profile: age, smoking status, prevalent diabetes mellitus, blood pressure medication, and systolic blood pressure.

Supplementary Table 5.6 Details of hyperparameter optimization and evaluation metrics in the female group in XGBoost model

Features	General parameters	Booster parameters						Learning task parameters	Evaluation metrics
	Booster	Maximum tree depth (3-10)	Boosting learning rate, eta (0.01-0.3)	Subsample ratio of the training instance (0.5-1)	Percentage of features will be used for building each tree (0.5-1)	Minimum sum of instance weight, hessian, needed in a child (0-10)	Number of boosted trees to fit (maximum 5000)	The learning task and the corresponding learning objective	Negative log-likelihood function
Age and 249 metabolites	gbtree	3	0.059378319	0.769299388	0.555246723	5	69	survival:cox	9.244404392
Age		3	0.022880952	0.994116934	0.694739677	7	230		9.300149029
Age and metabolites* selected by Boruta algorithm		3	0.123098386	0.785967657	0.608446381	6	44		9.249740905
FSRP#		3	0.09984811	0.679802186	0.681437152	4	45		9.237691353
FSRP# and metabolites* selected by Boruta algorithm		3	0.133878478	0.914119687	0.66763365	0	46		9.212728125

* Ratio of omega-6 fatty acids to total fatty acids, ratio of saturated fatty acids to total fatty acids, creatinine, albumin, glycoprotein acetyls, triglycerides in large LDL, and triglycerides to total lipids ratio in IDL.
# Framingham Stroke Risk Profile: age, smoking status, prevalent diabetes mellitus, blood pressure medication, and systolic blood pressure.

Supplementary Table 6.1 STROBE-MR checklist of recommended items to address in reports of Mendelian randomization studies

Item No.	Section	Checklist item	Chapter
		Title and Abstract	
1	Title and abstract	Indicate mendelian randomization (MR) as the study's design in the title and/or the abstract if that is a main purpose of the study.	Abstract
		Introduction	
2	Background	Explain the scientific background and rationale for the reported study. What is the exposure? Is a potential causal relationship between exposure and outcome plausible? Justify why MR is a helpful method to address the study question.	6.1
3	Objectives	State specific objectives clearly, including prespecified causal hypotheses (if any). State that MR is a method that, under specific assumptions, intends to estimate causal effects.	6.1
		Methods	
4	Study design and data sources	Present key elements of the study design early in the article. Consider including a table listing sources of data for all phases of the study. For each data source contributing to the analysis, describe the following:	
	a	Setting: Describe the study design and the underlying population, if possible. Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection, when available.	6.1
	b	Participants: Report the eligibility criteria and the sources and methods of selection of participants. Report the sample size and whether any power or sample size calculations were carried out prior to the main analysis.	6.2.2
	c	Describe measurement, quality control, and selection of genetic variants.	3.2, 3.3, 6.2.1, 6.2.2
	d	For each exposure, outcome, and other relevant variables, describe methods of assessment and diagnostic criteria for diseases.	3.2, 3.3, 3.4, 3.5
	e	Provide details of ethics committee approval and participant informed consent, if relevant.	3.1
5	Assumptions	Explicitly state the 3 core instrumental variable (IV) assumptions for the main analysis (relevance, independence, and exclusion restriction),	6.2.2, 6.2.3, 6.2.4

Item No.	Section	Checklist item	Chapter
		as well assumptions for any additional or sensitivity analysis.	
6	Statistical methods: main analysis	Describe statistical methods and statistics used.	
	a	Describe how quantitative variables were handled in the analyses (ie, scale, units, model).	3.2, 3.5
	b	Describe how genetic variants were handled in the analyses and, if applicable, how their weights were selected.	3.3, 6.2.1
	c	Describe the MR estimator (eg, 2-stage least squares, Wald ratio) and related statistics. Detail the included covariates and, in case of 2-sample MR, whether the same covariate set was used for adjustment in the 2 samples.	6.2.2
	d	Explain how missing data were addressed.	3.6
	e	If applicable, indicate how multiple testing was addressed.	6.2.2, 6.2.3, 6.2.4
7	Assessment of assumptions	Describe any methods or prior knowledge used to assess the assumptions or justify their validity.	6.2.2, 6.2.3, 6.2.4
8	Sensitivity analyses and additional analyses	Describe any sensitivity analyses or additional analyses performed (eg, comparison of effect estimates from different approaches, independent replication, bias analytic techniques, validation of instruments, simulations).	6.2.2, 6.2.3, 6.2.4
9	Software and preregistration		
	a	Name statistical software and package(s), including version and settings used.	6.2.2, 6.2.3, 6.2.4
	b	State whether the study protocol and details were preregistered (as well as when and where).	NA
Results			
10	Descriptive data		
	a	Report the numbers of individuals at each stage of included studies and reasons for exclusion. Consider use of a flow diagram.	6.2.1
	b	Report summary statistics for phenotypic exposure(s), outcome(s), and other relevant variables (eg, means, SDs, proportions).	6.3.1
	c	If the data sources include meta-analyses of previous studies, provide the assessments of heterogeneity across these studies.	NA
	d	For 2-sample MR:	6.4

Item No.	Section	Checklist item	Chapter
		i. Provide justification of the similarity of the genetic variant–exposure associations between the exposure and outcome samples. ii. Provide information on the number of individuals who overlap between the exposure and outcome studies.	
11	Main results		
	a	Report the associations between genetic variant and exposure and between genetic variant and outcome, preferably on an interpretable scale.	6.3.3
	b	Report MR estimates of the relationship between exposure and outcome and the measures of uncertainty from the MR analysis, on an interpretable scale, such as odds ratio or relative risk per SD difference.	6.3.3
	c	If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period.	NA
	d	Consider plots to visualize results (eg, forest plot, scatterplot of associations between genetic variants and outcome vs between genetic variants and exposure).	6.3.3
12	Assessment of assumptions		
	a	Report the assessment of the validity of the assumptions.	6.3.3
	b	Report any additional statistics (eg, assessments of heterogeneity across genetic variants, such as I^2, Q statistic, or E-value).	6.3.3
13	Sensitivity analyses and additional analyses		
	a	Report any sensitivity analyses to assess the robustness of the main results to violations of the assumptions.	6.3.4, 6.3.5
	b	Report results from other sensitivity analyses or additional analyses.	6.3.4, 6.3.5
	c	Report any assessment of the direction of the causal relationship (eg, bidirectional MR).	6.3.5
	d	When relevant, report and compare with estimates from non-MR analyses.	6.3.3
	e	Consider additional plots to visualize results (eg, leave-one-out analyses).	6.3.5
Discussion			

Item No.	Section	Checklist item	Chapter
14	Key results	Summarize key results with reference to study objectives.	6.4
15	Limitations	Discuss limitations of the study, taking into account the validity of the IV assumptions, other sources of potential bias, and imprecision. Discuss both direction and magnitude of any potential bias and any efforts to address them.	6.4
16	Interpretation		
	a	Meaning: Give a cautious overall interpretation of results in the context of their limitations and in comparison with other studies.	6.4
	b	Mechanism: Discuss underlying biological mechanisms that could drive a potential causal relationship between the investigated exposure and the outcome, and whether the gene-environment equivalence assumption is reasonable. Use causal language carefully, clarifying that IV estimates may provide causal effects only under certain assumptions.	6.4
	c	Clinical relevance: Discuss whether the results have clinical or public policy relevance, and to what extent they inform effect sizes of possible interventions.	6.4
17	Generalizability	Discuss the generalizability of the study results (a) to other populations, (b) across other exposure periods/timings, and (c) across other levels of exposure.	6.4
Other Information			
18	Funding	Describe sources of funding and the role of funders in the present study and, if applicable, sources of funding for the databases and original study or studies on which the present study is based.	NA
19	Data and data sharing	Provide the data used to perform all analyses or report where and how the data can be accessed, and reference these sources in the article. Provide the statistical code needed to reproduce the results in the article or report whether the code is publicly accessible and, if so, where.	NA
20	Conflicts of interest	All authors should declare all potential conflicts of interest.	NA

Supplementary Table 6.2 Genome-wide significant results for 47 metabolites

ID	SNP	Metabolite	Beta	SE	p value	A1	A2	EAF	Chr	Position	Distance*	Classification	IV	Novel variant
1	rs75406471	HDL_P	-0.031	0.004	1.26E-17	A	G	0.156	10	5257647	1329	cis	Yes	No
1	rs11239571	HDL_P	0.027	0.003	7.49E-19	A	G	0.242	10	46087892	2461	cis	Yes	No
1	rs16926246	HDL_P	-0.022	0.004	2.04E-08	T	C	0.130	10	71093392	30004	cis	Yes	No

ID	SNP	Metabolite	Beta	SE	p value	A1	A2	EAF	Chr	Position	Distance*	Classification	IV	Novel variant
1	rs6858	HDL_P	-0.015	0.003	4.58E-08	T	C	0.412	10	81142365	280	cis	Yes	No
1	rs2862954	HDL_P	0.019	0.003	1.93E-13	C	T	0.500	10	101912064	8311	cis	Yes	No
1	rs2792751	HDL_P	-0.037	0.003	2.75E-36	C	T	0.726	10	113940329	11233	cis	Yes	No
1	rs17672352	HDL_P	-0.025	0.003	1.23E-13	T	C	0.183	10	114028480	14677	cis	Yes	No
1	rs440150	HDL_P	0.031	0.005	9.78E-12	A	G	0.907	11	2936657	11612	cis	Yes	No
1	rs7118500	HDL_P	0.017	0.003	1.98E-10	A	C	0.554	11	10284762	9043	cis	Yes	No
1	rs77682558	HDL_P	0.042	0.007	4.19E-09	A	G	0.036	11	47276818	6360	cis	Yes	Yes
1	rs35184771	HDL_P	-0.034	0.003	6.40E-35	T	G	0.352	11	47475189	4493	cis	Yes	No
1	rs174581	HDL_P	-0.034	0.003	3.13E-35	A	G	0.355	11	61606683	9892	cis	Yes	No
1	rs72997616	HDL_P	-0.068	0.005	2.77E-50	A	C	0.093	11	75474195	3637	cis	Yes	No
1	rs12575636	HDL_P	-0.021	0.003	8.48E-10	G	T	0.188	11	95311260	7055	cis	Yes	No
1	rs117133249	HDL_P	0.041	0.007	1.96E-08	A	G	0.035	11	116524965	14825	cis	Yes	No
1	rs2542052	HDL_P	-0.024	0.003	3.81E-18	C	A	0.641	11	116699984	437	cis	No	Yes
1	rs12721030	HDL_P	0.074	0.003	3.97E-106	T	C	0.193	11	116705278	1554	cis	Yes	No
1	rs1622638	HDL_P	-0.015	0.003	2.33E-08	A	G	0.394	11	121800971	8167	cis	Yes	No
1	rs7134375	HDL_P	0.02	0.003	2.87E-13	A	C	0.431	12	20473758	14	cis	Yes	No
1	rs60318946	HDL_P	0.025	0.004	1.83E-09	G	A	0.118	12	21326728	42599	cis	No	Yes
1	rs3809114	HDL_P	0.016	0.003	4.18E-09	A	G	0.564	12	57848639	2532	cis	Yes	No
1	rs10774779	HDL_P	0.018	0.003	6.48E-11	A	T	0.373	12	110016952	5566	cis	Yes	No
1	rs653178	HDL_P	0.025	0.003	2.97E-21	T	C	0.517	12	112007756	5068	cis	Yes	No
1	rs113946477	HDL_P	0.035	0.005	8.41E-11	A	G	0.070	12	123759548	2666	cis	No	No
1	rs1054852	HDL_P	0.023	0.003	7.44E-16	G	A	0.376	12	124496316	5543	cis	Yes	No
1	rs3741521	HDL_P	-0.03	0.003	7.11E-25	A	G	0.672	12	125264293	7015	cis	Yes	No
1	rs7136506	HDL_P	-0.025	0.003	2.74E-14	C	T	0.216	12	125326153	41060	cis	Yes	No
1	rs9604045	HDL_P	0.019	0.003	1.73E-09	T	G	0.251	13	113927208	3722	cis	No	No
1	rs12878212	HDL_P	0.017	0.003	5.15E-09	A	T	0.672	14	24826199	4835	cis	No	No
1	rs2494747	HDL_P	-0.027	0.003	8.25E-24	T	G	0.617	14	105258437	3658	cis	Yes	No
1	rs7170463	HDL_P	0.017	0.003	8.04E-09	G	A	0.311	15	41888918	24413	cis	Yes	No
1	rs11854242	HDL_P	-0.018	0.003	8.93E-10	T	C	0.273	15	58571910	4658	cis	Yes	No
1	rs261290	HDL_P	-0.061	0.003	3.92E-107	C	T	0.653	15	58678720	24047	cis	Yes	No
1	rs633695	HDL_P	0.049	0.003	2.88E-62	G	A	0.292	15	58725839	20686	cis	Yes	No
1	rs450962	HDL_P	-0.02	0.003	3.42E-10	G	A	0.293	16	28413517	1644	cis	No	No
1	rs80072323	HDL_P	-0.081	0.005	7.87E-72	A	G	0.096	16	56892769	338	cis	Yes	No
1	rs117738782	HDL_P	-0.108	0.017	1.49E-10	C	T	0.007	16	56946502	6055	cis	No	Yes
1	rs9989419	HDL_P	0.104	0.003	1.00E-200	G	A	0.606	16	56985139	6037	cis	Yes	No
1	rs117250566	HDL_P	-0.089	0.009	1.07E-21	A	G	0.021	16	57224563	4513	cis	Yes	No
1	rs4986970	HDL_P	-0.087	0.007	8.89E-33	T	A	0.035	16	67976320	1713	cis	Yes	No
1	rs2925979	HDL_P	0.027	0.003	2.80E-20	C	T	0.699	16	81534790	16638	cis	Yes	No
1	rs4790310	HDL_P	-0.017	0.003	1.13E-10	T	C	0.574	17	1966920	8029	cis	Yes	Yes
1	rs2270445	HDL_P	0.016	0.003	1.98E-09	G	A	0.490	17	8219478	4970	cis	Yes	No
1	rs704	HDL_P	-0.022	0.003	1.89E-16	A	G	0.476	17	26694861	171	cis	Yes	No
1	rs2643194	HDL_P	0.021	0.003	7.26E-14	T	C	0.692	17	37853048	1	cis	Yes	No
1	rs72836561	HDL_P	-0.137	0.008	1.48E-73	T	C	0.032	17	41926126	1609	cis	Yes	No
1	rs34931250	HDL_P	-0.048	0.006	1.63E-18	T	C	0.062	17	66879927	39268	cis	Yes	No
1	rs4129767	HDL_P	0.029	0.003	2.30E-28	A	G	0.489	17	76403984	7967	cis	Yes	No
1	rs34250533	HDL_P	-0.026	0.004	3.66E-11	T	A	0.128	18	19565847	4942	cis	Yes	Yes
1	rs150237291	HDL_P	0.05	0.009	2.45E-08	C	T	0.022	18	46781014	48731	cis	Yes	No
1	rs77960347	HDL_P	0.301	0.011	1.29E-152	G	A	0.014	18	47109955	22885	cis	No	No
1	rs4939883	HDL_P	0.096	0.003	8.71E-171	C	T	0.820	18	47167214	9505	cis	Yes	No
1	rs34054923	HDL_P	0.035	0.004	6.61E-21	C	T	0.146	18	47272512	4078	cis	Yes	No

ID	SNP	Metabolite	Beta	SE	p value	A1	A2	EAF	Chr	Position	Distance*	Classification	IV	Novel variant
1	rs688671	HDL_P	-0.018	0.003	7.51E-10	G	A	0.269	18	57867526	3737	cis	Yes	No
1	rs12976739	HDL_P	0.024	0.003	1.55E-18	A	G	0.394	19	8461663	6093	cis	Yes	No
1	rs737338	HDL_P	-0.11	0.007	1.17E-51	T	C	0.034	19	11347657	520	cis	Yes	No
1	rs58542926	HDL_P	-0.046	0.005	1.05E-19	T	C	0.076	19	19379549	4628	cis	Yes	No
1	rs187429064	HDL_P	-0.068	0.012	1.22E-08	G	A	0.014	19	19380513	3664	cis	No	No
1	rs12721046	HDL_P	-0.032	0.004	5.98E-19	A	G	0.159	19	45421254	3749	cis	Yes	No
1	rs5167	HDL_P	0.035	0.003	8.89E-36	G	T	0.351	19	45448465	826	cis	Yes	No
1	rs1645788	HDL_P	0.033	0.003	2.93E-25	G	A	0.230	19	54808174	1248	cis	Yes	No
1	rs71576588	HDL_P	-0.016	0.003	2.48E-09	G	A	0.584	1	965009	9508	cis	No	No
1	rs71640337	HDL_P	-0.045	0.005	3.81E-21	A	G	0.088	1	27004388	14688	cis	Yes	No
1	rs3768321	HDL_P	-0.038	0.003	1.40E-29	T	G	0.198	1	40035928	2747	cis	No	No
1	rs12751443	HDL_P	-0.017	0.003	3.52E-09	T	C	0.302	1	46324191	10924	cis	No	Yes
1	rs11208004	HDL_P	-0.043	0.003	3.10E-54	A	G	0.351	1	63145439	8617	cis	Yes	No
1	rs57274629	HDL_P	0.018	0.003	4.04E-11	G	A	0.358	1	66110292	59333	cis	No	No
1	rs140584594	HDL_P	0.031	0.003	7.32E-26	G	A	0.730	1	110232983	2546	cis	No	No
1	rs11101992	HDL_P	-0.019	0.003	1.87E-09	C	A	0.243	1	110266754	1687	cis	Yes	No
1	rs198325	HDL_P	0.038	0.003	2.51E-33	T	C	0.222	1	150937405	315	cis	Yes	No
1	rs4656292	HDL_P	0.024	0.003	2.99E-18	G	A	0.620	1	161194641	1087	cis	Yes	No
1	rs10913592	HDL_P	0.019	0.003	1.49E-11	A	G	0.380	1	178631012	8605	cis	Yes	No
1	rs7529037	HDL_P	-0.017	0.003	1.48E-09	C	G	0.351	1	205637501	10515	cis	Yes	Yes
1	rs2642438	HDL_P	0.03	0.003	2.20E-24	G	A	0.703	1	220970028	9695	cis	Yes	No
1	rs2281718	HDL_P	0.044	0.003	3.37E-59	T	A	0.614	1	230297778	42301	cis	Yes	No
1	rs484084	HDL_P	0.016	0.003	5.55E-09	C	T	0.520	1	234857676	2112	cis	Yes	No
1	rs1132274	HDL_P	-0.021	0.004	3.66E-09	A	C	0.156	20	17596155	15702	cis	Yes	No
1	rs2425026	HDL_P	-0.017	0.003	1.31E-10	T	C	0.428	20	33847253	4736	cis	Yes	Yes
1	rs1800961	HDL_P	-0.138	0.008	5.97E-73	T	C	0.031	20	43042364	5603	cis	Yes	No
1	rs3817731	HDL_P	-0.019	0.003	3.12E-12	T	A	0.627	20	44518843	59	cis	Yes	Yes
1	rs4239651	HDL_P	0.026	0.003	1.13E-15	C	T	0.792	20	46340596	33950	cis	Yes	No
1	rs3859588	HDL_P	-0.019	0.003	4.10E-09	A	T	0.216	20	46476143	613	cis	Yes	No
1	rs235374	HDL_P	-0.018	0.003	1.48E-11	C	G	0.435	21	46294986	1094	cis	No	No
1	rs9980080	HDL_P	-0.019	0.003	2.49E-09	T	C	0.240	21	46899450	38818	cis	No	No
1	rs2256609	HDL_P	-0.038	0.003	1.39E-29	G	A	0.191	22	21925017	19266	cis	Yes	No
1	rs3747207	HDL_P	-0.019	0.003	5.41E-09	A	G	0.215	22	44324855	4350	cis	Yes	No
1	rs6542680	HDL_P	-0.021	0.003	1.17E-09	T	C	0.820	2	3640142	2279	cis	No	No
1	rs35135293	HDL_P	-0.025	0.003	1.26E-21	T	C	0.519	2	20363666	7847	cis	Yes	No
1	rs676210	HDL_P	0.045	0.003	7.68E-43	A	G	0.203	2	21231524	7734	cis	Yes	No
1	rs1260326	HDL_P	-0.05	0.003	4.76E-76	C	T	0.608	2	27730940	11233	cis	Yes	No
1	rs62135193	HDL_P	-0.015	0.003	2.53E-08	T	C	0.540	2	48755477	1044	cis	Yes	Yes
1	rs2302643	HDL_P	0.015	0.003	3.48E-08	A	G	0.446	2	64881229	9571	cis	Yes	Yes
1	rs7572242	HDL_P	-0.016	0.003	8.96E-09	T	C	0.403	2	128625164	9434	cis	Yes	No
1	rs72926946	HDL_P	-0.016	0.003	2.09E-08	A	C	0.296	2	203477868	1068	cis	Yes	No
1	rs715	HDL_P	-0.019	0.003	1.07E-10	C	T	0.312	2	211543055	56833	cis	Yes	No
1	rs78058190	HDL_P	-0.044	0.007	9.78E-11	A	G	0.050	2	219699999	3189	cis	Yes	No
1	rs2203452	HDL_P	-0.025	0.003	1.84E-19	G	A	0.648	2	227094758	13319	cis	Yes	No
1	rs74780677	HDL_P	-0.076	0.011	4.72E-12	G	A	0.015	3	48601774	559	cis	No	No
1	rs11921360	HDL_P	0.016	0.003	3.33E-09	A	C	0.633	3	119612055	23687	cis	Yes	No
1	rs645040	HDL_P	-0.027	0.003	9.71E-18	T	G	0.774	3	135926622	1864	cis	Yes	No
1	rs2871960	HDL_P	-0.016	0.003	2.26E-09	C	A	0.446	3	141121814	10200	cis	Yes	Yes
1	rs16864251	HDL_P	0.018	0.003	5.52E-09	A	G	0.258	3	152116954	50106	cis	Yes	No
1	rs13322435	HDL_P	0.019	0.003	9.10E-12	G	A	0.402	3	156795468	3785	cis	Yes	No

ID	SNP	Metabolite	Beta	SE	p value	A1	A2	EAF	Chr	Position	Distance*	Classification	IV	Novel variant
1	rs10513801	HDL_P	-0.023	0.004	4.98E-09	G	T	0.136	3	185822353	5753	cis	Yes	No
1	rs13108218	HDL_P	-0.016	0.003	6.32E-09	G	A	0.619	4	3443931	235	cis	Yes	No
1	rs10019888	HDL_P	-0.024	0.004	4.58E-11	G	A	0.164	4	26062990	2152	cis	Yes	No
1	rs35088884	HDL_P	0.023	0.003	1.04E-15	C	A	0.658	4	69357050	26314	cis	No	No
1	rs13101719	HDL_P	0.019	0.003	1.06E-09	A	T	0.418	4	69446431	4299	cis	No	No
1	rs4693557	HDL_P	0.016	0.003	1.72E-08	G	A	0.582	4	83998642	1531	cis	Yes	No
1	rs13107325	HDL_P	-0.06	0.005	2.13E-32	T	C	0.074	4	103188709	50606	cis	Yes	No
1	rs7670387	HDL_P	0.015	0.003	2.10E-08	C	G	0.420	4	123522537	2992	cis	No	Yes
1	rs888655	HDL_P	0.023	0.003	3.59E-15	G	A	0.719	5	72917439	4230	cis	Yes	No
1	rs6595799	HDL_P	-0.018	0.003	4.11E-09	T	C	0.733	5	127408683	10774	cis	No	No
1	rs6871098	HDL_P	-0.015	0.003	3.55E-08	G	C	0.633	5	153373345	8093	cis	No	No
1	rs9258375	HDL_P	-0.035	0.004	3.63E-18	G	A	0.129	6	29752808	5922	cis	Yes	No
1	rs142804168	HDL_P	-0.049	0.005	1.45E-21	A	G	0.117	6	32487361	10702	cis	No	No
1	rs112692477	HDL_P	0.032	0.003	1.17E-29	A	G	0.580	6	32506042	7777	cis	No	No
1	rs78121911	HDL_P	-0.028	0.004	2.16E-10	C	T	0.097	6	34503739	20370	cis	Yes	Yes
1	rs9471975	HDL_P	0.021	0.003	3.46E-15	C	T	0.583	6	42919222	5280	cis	Yes	No
1	rs571848809	HDL_P	-0.025	0.004	5.71E-09	A	G	0.108	6	161005388	64049	cis	No	No
1	rs836558	HDL_P	0.017	0.003	1.02E-10	A	C	0.458	7	6447991	33832	cis	Yes	No
1	rs10242866	HDL_P	-0.023	0.003	2.45E-17	T	C	0.396	7	17920613	19292	cis	Yes	No
1	rs12533197	HDL_P	-0.016	0.003	8.43E-10	G	T	0.438	7	36192606	256	cis	Yes	No
1	rs11773763	HDL_P	0.018	0.003	2.06E-10	T	C	0.327	7	50271499	25363	cis	No	No
1	rs7783857	HDL_P	-0.017	0.003	6.67E-09	G	C	0.281	7	130439058	20090	cis	Yes	No
1	rs6467595	HDL_P	0.017	0.003	4.08E-08	T	C	0.772	7	135247752	5086	cis	Yes	Yes
1	rs4240624	HDL_P	0.108	0.005	4.99E-123	A	G	0.908	8	9184231	1143	cis	Yes	No
1	rs59347135	HDL_P	-0.064	0.007	5.19E-23	G	C	0.046	8	19750044	9183	cis	Yes	No
1	rs139915535	HDL_P	-0.14	0.01	8.98E-45	G	A	0.018	8	19766233	7004	cis	Yes	No
1	rs15285	HDL_P	0.084	0.003	1.08E-179	T	C	0.284	8	19824667	65438	cis	Yes	No
1	rs10504477	HDL_P	-0.017	0.003	9.81E-10	C	T	0.408	8	71338185	1373	cis	Yes	No
1	rs2737214	HDL_P	0.028	0.003	3.79E-25	G	C	0.558	8	116624987	103157	cis	No	No
1	rs10109303	HDL_P	-0.017	0.003	3.90E-09	T	A	0.709	8	121873486	34415	cis	Yes	No
1	rs35309034	HDL_P	0.017	0.003	3.67E-08	A	C	0.239	8	126498265	19083	cis	Yes	Yes
1	rs7817574	HDL_P	0.027	0.003	7.67E-15	C	T	0.186	8	144302570	52	cis	Yes	No
1	rs686030	HDL_P	0.038	0.004	7.39E-24	A	C	0.859	9	15304782	2575	cis	Yes	No
1	rs2297402	HDL_P	-0.062	0.009	7.29E-12	T	C	0.024	9	107579880	53469	cis	Yes	No
1	rs2066714	HDL_P	0.033	0.004	1.53E-16	C	T	0.127	9	107586753	60342	cis	Yes	No
1	rs11789603	HDL_P	0.053	0.004	2.20E-35	T	C	0.109	9	107647019	22163	cis	Yes	No
1	rs2740488	HDL_P	-0.051	0.003	4.87E-63	C	A	0.264	9	107661742	7440	cis	Yes	No
1	rs635634	HDL_P	0.032	0.003	1.22E-20	T	C	0.185	9	136155000	3559	cis	Yes	No
2	rs7924036	VLDL_size	-0.036	0.003	3.39E-43	T	G	0.502	10	65191645	1296	cis	Yes	No
2	rs8211	VLDL_size	0.025	0.003	5.10E-21	T	C	0.616	10	94819053	1501	cis	Yes	Yes
2	rs10786114	VLDL_size	-0.031	0.004	2.43E-15	T	C	0.875	10	95309022	17399	cis	Yes	Yes
2	rs117779480	VLDL_size	-0.029	0.005	1.67E-08	T	C	0.069	10	104080583	59595	cis	Yes	Yes
2	rs2792703	VLDL_size	0.021	0.003	1.05E-13	T	C	0.721	10	113931690	19872	cis	Yes	Yes
2	rs2484294	VLDL_size	-0.018	0.003	1.52E-09	A	G	0.736	10	115792062	5651	cis	Yes	Yes
2	rs11199840	VLDL_size	-0.017	0.003	3.24E-11	G	A	0.550	10	122990338	35685	cis	Yes	Yes
2	rs2098839	VLDL_size	0.016	0.003	9.38E-09	C	T	0.670	11	10676987	38547	cis	Yes	Yes
2	rs6486122	VLDL_size	0.016	0.003	1.49E-08	T	C	0.693	11	13361524	13230	cis	Yes	Yes
2	rs11030107	VLDL_size	0.019	0.003	3.89E-10	G	A	0.265	11	27694835	38509	cis	Yes	Yes
2	rs10838681	VLDL_size	-0.023	0.003	1.25E-14	A	G	0.269	11	47275064	4606	cis	Yes	Yes
2	rs174564	VLDL_size	0.069	0.003	1.77E-139	G	A	0.351	11	61588305	5592	cis	Yes	No

ID	SNP	Metabolite	Beta	SE	p value	A1	A2	EAF	Chr	Position	Distance*	Classification	IV	Novel variant
2	rs12294913	VLDL_size	0.063	0.006	2.13E-23	G	C	0.048	11	64004643	2586	cis	No	Yes
2	rs10750766	VLDL_size	0.02	0.003	2.52E-12	A	C	0.712	11	65473798	5668	cis	Yes	No
2	rs11824135	VLDL_size	0.161	0.005	1.00E-200	T	G	0.066	11	116559553	42388	cis	Yes	Yes
2	rs12421652	VLDL_size	-0.045	0.004	1.62E-31	T	G	0.130	11	116625875	17827	cis	Yes	Yes
2	rs2070665	VLDL_size	-0.147	0.005	1.20E-189	G	A	0.927	11	116707684	653	cis	Yes	Yes
2	rs117604511	VLDL_size	0.051	0.009	1.19E-08	C	A	0.023	11	117382799	99932	cis	Yes	Yes
2	rs76895963	VLDL_size	-0.092	0.01	9.62E-20	G	T	0.020	12	4384844	76	cis	Yes	Yes
2	rs11045172	VLDL_size	-0.027	0.003	7.17E-17	C	A	0.198	12	20470221	3521	cis	Yes	Yes
2	rs11045866	VLDL_size	0.021	0.004	7.26E-09	C	A	0.151	12	21367382	83253	cis	No	Yes
2	rs6582586	VLDL_size	0.015	0.003	8.72E-09	A	G	0.400	12	46416742	18273	cis	Yes	Yes
2	rs3809114	VLDL_size	-0.021	0.003	1.27E-15	A	G	0.564	12	57848639	2532	cis	Yes	Yes
2	rs73125460	VLDL_size	-0.023	0.004	1.57E-08	T	C	0.117	12	58094497	6872	cis	Yes	Yes
2	rs11174498	VLDL_size	-0.021	0.003	9.10E-10	C	T	0.169	12	62865345	4747	cis	Yes	Yes
2	rs2137537	VLDL_size	0.015	0.003	7.97E-09	C	T	0.544	12	71113087	444	cis	Yes	Yes
2	rs10861661	VLDL_size	0.019	0.003	1.19E-10	C	A	0.247	12	107174646	5911	cis	Yes	Yes
2	rs2393792	VLDL_size	0.016	0.003	1.90E-10	A	G	0.481	12	121406293	10052	cis	Yes	Yes
2	rs10773000	VLDL_size	-0.016	0.003	5.64E-09	T	G	0.333	12	123736084	7522	cis	Yes	Yes
2	rs4930724	VLDL_size	-0.025	0.003	9.99E-20	C	T	0.330	12	124423817	3063	cis	Yes	Yes
2	rs10773049	VLDL_size	-0.03	0.003	9.05E-30	C	T	0.395	12	124506631	15858	cis	Yes	No
2	rs1928496	VLDL_size	0.017	0.003	3.63E-09	T	C	0.744	13	31012904	6438	cis	Yes	No
2	rs2532972	VLDL_size	0.015	0.003	7.88E-09	C	T	0.405	13	50980822	12938	cis	No	Yes
2	rs9561643	VLDL_size	0.017	0.003	4.61E-10	C	A	0.313	13	95253131	979	cis	Yes	No
2	rs7140110	VLDL_size	0.022	0.003	2.60E-14	C	T	0.297	13	114544024	23024	cis	Yes	Yes
2	rs60852772	VLDL_size	-0.018	0.003	2.99E-09	C	T	0.252	13	114824211	343	cis	Yes	Yes
2	rs11634582	VLDL_size	-0.017	0.003	1.40E-08	T	C	0.272	15	41896257	17074	cis	Yes	Yes
2	rs139974673	VLDL_size	0.112	0.008	2.15E-40	C	T	0.025	15	44027885	1401	cis	Yes	No
2	rs1318175	VLDL_size	0.081	0.004	2.63E-113	T	C	0.159	15	58586129	9559	cis	Yes	Yes
2	rs4775033	VLDL_size	0.047	0.004	1.04E-30	T	A	0.119	15	58644087	58680	cis	Yes	Yes
2	rs1601933	VLDL_size	0.084	0.003	1.00E-200	T	C	0.468	15	58671559	31208	cis	Yes	Yes
2	rs11632618	VLDL_size	-0.095	0.005	4.17E-78	A	G	0.070	15	58724706	21819	cis	Yes	No
2	rs77516617	VLDL_size	0.042	0.006	9.74E-13	A	G	0.054	15	59058720	4670	cis	Yes	No
2	rs4781137	VLDL_size	0.015	0.003	6.05E-09	C	T	0.489	16	11774950	12638	cis	Yes	Yes
2	rs11075253	VLDL_size	-0.032	0.003	3.69E-29	A	C	0.296	16	15148646	1289	cis	Yes	No
2	rs749671	VLDL_size	-0.015	0.003	1.12E-08	A	G	0.370	16	31088347	2603	cis	Yes	Yes
2	rs247616	VLDL_size	-0.028	0.003	2.02E-24	T	C	0.325	16	56989590	1586	cis	Yes	No
2	rs2937124	VLDL_size	-0.018	0.003	7.34E-11	T	C	0.363	16	69862109	42092	cis	Yes	Yes
2	rs34682685	VLDL_size	0.035	0.004	5.61E-17	A	G	0.105	16	72096227	819	cis	Yes	No
2	rs2925979	VLDL_size	-0.028	0.003	3.97E-23	C	T	0.699	16	81534790	16638	cis	Yes	No
2	rs11078597	VLDL_size	0.019	0.003	1.32E-08	C	T	0.186	17	1618363	1453	cis	Yes	Yes
2	rs9895028	VLDL_size	0.038	0.005	1.41E-12	G	A	0.062	17	17489606	5135	cis	No	No
2	rs72836561	VLDL_size	0.148	0.007	2.53E-89	T	C	0.032	17	41926126	1609	cis	Yes	No
2	rs117467721	VLDL_size	0.044	0.007	1.25E-09	T	C	0.035	17	42099990	1323	cis	Yes	Yes
2	rs11079868	VLDL_size	-0.019	0.003	4.62E-13	G	A	0.549	17	47357969	45293	cis	Yes	Yes
2	rs1292063	VLDL_size	-0.02	0.003	2.28E-10	A	T	0.786	17	57907768	10858	cis	No	Yes
2	rs1801689	VLDL_size	-0.134	0.008	5.15E-69	C	A	0.030	17	64210580	16768	cis	No	No
2	rs9892166	VLDL_size	0.016	0.003	2.23E-08	G	C	0.735	17	73373135	3317	cis	Yes	Yes
2	rs164011	VLDL_size	-0.017	0.003	4.16E-10	A	G	0.318	17	74273165	5000	cis	Yes	Yes
2	rs4969145	VLDL_size	-0.02	0.003	5.66E-14	C	T	0.581	17	76406170	10153	cis	Yes	No
2	rs1808579	VLDL_size	-0.015	0.003	2.06E-09	T	C	0.487	18	21104888	21427	cis	No	Yes
2	rs10405423	VLDL_size	0.02	0.003	2.78E-13	A	C	0.659	19	7211311	14019	cis	Yes	Yes

ID	SNP	Metabolite	Beta	SE	p value	A1	A2	EAF	Chr	Position	Distance*	Classification	IV	Novel variant
2	rs4804833	VLDL_size	-0.023	0.003	1.24E-17	G	A	0.619	19	7970635	1885	cis	Yes	Yes
2	rs116843064	VLDL_size	-0.209	0.009	8.11E-110	A	G	0.019	19	8429323	1149	cis	Yes	No
2	rs1044821	VLDL_size	0.016	0.003	2.10E-08	T	C	0.284	19	18307108	3079	cis	Yes	Yes
2	rs58542926	VLDL_size	-0.098	0.005	1.36E-88	T	C	0.076	19	19379549	4628	cis	Yes	No
2	rs182611493	VLDL_size	-0.124	0.012	3.39E-24	G	A	0.013	19	19458388	26749	cis	Yes	No
2	rs62102718	VLDL_size	0.021	0.003	7.89E-13	T	A	0.284	19	33891013	26421	cis	Yes	Yes
2	rs483082	VLDL_size	0.057	0.003	9.53E-79	T	G	0.239	19	45416178	1325	cis	Yes	No
2	rs5112	VLDL_size	0.066	0.003	5.36E-126	G	C	0.533	19	45430280	46	cis	No	No
2	rs11669173	VLDL_size	-0.018	0.003	2.71E-12	A	G	0.506	19	45492559	12159	cis	Yes	Yes
2	rs35903319	VLDL_size	0.02	0.004	2.46E-08	T	C	0.178	19	56119837	2823	cis	Yes	Yes
2	rs114165349	VLDL_size	0.084	0.009	2.59E-22	C	G	0.023	1	27021913	520	cis	Yes	No
2	rs66848709	VLDL_size	0.023	0.003	1.99E-13	C	T	0.217	1	39612557	7410	cis	Yes	Yes
2	rs12088739	VLDL_size	-0.026	0.005	6.46E-09	G	A	0.090	1	51506886	18622	cis	Yes	Yes
2	rs6699195	VLDL_size	-0.014	0.003	2.65E-08	C	T	0.501	1	54841838	28679	cis	Yes	Yes
2	rs1748195	VLDL_size	-0.062	0.003	6.30E-116	G	C	0.356	1	63049593	13597	cis	Yes	No
2	rs2802525	VLDL_size	-0.019	0.003	2.41E-08	G	A	0.813	1	98549810	2937	cis	Yes	Yes
2	rs949464	VLDL_size	-0.016	0.003	1.09E-09	T	C	0.505	1	156272271	3172	cis	Yes	Yes
2	rs6694406	VLDL_size	-0.017	0.003	3.47E-10	A	G	0.347	1	178532010	1961	cis	Yes	Yes
2	rs2821226	VLDL_size	-0.015	0.003	1.28E-08	G	A	0.526	1	203517292	8088	cis	Yes	Yes
2	rs11118310	VLDL_size	0.019	0.003	1.45E-13	T	A	0.591	1	219637671	45046	cis	Yes	No
2	rs61830291	VLDL_size	0.032	0.004	5.90E-13	C	A	0.098	1	221001142	1410	cis	Yes	Yes
2	rs10779835	VLDL_size	-0.045	0.003	6.39E-64	C	T	0.614	1	230299949	44472	cis	Yes	No
2	rs6065906	VLDL_size	0.066	0.003	6.79E-88	C	T	0.186	20	44554015	9306	cis	Yes	No
2	rs6066149	VLDL_size	-0.016	0.003	4.09E-08	A	G	0.283	20	45602638	23596	cis	Yes	Yes
2	rs133023	VLDL_size	-0.018	0.003	3.68E-12	C	T	0.472	22	38573629	23144	cis	Yes	Yes
2	rs3747207	VLDL_size	-0.027	0.003	8.25E-18	A	G	0.215	22	44324855	4350	cis	Yes	Yes
2	rs676210	VLDL_size	-0.086	0.003	1.25E-157	A	G	0.203	2	21231524	7734	cis	Yes	No
2	rs483490	VLDL_size	0.018	0.003	1.92E-08	G	A	0.212	2	25756264	26359	cis	Yes	Yes
2	rs4665972	VLDL_size	-0.084	0.003	1.00E-200	C	T	0.608	2	27598097	4733	cis	Yes	No
2	rs10210970	VLDL_size	0.026	0.004	1.13E-11	T	C	0.130	2	28646847	1964	cis	Yes	Yes
2	rs1009358	VLDL_size	-0.018	0.003	1.94E-11	C	T	0.373	2	65276452	6914	cis	Yes	Yes
2	rs6734238	VLDL_size	-0.016	0.003	9.40E-10	G	A	0.403	2	113841030	3351	cis	No	Yes
2	rs35932591	VLDL_size	0.022	0.004	3.07E-08	T	C	0.120	2	121310704	146	cis	Yes	Yes
2	rs13389219	VLDL_size	-0.039	0.003	2.94E-49	T	C	0.394	2	165528876	15229	cis	Yes	No
2	rs9751327	VLDL_size	-0.015	0.003	3.40E-08	A	G	0.596	2	219179995	6973	cis	Yes	Yes
2	rs78058190	VLDL_size	0.073	0.007	4.31E-28	A	G	0.050	2	219699999	3189	cis	Yes	No
2	rs2943648	VLDL_size	0.041	0.003	2.18E-50	G	A	0.647	2	227100490	19051	cis	Yes	No
2	rs2881654	VLDL_size	-0.028	0.004	7.26E-13	A	G	0.122	3	12396955	27452	cis	Yes	No
2	rs9968117	VLDL_size	-0.034	0.004	1.37E-17	T	C	0.121	3	52838654	9869	cis	Yes	No
2	rs9844972	VLDL_size	0.041	0.005	1.27E-15	C	G	0.071	3	150097635	20232	cis	Yes	Yes
2	rs79287178	VLDL_size	0.065	0.008	1.09E-16	A	G	0.031	3	172294500	18896	cis	Yes	Yes
2	rs6599391	VLDL_size	0.022	0.004	6.55E-09	G	T	0.137	4	979179	1503	cis	No	Yes
2	rs13108218	VLDL_size	-0.033	0.003	1.93E-35	G	A	0.619	4	3443931	235	cis	Yes	No
2	rs1902176	VLDL_size	0.015	0.003	3.45E-08	G	T	0.638	4	15099106	3387	cis	No	No
2	rs71603401	VLDL_size	0.022	0.004	9.99E-09	G	A	0.135	4	18034463	10963	cis	Yes	Yes
2	rs10019888	VLDL_size	0.024	0.004	3.57E-12	G	A	0.164	4	26062990	2152	cis	Yes	Yes
2	rs1471251	VLDL_size	0.028	0.003	7.91E-26	T	A	0.398	4	87976359	94404	cis	Yes	No
2	rs1377290	VLDL_size	0.014	0.003	3.58E-08	T	C	0.460	4	89719368	89594	cis	Yes	Yes
2	rs2035816	VLDL_size	-0.028	0.005	3.91E-09	G	A	0.083	4	100508556	23317	cis	Yes	Yes
2	rs4691378	VLDL_size	-0.019	0.003	1.36E-11	G	A	0.316	4	157704459	58051	cis	Yes	Yes

ID	SNP	Metabolite	Beta	SE	p value	A1	A2	EAF	Chr	Position	Distance*	Classification	IV	Novel variant
2	rs7735249	VLDL_size	0.023	0.004	3.04E-08	G	C	0.113	5	53310139	33998	cis	Yes	Yes
2	rs593979	VLDL_size	-0.016	0.003	3.15E-09	C	T	0.401	5	55776404	1543	cis	Yes	Yes
2	rs3936511	VLDL_size	0.045	0.003	1.19E-43	G	A	0.191	5	55860781	27127	cis	Yes	No
2	rs4976033	VLDL_size	0.02	0.003	5.00E-14	G	A	0.401	5	67714246	4454	cis	Yes	Yes
2	rs1818351	VLDL_size	0.016	0.003	2.54E-08	C	T	0.706	5	90252077	26596	cis	No	Yes
2	rs258051	VLDL_size	-0.018	0.003	1.52E-10	A	C	0.705	5	118767591	13819	cis	Yes	Yes
2	rs72801474	VLDL_size	-0.028	0.004	5.84E-10	A	G	0.093	5	132444128	16053	cis	Yes	Yes
2	rs6882345	VLDL_size	0.027	0.003	4.04E-23	A	G	0.635	5	156397673	7435	cis	Yes	No
2	rs75049939	VLDL_size	0.016	0.003	6.91E-09	G	T	0.305	5	173334219	4465	cis	Yes	Yes
2	rs7756992	VLDL_size	0.019	0.003	2.36E-10	G	A	0.265	6	20679709	44176	cis	Yes	Yes
2	rs2523740	VLDL_size	-0.015	0.003	1.11E-08	A	G	0.390	6	30224362	2962	cis	Yes	Yes
2	rs28752924	VLDL_size	0.022	0.003	2.40E-15	C	T	0.441	6	31303922	17726	cis	No	Yes
2	rs557699552	VLDL_size	0.069	0.009	4.98E-16	A	T	0.025	6	31765865	2335	cis	No	Yes
2	rs28383314	VLDL_size	0.032	0.003	2.96E-33	C	T	0.625	6	32587213	3686	cis	Yes	No
2	rs9273338	VLDL_size	-0.016	0.003	1.75E-08	T	A	0.435	6	32623393	4263	cis	No	Yes
2	rs998584	VLDL_size	0.035	0.003	8.24E-42	A	C	0.481	6	43757896	6984	cis	Yes	No
2	rs9353320	VLDL_size	-0.015	0.003	4.62E-08	G	C	0.602	6	86244848	36071	cis	Yes	Yes
2	rs4946812	VLDL_size	-0.017	0.003	1.38E-09	A	G	0.324	6	107431688	5030	cis	Yes	Yes
2	rs3951069	VLDL_size	-0.015	0.003	7.88E-09	G	A	0.432	6	109131312	7570	cis	Yes	Yes
2	rs2800709	VLDL_size	0.029	0.003	2.16E-29	G	T	0.521	6	127439297	518	cis	Yes	No
2	rs605066	VLDL_size	-0.029	0.003	2.39E-28	T	C	0.591	6	139829666	26151	cis	Yes	No
2	rs3963973	VLDL_size	-0.069	0.01	1.55E-11	C	G	0.019	6	160586274	1486	cis	Yes	Yes
2	rs9295127	VLDL_size	-0.289	0.009	1.00E-200	A	C	0.020	6	160751123	18286	cis	Yes	No
2	rs146184004	VLDL_size	-0.107	0.008	2.06E-38	T	C	0.029	6	160766334	3075	cis	Yes	Yes
2	rs6938647	VLDL_size	-0.078	0.003	6.97E-132	C	A	0.782	6	160986915	45576	cis	Yes	Yes
2	rs143725668	VLDL_size	0.021	0.004	5.99E-09	G	A	0.202	6	161684237	10855	cis	No	Yes
2	rs62433138	VLDL_size	0.028	0.003	1.74E-26	C	T	0.431	7	1050877	11784	cis	No	No
2	rs2254737	VLDL_size	0.017	0.003	2.90E-10	T	C	0.418	7	17811797	92623	cis	Yes	Yes
2	rs4722551	VLDL_size	-0.038	0.004	1.16E-26	C	T	0.158	7	25991826	692	cis	Yes	Yes
2	rs1406754	VLDL_size	-0.018	0.003	7.92E-12	T	G	0.401	7	26396198	37509	cis	Yes	Yes
2	rs10264306	VLDL_size	0.027	0.003	1.66E-16	G	A	0.197	7	28239466	19051	cis	Yes	No
2	rs217379	VLDL_size	-0.015	0.003	4.65E-09	T	C	0.444	7	44607172	6450	cis	Yes	Yes
2	rs55682112	VLDL_size	0.015	0.003	2.89E-08	G	T	0.343	7	66056346	1088	cis	Yes	Yes
2	rs714052	VLDL_size	-0.119	0.004	1.00E-200	G	A	0.125	7	72864869	16760	cis	Yes	No
2	rs72555385	VLDL_size	0.068	0.006	3.90E-29	G	A	0.049	7	73123473	2173	cis	Yes	No
2	rs13228906	VLDL_size	-0.03	0.003	1.22E-31	A	G	0.493	7	130441884	22916	cis	Yes	Yes
2	rs2980755	VLDL_size	-0.037	0.003	9.74E-44	G	A	0.461	8	8363683	44682	cis	Yes	No
2	rs4922079	VLDL_size	-0.026	0.003	3.08E-16	C	T	0.776	8	19518499	17550	cis	Yes	Yes
2	rs144503444	VLDL_size	0.156	0.01	3.24E-58	C	T	0.019	8	19773155	13926	cis	No	No
2	rs286	VLDL_size	-0.173	0.005	1.00E-200	T	A	0.086	8	19815256	56027	cis	Yes	No
2	rs308	VLDL_size	-0.151	0.009	8.87E-61	G	T	0.021	8	19817476	58247	cis	Yes	Yes
2	rs13253288	VLDL_size	-0.018	0.003	4.61E-11	C	G	0.663	8	59338856	387	cis	Yes	Yes
2	rs28446899	VLDL_size	0.03	0.005	1.13E-09	T	C	0.077	8	72396213	34282	cis	Yes	Yes
2	rs2245019	VLDL_size	-0.018	0.003	2.64E-12	A	C	0.560	8	116622906	101076	cis	Yes	Yes
2	rs28601761	VLDL_size	-0.065	0.003	1.74E-129	G	C	0.421	8	126500031	20849	cis	No	No
2	rs3860846	VLDL_size	0.026	0.003	2.40E-18	T	C	0.275	8	126631687	28787	cis	No	Yes
2	rs7870869	VLDL_size	0.016	0.003	2.37E-08	G	A	0.308	9	1036193	6414	cis	No	Yes
2	rs7047907	VLDL_size	0.02	0.003	4.85E-11	A	G	0.745	9	86368660	45786	cis	Yes	Yes
2	rs62565259	VLDL_size	-0.02	0.003	1.25E-08	T	C	0.174	9	102162570	41389	cis	Yes	Yes
2	rs2808407	VLDL_size	0.017	0.003	3.69E-09	T	G	0.268	9	126295298	41134	cis	Yes	Yes

ID	SNP	Metabolite	Beta	SE	p value	A1	A2	EAF	Chr	Position	Distance*	Classification	IV	Novel variant
2	rs2519093	VLDL_size	-0.022	0.003	5.43E-11	T	C	0.185	9	136141870	9569	cis	Yes	Yes
2	rs11999525	VLDL_size	0.021	0.003	3.71E-13	A	G	0.261	9	139371342	6604	cis	Yes	Yes
3	rs74827952	Omega_3	-0.034	0.004	1.08E-20	A	G	0.157	10	5245196	3935	cis	Yes	No
3	rs792456	Omega_3	0.019	0.003	3.46E-11	T	C	0.702	10	22174541	19690	cis	Yes	No
3	rs7075901	Omega_3	0.02	0.003	1.35E-13	A	C	0.471	10	65280994	166	cis	Yes	No
3	rs1782652	Omega_3	0.018	0.003	6.84E-11	A	T	0.379	10	81074125	33099	cis	Yes	No
3	rs2068888	Omega_3	-0.028	0.003	2.59E-26	A	G	0.447	10	94839642	6409	cis	Yes	No
3	rs11596236	Omega_3	0.015	0.003	2.95E-08	A	C	0.353	10	96672375	26039	cis	Yes	No
3	rs7901791	Omega_3	-0.018	0.003	4.27E-09	G	C	0.744	10	102645384	16913	cis	Yes	No
3	rs74563318	Omega_3	-0.047	0.007	3.57E-11	A	C	0.036	10	114124977	1897	cis	Yes	No
3	rs11024743	Omega_3	0.018	0.003	9.46E-09	G	A	0.741	11	18653712	2421	cis	Yes	Yes
3	rs4572092	Omega_3	-0.029	0.005	1.03E-09	A	G	0.093	11	47094546	31049	cis	Yes	No
3	rs141443145	Omega_3	0.041	0.007	5.00E-10	T	A	0.048	11	50153135	9404	cis	Yes	No
3	rs2905502	Omega_3	-0.044	0.005	6.98E-21	C	G	0.095	11	60848957	5455	cis	Yes	No
3	rs4938998	Omega_3	0.029	0.004	6.16E-14	T	C	0.157	11	61017871	9223	cis	Yes	No
3	rs147419504	Omega_3	0.117	0.012	1.16E-22	T	C	0.015	11	61450428	2564	cis	Yes	No
3	rs198457	Omega_3	0.117	0.004	1.00E-200	T	C	0.187	11	61471678	23814	cis	Yes	No
3	rs79519287	Omega_3	-0.159	0.006	4.61E-155	G	A	0.059	11	61531810	6660	cis	Yes	No
3	rs11825591	Omega_3	-0.138	0.008	1.21E-65	T	C	0.030	11	61693684	5942	cis	Yes	Yes
3	rs199789685	Omega_3	0.062	0.003	1.51E-90	T	C	0.627	11	61866297	25015	cis	No	Yes
3	rs673335	Omega_3	-0.069	0.004	6.25E-75	C	T	0.158	11	75450576	2597	cis	Yes	No
3	rs17119701	Omega_3	0.043	0.007	5.39E-09	G	A	0.036	11	116331582	39087	cis	Yes	Yes
3	rs964184	Omega_3	-0.128	0.004	1.00E-200	C	G	0.868	11	116648917	4811	cis	Yes	No
3	rs144018203	Omega_3	0.14	0.014	1.44E-24	C	G	0.011	11	116916060	9292	cis	No	No
3	rs76895963	Omega_3	-0.07	0.01	1.08E-11	G	T	0.020	12	4384844	76	cis	Yes	No
3	rs7397189	Omega_3	-0.026	0.003	4.22E-17	G	T	0.243	12	57728776	38508	cis	No	No
3	rs2393776	Omega_3	-0.02	0.003	3.19E-13	A	G	0.623	12	121424406	5634	cis	Yes	No
3	rs2937330	Omega_3	-0.022	0.003	1.50E-13	G	C	0.724	13	56009212	32175	cis	Yes	No
3	rs7140110	Omega_3	0.02	0.003	3.33E-12	C	T	0.297	13	114544024	23024	cis	Yes	No
3	rs12016920	Omega_3	-0.02	0.003	1.66E-09	C	T	0.195	13	114639025	7060	cis	Yes	No
3	rs7350719	Omega_3	0.016	0.003	4.34E-09	T	C	0.502	14	70665113	8915	cis	No	No
3	rs139974673	Omega_3	0.107	0.009	8.01E-36	C	T	0.025	15	44027885	1401	cis	Yes	No
3	rs10162642	Omega_3	-0.043	0.003	2.20E-39	A	G	0.210	15	58577163	593	cis	Yes	No
3	rs1601935	Omega_3	-0.099	0.003	1.00E-200	T	G	0.654	15	58671765	31002	cis	Yes	No
3	rs12908761	Omega_3	0.034	0.005	7.32E-14	G	A	0.098	15	58715647	12878	cis	Yes	No
3	rs1077835	Omega_3	0.126	0.003	1.00E-200	G	A	0.220	15	58723426	20657	cis	Yes	No
3	rs199782954	Omega_3	0.037	0.003	9.21E-30	A	G	0.282	16	14886131	22198	cis	No	No
3	rs77927814	Omega_3	-0.08	0.003	9.73E-169	C	T	0.297	16	15124327	9981	cis	No	No
3	rs4986970	Omega_3	-0.053	0.007	1.69E-13	T	A	0.035	16	67976320	1713	cis	Yes	No
3	rs907035	Omega_3	-0.02	0.003	5.44E-14	A	G	0.538	16	83988964	12884	cis	Yes	No
3	rs1050541	Omega_3	-0.018	0.003	3.68E-11	G	T	0.558	17	7560835	10889	cis	Yes	Yes
3	rs1242503	Omega_3	-0.027	0.003	1.62E-18	A	C	0.766	17	17368149	4390	cis	Yes	No
3	rs228754	Omega_3	-0.027	0.003	6.14E-15	A	C	0.824	17	42116056	13464	cis	Yes	No
3	rs4597358	Omega_3	-0.031	0.003	1.08E-22	G	C	0.228	17	44110670	2008	cis	Yes	No
3	rs78057960	Omega_3	-0.019	0.003	5.45E-11	T	C	0.292	17	73779075	1612	cis	No	No
3	rs77960347	Omega_3	0.162	0.011	1.25E-45	G	A	0.014	18	47109955	22885	cis	No	No
3	rs9304381	Omega_3	0.047	0.003	1.21E-41	T	C	0.820	18	47158234	18485	cis	Yes	No
3	rs78277979	Omega_3	0.023	0.004	1.34E-09	A	C	0.145	18	47276942	8508	cis	No	Yes
3	rs116843064	Omega_3	-0.078	0.01	3.67E-16	A	G	0.019	19	8429323	1149	cis	Yes	No
3	rs56048141	Omega_3	-0.054	0.01	4.02E-08	T	C	0.019	19	11317744	1126	cis	Yes	No

ID	SNP	Metabolite	Beta	SE	p value	A1	A2	EAF	Chr	Position	Distance*	Classification	IV	Novel variant
3	rs2269869	Omega_3	-0.017	0.003	4.75E-08	A	G	0.759	19	19109545	34838	cis	Yes	No
3	rs72999033	Omega_3	-0.172	0.005	1.00E-200	T	C	0.065	19	19366632	184	cis	Yes	No
3	rs182611493	Omega_3	-0.217	0.013	1.55E-67	G	A	0.013	19	19458388	26749	cis	Yes	No
3	rs157592	Omega_3	0.042	0.003	2.84E-33	C	A	0.187	19	45424514	4683	cis	No	No
3	rs5112	Omega_3	0.053	0.003	2.73E-77	G	C	0.533	19	45430280	46	cis	No	No
3	rs1132899	Omega_3	0.028	0.003	4.47E-26	C	T	0.508	19	45448036	1255	cis	Yes	No
3	rs563028852	Omega_3	0.024	0.003	3.32E-13	C	T	0.272	19	54672549	4394	cis	No	No
3	rs77784287	Omega_3	0.02	0.003	3.79E-11	T	C	0.326	1	2332666	9451	cis	No	No
3	rs1168093	Omega_3	0.07	0.003	1.01E-142	A	G	0.644	1	63116227	8888	cis	No	No
3	rs660240	Omega_3	0.027	0.003	4.18E-17	C	T	0.785	1	109817838	7969	cis	Yes	No
3	rs140584594	Omega_3	-0.023	0.003	2.78E-14	G	A	0.730	1	110232983	2546	cis	No	No
3	rs198325	Omega_3	0.019	0.003	2.93E-09	T	C	0.222	1	150937405	315	cis	Yes	Yes
3	rs652108	Omega_3	-0.02	0.004	1.70E-08	T	C	0.158	1	230312704	57227	cis	Yes	No
3	rs11167275	Omega_3	-0.027	0.004	5.25E-10	C	T	0.100	20	34340275	460	cis	Yes	No
3	rs1883711	Omega_3	0.063	0.008	3.40E-16	C	G	0.032	20	39179822	33413	cis	Yes	No
3	rs6093446	Omega_3	0.017	0.003	2.71E-09	A	G	0.287	20	39780932	14288	cis	Yes	Yes
3	rs1800961	Omega_3	-0.052	0.008	7.65E-12	T	C	0.031	20	43042364	5603	cis	Yes	Yes
3	rs13049280	Omega_3	-0.024	0.003	5.00E-18	G	C	0.369	21	40594656	2924	cis	Yes	No
3	rs139052	Omega_3	0.018	0.003	3.07E-09	A	C	0.749	22	44327012	6507	cis	Yes	No
3	rs11895352	Omega_3	-0.023	0.003	1.19E-18	T	C	0.477	2	20367135	7971	cis	Yes	No
3	rs62120845	Omega_3	-0.02	0.003	2.37E-11	G	C	0.730	2	21204430	10348	cis	No	Yes
3	rs4635554	Omega_3	0.036	0.003	1.42E-37	G	T	0.340	2	21389659	32108	cis	Yes	No
3	rs1260326	Omega_3	-0.08	0.003	1.25E-188	C	T	0.608	2	27730940	11233	cis	Yes	No
3	rs2441383	Omega_3	-0.018	0.003	1.78E-11	C	T	0.421	2	61811836	2425	cis	No	No
3	rs1446585	Omega_3	0.02	0.003	3.82E-10	G	A	0.226	2	136407479	6669	cis	Yes	No
3	rs10195252	Omega_3	-0.019	0.003	2.26E-12	C	T	0.406	2	165513091	31014	cis	Yes	No
3	rs695873	Omega_3	0.074	0.012	2.11E-09	A	C	0.012	2	219266810	558	cis	Yes	Yes
3	rs13098031	Omega_3	-0.02	0.003	6.81E-11	T	G	0.267	3	12239931	32788	cis	Yes	No
3	rs352143	Omega_3	-0.019	0.003	2.05E-09	C	T	0.215	3	52264907	298	cis	Yes	No
3	rs900048	Omega_3	0.02	0.003	2.05E-11	T	C	0.710	3	136218930	69904	cis	Yes	No
3	rs5394	Omega_3	0.025	0.004	3.12E-08	A	G	0.096	3	170744893	360	cis	Yes	No
3	rs79287178	Omega_3	0.055	0.008	4.81E-12	A	G	0.031	3	172294500	18896	cis	Yes	No
3	rs13108218	Omega_3	-0.021	0.003	3.37E-14	G	A	0.619	4	3443931	235	cis	Yes	No
3	rs74486752	Omega_3	0.023	0.004	9.55E-10	G	C	0.144	4	46438404	38842	cis	No	No
3	rs34707604	Omega_3	0.045	0.003	4.01E-45	C	T	0.261	4	69491456	3160	cis	Yes	No
3	rs1349852	Omega_3	0.017	0.003	1.27E-10	C	A	0.475	4	69533217	3152	cis	No	No
3	rs6854749	Omega_3	0.02	0.003	6.72E-10	T	A	0.201	4	87968798	86843	cis	Yes	No
3	rs7735249	Omega_3	0.029	0.004	7.92E-12	G	C	0.113	5	53310139	33998	cis	Yes	No
3	rs28650790	Omega_3	0.021	0.003	2.32E-10	T	C	0.191	5	55861464	27810	cis	Yes	No
3	rs11242109	Omega_3	0.016	0.003	9.28E-10	T	G	0.478	5	131677047	24243	cis	Yes	No
3	rs6882345	Omega_3	0.03	0.003	3.70E-28	A	G	0.635	5	156397673	7435	cis	Yes	No
3	rs2734961	Omega_3	-0.024	0.004	1.98E-10	T	C	0.178	6	29849380	4011	cis	No	Yes
3	rs4947311	Omega_3	0.021	0.003	3.82E-15	T	C	0.487	6	31326166	1741	cis	No	Yes
3	rs13191810	Omega_3	0.034	0.003	4.76E-33	T	C	0.313	6	32586831	4068	cis	Yes	No
3	rs36018387	Omega_3	-0.031	0.004	5.86E-13	T	C	0.106	6	35386872	17577	cis	Yes	No
3	rs3957357	Omega_3	-0.024	0.003	2.59E-19	G	A	0.565	6	52668687	28	cis	No	No
3	rs3756772	Omega_3	0.016	0.003	1.84E-09	T	C	0.401	6	116325142	29911	cis	Yes	Yes
3	rs71562509	Omega_3	-0.016	0.003	1.44E-09	T	G	0.593	6	139835423	31908	cis	No	No
3	rs41267807	Omega_3	0.063	0.01	2.13E-10	C	T	0.018	6	160952816	11477	cis	Yes	No
3	rs186696265	Omega_3	-0.132	0.011	7.06E-33	T	C	0.015	6	161111700	7483	cis	Yes	No

ID	SNP	Metabolite	Beta	SE	p value	A1	A2	EAF	Chr	Position	Distance*	Classification	IV	Novel variant
3	rs1835346	Omega_3	-0.051	0.009	2.72E-09	G	A	0.025	6	161162290	16259	cis	Yes	No
3	rs576405094	Omega_3	0.029	0.004	5.45E-11	T	C	0.110	7	1099162	264	cis	No	No
3	rs4722551	Omega_3	-0.028	0.004	3.19E-14	C	T	0.158	7	25991826	692	cis	Yes	No
3	rs113585586	Omega_3	-0.024	0.004	1.71E-09	T	G	0.129	7	44798315	10134	cis	Yes	No
3	rs147440796	Omega_3	0.047	0.008	2.44E-08	C	T	0.033	7	72823767	5547	cis	No	Yes
3	rs55747707	Omega_3	-0.065	0.003	1.34E-87	A	G	0.204	7	73037366	1506	cis	Yes	No
3	rs41301394	Omega_3	0.02	0.003	7.93E-12	T	C	0.283	7	75612803	11170	cis	Yes	No
3	rs705379	Omega_3	-0.018	0.003	3.04E-11	A	G	0.476	7	94953895	50	cis	Yes	No
3	rs66497154	Omega_3	0.016	0.003	3.71E-08	C	T	0.275	7	151402902	18018	cis	Yes	Yes
3	rs1124175	Omega_3	0.023	0.004	6.66E-10	T	G	0.146	7	155015124	510	cis	Yes	No
3	rs4841132	Omega_3	0.037	0.005	3.53E-16	G	A	0.908	8	9183596	508	cis	No	No
3	rs117199990	Omega_3	-0.054	0.004	1.20E-34	T	C	0.101	8	19820916	61687	cis	Yes	No
3	rs55704109	Omega_3	-0.016	0.003	2.95E-09	A	G	0.493	8	59323247	670	cis	No	Yes
3	rs4876611	Omega_3	0.021	0.003	1.45E-12	G	A	0.717	8	116671848	150018	cis	Yes	No
3	rs28601761	Omega_3	-0.087	0.003	1.00E-200	G	C	0.421	8	126500031	20849	cis	No	No
3	rs3860846	Omega_3	0.02	0.003	1.10E-11	T	C	0.275	8	126631687	28787	cis	No	No
3	rs13438888	Omega_3	-0.015	0.003	3.19E-08	T	A	0.510	8	141841942	9428	cis	No	Yes
3	rs34433667	Omega_3	-0.015	0.003	4.75E-08	C	A	0.345	9	18904296	69370	cis	Yes	Yes
3	rs2740488	Omega_3	-0.022	0.003	2.39E-13	C	A	0.264	9	107661742	7440	cis	Yes	No
3	rs2236514	Omega_3	-0.019	0.003	7.34E-12	G	C	0.661	9	139572068	7013	cis	Yes	Yes

The results for the other 44 metabolites are provided in the appendix file, “Supplementary Table 6.2.”

*The distance is the absolute base-pair difference between each single nucleotide polymorphism (SNP) position and the Transcription Start Site (TSS) of a gene. SE, standard error. Chr, chromosome. A1, effect allele. A2, other allele. EAF, effect allele frequency. IV, if the instrumental variable was used in Mendelian randomisation.

Abbreviations are in line with **Supplementary Table 3.1**.

Supplementary Table 6.3 Number of genome-wide significant single nucleotide polymorphisms identified for 47 metabolites

ID	Metabolite	Number of significant SNPs	Number of novel SNPs	
1	HDL_P	141	17	12.1%
2	VLDL_size	164	112	68.3%
3	Omega_3	118	20	16.9%
4	PUFA_pct	103	4	3.9%
5	LA_pct	101	35	34.7%
6	Phe	20	11	55.0%
7	Glucose	60	11	18.3%
8	Creatinine	158	36	22.8%
9	Albumin	91	49	53.8%
10	GlycA	137	52	38.0%
11	XXL_VLDL_L	149	112	75.2%
12	XXL_VLDL_FC	144	110	76.4%
13	XS_VLDL_C	125	98	78.4%
14	XS_VLDL_CE	116	46	39.7%
15	XL_HDL_L	173	139	80.3%
16	XL_HDL_C	191	156	81.7%
17	XL_HDL_FC	160	130	81.3%
18	M_HDL_L	137	96	70.1%
19	M_HDL_PL	131	93	71.0%
20	S_HDL_P	105	23	21.9%
21	S_HDL_L	106	85	80.2%
22	S_HDL_PL	119	96	80.7%
23	S_HDL_C	109	84	77.1%
24	S_HDL_CE	100	39	39.0%
25	XXL_VLDL_C_pct	51	21	41.2%
26	XXL_VLDL_CE_pct	50	15	30.0%
27	XL_VLDL_C_pct	149	65	43.6%
28	XL_VLDL_CE_pct	145	57	39.3%

ID	Metabolite	Number of significant SNPs	Number of novel SNPs	
29	XL_VLDL_TG_pct	122	49	40.2%
30	M_VLDL_PL_pct	137	56	40.9%
31	M_VLDL_C_pct	177	86	48.6%
32	M_VLDL_CE_pct	176	69	39.2%
33	M_VLDL_FC_pct	158	71	44.9%
34	M_VLDL_TG_pct	169	89	52.7%
35	S_VLDL_C_pct	144	66	45.8%
36	S_VLDL_CE_pct	123	42	34.1%
37	S_VLDL_FC_pct	158	69	43.7%
38	S_VLDL_TG_pct	153	74	48.4%
39	XS_VLDL_C_pct	171	74	43.3%
40	XS_VLDL_CE_pct	177	70	39.5%
41	XS_VLDL_TG_pct	176	90	51.1%
42	M_LDL_CE_pct	84	23	27.4%
43	S_LDL_PL_pct	99	32	32.3%
44	S_LDL_CE_pct	98	27	27.6%
45	XL_HDL_FC_pct	169	84	49.7%
46	L_HDL_FC_pct	150	72	48.0%
47	S_HDL_FC_pct	143	70	49.0%
	Total	6137	3025	49.3%

SNP, single nucleotide polymorphism. Abbreviations are in line with **Supplementary Table 3.1**.

Supplementary Table 6.4 Number of unique genome-wide significant single nucleotide polymorphisms identified for 47 metabolites by chromosome

Chromosome	Unique single nucleotide polymorphisms
1	231
2	170
3	103
4	106
5	95
6	228
7	144
8	125
9	71
10	117
11	162
12	148
13	38
14	43
15	109
16	100
17	112
18	51
19	160
20	85
21	14
22	38

Supplementary Table 6.5 Top 20 single nucleotide polymorphisms with the largest number of associated metabolites

Rank	SNP	Number of associated metabolites	List of associated metabolites
1	rs72836561	36	HDL_P, VLDL_size, PUFA_pct, LA_pct, XXL_VLDL_L, XXL_VLDL_FC, XS_VLDL_C, XS_VLDL_CE, XL_HDL_L, XL_HDL_C, XL_HDL_FC, M_HDL_L, M_HDL_PL, S_HDL_P, S_HDL_L, S_HDL_PL, S_HDL_C, S_HDL_CE, XL_VLDL_C_pct, XL_VLDL_CE_pct, XL_VLDL_TG_pct, M_VLDL_PL_pct, M_VLDL_C_pct, M_VLDL_CE_pct, M_VLDL_FC_pct, M_VLDL_TG_pct, S_VLDL_C_pct, S_VLDL_CE_pct, S_VLDL_FC_pct, S_VLDL_TG_pct, XS_VLDL_C_pct, XS_VLDL_CE_pct, XS_VLDL_TG_pct, M_LDL_CE_pct, XL_HDL_FC_pct, L_HDL_FC_pct
2	rs76895963	33	VLDL_size, Omega_3, PUFA_pct, LA_pct, Glucose, GlycA, XXL_VLDL_L, XXL_VLDL_FC, XL_HDL_L, XL_HDL_C, XL_HDL_FC, S_HDL_P, S_HDL_L, S_HDL_PL, S_HDL_C, S_HDL_CE, XL_VLDL_C_pct, XL_VLDL_CE_pct, XL_VLDL_TG_pct, M_VLDL_PL_pct, M_VLDL_C_pct, M_VLDL_CE_pct, M_VLDL_FC_pct, M_VLDL_TG_pct, S_VLDL_C_pct, S_VLDL_CE_pct, S_VLDL_FC_pct, S_VLDL_TG_pct, XS_VLDL_C_pct, XS_VLDL_CE_pct, XS_VLDL_TG_pct, S_LDL_PL_pct, S_LDL_CE_pct
3	rs13108218	31	HDL_P, VLDL_size, Omega_3, Albumin, GlycA, XXL_VLDL_L, XXL_VLDL_FC, XL_HDL_L, XL_HDL_C, XL_HDL_FC, S_HDL_P, S_HDL_L, S_HDL_PL, S_HDL_C, S_HDL_CE, XL_VLDL_C_pct, XL_VLDL_CE_pct, M_VLDL_C_pct, M_VLDL_CE_pct, M_VLDL_FC_pct, M_VLDL_TG_pct, S_VLDL_C_pct, S_VLDL_FC_pct, S_VLDL_TG_pct, XS_VLDL_C_pct, XS_VLDL_CE_pct, XS_VLDL_TG_pct, M_LDL_CE_pct, S_LDL_PL_pct, S_LDL_CE_pct, L_HDL_FC_pct
4	rs1260326	27	HDL_P, Omega_3, LA_pct, Albumin, XXL_VLDL_L, XL_HDL_L, XL_HDL_C, XL_HDL_FC, M_HDL_L, M_HDL_PL, S_HDL_P, S_HDL_C, S_HDL_CE, XL_VLDL_C_pct, XL_VLDL_CE_pct, XL_VLDL_TG_pct, M_VLDL_C_pct, M_VLDL_CE_pct, M_VLDL_FC_pct, M_VLDL_TG_pct, S_VLDL_C_pct, S_VLDL_CE_pct, S_VLDL_FC_pct, S_VLDL_TG_pct, M_LDL_CE_pct, S_LDL_PL_pct, S_LDL_CE_pct
5	rs2925979	26	HDL_P, VLDL_size, PUFA_pct, XXL_VLDL_L, XXL_VLDL_FC, XL_HDL_L, XL_HDL_C, XL_HDL_FC, M_HDL_L, M_HDL_PL, XL_VLDL_C_pct, XL_VLDL_CE_pct, XL_VLDL_TG_pct, M_VLDL_PL_pct, M_VLDL_C_pct, M_VLDL_CE_pct, M_VLDL_FC_pct, M_VLDL_TG_pct, S_VLDL_C_pct, S_VLDL_FC_pct, S_VLDL_TG_pct, XS_VLDL_C_pct, XS_VLDL_CE_pct, XS_VLDL_TG_pct, XL_HDL_FC_pct, L_HDL_FC_pct
6	rs11591147	25	LA_pct, XS_VLDL_C, XS_VLDL_CE, XL_HDL_L, XL_HDL_C, XL_HDL_FC, XXL_VLDL_C_pct, XXL_VLDL_CE_pct, XL_VLDL_C_pct, XL_VLDL_CE_pct, XL_VLDL_TG_pct, M_VLDL_PL_pct, M_VLDL_C_pct, M_VLDL_CE_pct, M_VLDL_FC_pct, M_VLDL_TG_pct, S_VLDL_C_pct, S_VLDL_CE_pct, S_VLDL_FC_pct, S_VLDL_TG_pct, XS_VLDL_C_pct, XS_VLDL_CE_pct, XS_VLDL_TG_pct, L_HDL_FC_pct,

Rank	SNP	Number of associated metabolites	List of associated metabolites
			S_HDL_FC_pct
7	rs140584594	25	HDL_P, Omega_3, Phe, XS_VLDL_C, XS_VLDL_CE, XL_HDL_FC, M_HDL_L, M_HDL_PL, S_HDL_L, S_HDL_PL, S_HDL_CE, XL_VLDL_C_pct, XL_VLDL_CE_pct, XL_VLDL_TG_pct, M_VLDL_PL_pct, M_VLDL_C_pct, M_VLDL_CE_pct, M_VLDL_FC_pct, M_VLDL_TG_pct, S_VLDL_C_pct, S_VLDL_CE_pct, S_VLDL_TG_pct, S_LDL_PL_pct, L_HDL_FC_pct, S_HDL_FC_pct
8	rs1801689	25	VLDL_size, PUFA_pct, GlycA, XXL_VLDL_L, XXL_VLDL_FC, XS_VLDL_C, XS_VLDL_CE, XXL_VLDL_C_pct, XXL_VLDL_CE_pct, XL_VLDL_C_pct, XL_VLDL_CE_pct, XL_VLDL_TG_pct, M_VLDL_PL_pct, M_VLDL_C_pct, M_VLDL_CE_pct, M_VLDL_FC_pct, M_VLDL_TG_pct, S_VLDL_C_pct, S_VLDL_CE_pct, S_VLDL_FC_pct, S_VLDL_TG_pct, XS_VLDL_C_pct, XS_VLDL_CE_pct, XS_VLDL_TG_pct, L_HDL_FC_pct
9	rs58542926	25	HDL_P, VLDL_size, LA_pct, GlycA, XXL_VLDL_L, XXL_VLDL_FC, XS_VLDL_C, XS_VLDL_CE, XL_HDL_L, XL_HDL_C, XL_HDL_FC, M_HDL_L, M_HDL_PL, S_HDL_P, S_HDL_L, S_HDL_PL, S_HDL_C, S_HDL_CE, XL_VLDL_C_pct, XL_VLDL_CE_pct, M_VLDL_CE_pct, M_LDL_CE_pct, S_LDL_PL_pct, S_LDL_CE_pct, XL_HDL_FC_pct
10	rs4722551	24	VLDL_size, Omega_3, GlycA, XXL_VLDL_L, XXL_VLDL_FC, XXL_VLDL_C_pct, XXL_VLDL_CE_pct, XL_VLDL_C_pct, XL_VLDL_CE_pct, XL_VLDL_TG_pct, M_VLDL_PL_pct, M_VLDL_C_pct, M_VLDL_CE_pct, M_VLDL_FC_pct, M_VLDL_TG_pct, S_VLDL_C_pct, S_VLDL_CE_pct, S_VLDL_FC_pct, S_VLDL_TG_pct, XS_VLDL_C_pct, XS_VLDL_CE_pct, XS_VLDL_TG_pct, M_LDL_CE_pct, L_HDL_FC_pct
11	rs77542162	23	LA_pct, Albumin, XS_VLDL_C, XS_VLDL_CE, XXL_VLDL_C_pct, XXL_VLDL_CE_pct, XL_VLDL_C_pct, XL_VLDL_CE_pct, XL_VLDL_TG_pct, M_VLDL_PL_pct, M_VLDL_C_pct, M_VLDL_CE_pct, M_VLDL_FC_pct, M_VLDL_TG_pct, S_VLDL_C_pct, S_VLDL_CE_pct, S_VLDL_FC_pct, S_VLDL_TG_pct, XS_VLDL_C_pct, XS_VLDL_CE_pct, XS_VLDL_TG_pct, L_HDL_FC_pct, S_HDL_FC_pct
12	rs117733303	22	PUFA_pct, GlycA, S_HDL_L, S_HDL_PL, XXL_VLDL_C_pct, XXL_VLDL_CE_pct, XL_VLDL_TG_pct, M_VLDL_PL_pct, M_VLDL_C_pct, M_VLDL_CE_pct, M_VLDL_FC_pct, M_VLDL_TG_pct, S_VLDL_C_pct, S_VLDL_CE_pct, S_VLDL_FC_pct, S_VLDL_TG_pct, XS_VLDL_C_pct, XS_VLDL_CE_pct, XS_VLDL_TG_pct, M_LDL_CE_pct, S_LDL_PL_pct, S_LDL_CE_pct
13	rs144503444	22	VLDL_size, PUFA_pct, GlycA, XXL_VLDL_L, XXL_VLDL_FC, XL_HDL_L, XL_HDL_C, XL_HDL_FC, S_HDL_P, S_HDL_C, S_HDL_CE, XL_VLDL_C_pct, M_VLDL_PL_pct, M_VLDL_C_pct, M_VLDL_CE_pct, M_VLDL_FC_pct, S_VLDL_FC_pct, S_VLDL_TG_pct, XS_VLDL_C_pct, XS_VLDL_TG_pct, S_LDL_PL_pct,

Rank	SNP	Number of associated metabolites	List of associated metabolites
			S_LDL_CE_pct
14	rs9844972	22	VLDL_size, PUFA_pct, XXL_VLDL_L, XXL_VLDL_FC, XL_HDL_L, XL_HDL_C, XL_HDL_FC, XL_VLDL_C_pct, XL_VLDL_CE_pct, XL_VLDL_TG_pct, M_VLDL_PL_pct, M_VLDL_C_pct, M_VLDL_CE_pct, M_VLDL_FC_pct, M_VLDL_TG_pct, S_VLDL_C_pct, S_VLDL_FC_pct, S_VLDL_TG_pct, XS_VLDL_C_pct, XS_VLDL_CE_pct, XS_VLDL_TG_pct, L_HDL_FC_pct
15	rs10750766	21	VLDL_size, GlycA, XXL_VLDL_L, XL_HDL_C, XXL_VLDL_C_pct, XXL_VLDL_CE_pct, XL_VLDL_C_pct, XL_VLDL_CE_pct, XL_VLDL_TG_pct, M_VLDL_PL_pct, M_VLDL_C_pct, M_VLDL_CE_pct, M_VLDL_FC_pct, M_VLDL_TG_pct, S_VLDL_C_pct, S_VLDL_CE_pct, S_VLDL_FC_pct, S_VLDL_TG_pct, XS_VLDL_C_pct, XS_VLDL_CE_pct, XS_VLDL_TG_pct
16	rs2618566	21	XS_VLDL_CE, XL_HDL_L, XL_HDL_C, XL_HDL_FC, XL_VLDL_C_pct, XL_VLDL_CE_pct, XL_VLDL_TG_pct, M_VLDL_PL_pct, M_VLDL_C_pct, M_VLDL_CE_pct, M_VLDL_FC_pct, M_VLDL_TG_pct, S_VLDL_C_pct, S_VLDL_CE_pct, S_VLDL_FC_pct, S_VLDL_TG_pct, XS_VLDL_C_pct, XS_VLDL_CE_pct, XS_VLDL_TG_pct, L_HDL_FC_pct, S_HDL_FC_pct
17	rs7534572	21	XS_VLDL_C, XS_VLDL_CE, XL_HDL_C, XL_HDL_FC, XL_VLDL_C_pct, XL_VLDL_CE_pct, XL_VLDL_TG_pct, M_VLDL_PL_pct, M_VLDL_C_pct, M_VLDL_CE_pct, M_VLDL_FC_pct, M_VLDL_TG_pct, S_VLDL_C_pct, S_VLDL_CE_pct, S_VLDL_FC_pct, S_VLDL_TG_pct, XS_VLDL_C_pct, XS_VLDL_CE_pct, XS_VLDL_TG_pct, L_HDL_FC_pct, S_HDL_FC_pct
18	rs77960347	21	HDL_P, Omega_3, XXL_VLDL_L, XXL_VLDL_FC, XS_VLDL_C, XS_VLDL_CE, XL_HDL_L, M_HDL_L, M_HDL_PL, S_HDL_P, S_HDL_L, S_HDL_PL, S_HDL_C, S_HDL_CE, S_VLDL_CE_pct, XS_VLDL_C_pct, XS_VLDL_CE_pct, M_LDL_CE_pct, S_LDL_PL_pct, S_LDL_CE_pct, XL_HDL_FC_pct
19	rs998584	21	VLDL_size, PUFA_pct, GlycA, XXL_VLDL_L, XXL_VLDL_FC, XL_HDL_L, XL_HDL_C, XL_HDL_FC, XL_VLDL_C_pct, XL_VLDL_CE_pct, M_VLDL_C_pct, M_VLDL_CE_pct, M_VLDL_TG_pct, S_VLDL_C_pct, S_VLDL_FC_pct, S_VLDL_TG_pct, XS_VLDL_C_pct, XS_VLDL_CE_pct, S_LDL_PL_pct, S_LDL_CE_pct, XL_HDL_FC_pct
20	rs577721086	20	PUFA_pct, LA_pct, XL_HDL_L, XL_HDL_C, XL_HDL_FC, XL_VLDL_C_pct, XL_VLDL_CE_pct, XL_VLDL_TG_pct, M_VLDL_PL_pct, M_VLDL_C_pct, M_VLDL_CE_pct, M_VLDL_FC_pct, M_VLDL_TG_pct, S_VLDL_C_pct, S_VLDL_FC_pct, S_VLDL_TG_pct, XS_VLDL_C_pct, XS_VLDL_CE_pct, XS_VLDL_TG_pct, L_HDL_FC_pct

Abbreviations are in line with **Supplementary Table 3.1**.

Supplementary Table 6.6 All results from Mendelian randomisation analyses of association of metabolites with ischaemic stroke

ID	Metabolite	Method	N_{SNP}	Beta	SE	p value	FDR	OR (95% CI)
1	HDL_P	Inverse variance weighted	113	-0.076	0.039	4.78E-02	1.41E-01	0.93 (0.86-1.00)
1	HDL_P	MR Egger	113	-0.089	0.070	2.03E-01	3.17E-01	0.91 (0.80-1.05)
1	HDL_P	Simple mode	113	-0.146	0.068	3.34E-02	2.18E-01	0.86 (0.76-0.99)
1	HDL_P	Weighted median	113	-0.084	0.038	2.89E-02	1.51E-01	0.92 (0.85-0.99)
1	HDL_P	Weighted mode	113	-0.106	0.042	1.31E-02	1.05E-01	0.90 (0.83-0.98)
2	VLDL_size	Inverse variance weighted	143	0.042	0.027	1.24E-01	2.78E-01	1.04 (0.99-1.10)
2	VLDL_size	MR Egger	143	-0.044	0.042	3.03E-01	4.06E-01	0.96 (0.88-1.04)
2	VLDL_size	Simple mode	143	0.265	0.078	8.61E-04	4.05E-02	1.30 (1.12-1.52)
2	VLDL_size	Weighted median	143	0.008	0.032	8.13E-01	8.49E-01	1.01 (0.95-1.07)
2	VLDL_size	Weighted mode	143	-0.030	0.036	4.01E-01	5.23E-01	0.97 (0.90-1.04)
3	Omega_3	Inverse variance weighted	88	0.016	0.031	6.07E-01	7.92E-01	1.02 (0.96-1.08)
3	Omega_3	MR Egger	88	-0.023	0.047	6.24E-01	6.82E-01	0.98 (0.89-1.07)
3	Omega_3	Simple mode	88	0.093	0.070	1.88E-01	4.41E-01	1.10 (0.96-1.26)
3	Omega_3	Weighted median	88	-0.017	0.038	6.50E-01	8.04E-01	0.98 (0.91-1.06)
3	Omega_3	Weighted mode	88	0.007	0.032	8.26E-01	8.82E-01	1.01 (0.95-1.07)
4	PUFA_pct	Inverse variance weighted	89	-0.014	0.039	7.11E-01	8.71E-01	0.99 (0.91-1.06)
4	PUFA_pct	MR Egger	89	0.118	0.064	6.88E-02	1.35E-01	1.13 (0.99-1.28)
4	PUFA_pct	Simple mode	89	-0.163	0.114	1.58E-01	4.29E-01	0.85 (0.68-1.06)
4	PUFA_pct	Weighted median	89	0.026	0.045	5.56E-01	7.26E-01	1.03 (0.94-1.12)
4	PUFA_pct	Weighted mode	89	0.065	0.048	1.80E-01	3.25E-01	1.07 (0.97-1.17)
5	LA_pct	Inverse variance weighted	88	-0.076	0.060	2.01E-01	4.11E-01	0.93 (0.82-1.04)
5	LA_pct	MR Egger	88	0.166	0.117	1.58E-01	2.65E-01	1.18 (0.94-1.49)
5	LA_pct	Simple mode	88	-0.098	0.109	3.73E-01	6.26E-01	0.91 (0.73-1.12)
5	LA_pct	Weighted median	88	0.019	0.057	7.34E-01	8.11E-01	1.02 (0.91-1.14)
5	LA_pct	Weighted mode	88	-0.003	0.060	9.62E-01	9.69E-01	1.00 (0.89-1.12)
6	Phe	Inverse variance weighted	17	-0.126	0.075	9.26E-02	2.29E-01	0.88 (0.76-1.02)
6	Phe	MR Egger	17	-0.027	0.132	8.40E-01	8.58E-01	0.97 (0.75-1.26)
6	Phe	Simple mode	17	-0.071	0.110	5.29E-01	7.02E-01	0.93 (0.75-1.16)
6	Phe	Weighted median	17	-0.088	0.062	1.55E-01	3.80E-01	0.92 (0.81-1.03)
6	Phe	Weighted mode	17	-0.077	0.060	2.14E-01	3.53E-01	0.93 (0.82-1.04)

ID	Metabolite	Method	N_{SNP}	Beta	SE	p value	FDR	OR (95% CI)
7	Glucose	Inverse variance weighted	50	0.176	0.060	3.60E-03	6.20E-02	1.19 (1.06-1.34)
7	Glucose	MR Egger	50	0.252	0.131	6.10E-02	1.34E-01	1.29 (0.99-1.66)
7	Glucose	Simple mode	50	0.132	0.140	3.48E-01	6.26E-01	1.14 (0.87-1.50)
7	Glucose	Weighted median	50	0.093	0.066	1.57E-01	3.80E-01	1.10 (0.96-1.25)
7	Glucose	Weighted mode	50	0.109	0.073	1.40E-01	3.14E-01	1.12 (0.97-1.29)
8	Creatinine	Inverse variance weighted	132	-0.007	0.036	8.38E-01	9.09E-01	0.99 (0.92-1.07)
8	Creatinine	MR Egger	132	-0.082	0.095	3.90E-01	4.82E-01	0.92 (0.76-1.11)
8	Creatinine	Simple mode	132	-0.231	0.134	8.64E-02	3.34E-01	0.79 (0.61-1.03)
8	Creatinine	Weighted median	132	-0.093	0.052	7.42E-02	3.19E-01	0.91 (0.82-1.01)
8	Creatinine	Weighted mode	132	-0.132	0.093	1.58E-01	3.24E-01	0.88 (0.73-1.05)
9	Albumin	Inverse variance weighted	75	-0.039	0.045	3.86E-01	6.02E-01	0.96 (0.88-1.05)
9	Albumin	MR Egger	75	-0.010	0.077	8.96E-01	8.96E-01	0.99 (0.85-1.15)
9	Albumin	Simple mode	75	-0.071	0.128	5.82E-01	7.20E-01	0.93 (0.72-1.20)
9	Albumin	Weighted median	75	-0.084	0.062	1.74E-01	3.80E-01	0.92 (0.81-1.04)
9	Albumin	Weighted mode	75	-0.071	0.060	2.40E-01	3.53E-01	0.93 (0.83-1.05)
10	GlycA	Inverse variance weighted	104	0.021	0.036	5.50E-01	7.61E-01	1.02 (0.95-1.10)
10	GlycA	MR Egger	104	-0.054	0.058	3.56E-01	4.65E-01	0.95 (0.85-1.06)
10	GlycA	Simple mode	104	0.020	0.103	8.46E-01	9.53E-01	1.02 (0.83-1.25)
10	GlycA	Weighted median	104	-0.049	0.043	2.53E-01	4.10E-01	0.95 (0.88-1.04)
10	GlycA	Weighted mode	104	-0.050	0.042	2.33E-01	3.53E-01	0.95 (0.88-1.03)
11	XXL_VLDL_L	Inverse variance weighted	124	0.038	0.031	2.30E-01	4.33E-01	1.04 (0.98-1.10)
11	XXL_VLDL_L	MR Egger	124	-0.086	0.046	6.53E-02	1.34E-01	0.92 (0.84-1.00)
11	XXL_VLDL_L	Simple mode	124	0.216	0.105	4.04E-02	2.18E-01	1.24 (1.01-1.52)
11	XXL_VLDL_L	Weighted median	124	-0.026	0.037	4.83E-01	6.67E-01	0.97 (0.91-1.05)
11	XXL_VLDL_L	Weighted mode	124	-0.048	0.036	1.89E-01	3.28E-01	0.95 (0.89-1.02)
12	XXL_VLDL_FC	Inverse variance weighted	123	0.006	0.025	8.18E-01	9.09E-01	1.01 (0.96-1.06)
12	XXL_VLDL_FC	MR Egger	123	-0.072	0.036	4.60E-02	1.08E-01	0.93 (0.87-1.00)
12	XXL_VLDL_FC	Simple mode	123	0.163	0.081	4.64E-02	2.18E-01	1.18 (1.00-1.38)
12	XXL_VLDL_FC	Weighted median	123	-0.035	0.033	2.84E-01	4.30E-01	0.97 (0.91-1.03)
12	XXL_VLDL_FC	Weighted mode	123	-0.071	0.031	2.10E-02	1.24E-01	0.93 (0.88-0.99)
13	XS_VLDL_C	Inverse variance weighted	103	0.080	0.033	1.50E-02	6.31E-02	1.08 (1.02-1.15)
13	XS_VLDL_C	MR Egger	103	0.160	0.050	1.68E-03	7.91E-03	1.17 (1.06-1.29)

ID	Metabolite	Method	N_{SNP}	Beta	SE	p value	FDR	OR (95% CI)
13	XS_VLDL_C	Simple mode	103	0.164	0.078	3.75E-02	2.18E-01	1.18 (1.01-1.37)
13	XS_VLDL_C	Weighted median	103	0.096	0.036	7.75E-03	9.10E-02	1.10 (1.03-1.18)
13	XS_VLDL_C	Weighted mode	103	0.122	0.047	9.97E-03	1.05E-01	1.13 (1.03-1.24)
14	XS_VLDL_CE	Inverse variance weighted	92	0.086	0.034	1.20E-02	6.29E-02	1.09 (1.02-1.17)
14	XS_VLDL_CE	MR Egger	92	0.163	0.052	2.11E-03	9.02E-03	1.18 (1.06-1.30)
14	XS_VLDL_CE	Simple mode	92	0.151	0.072	3.93E-02	2.18E-01	1.16 (1.01-1.34)
14	XS_VLDL_CE	Weighted median	92	0.089	0.040	2.51E-02	1.47E-01	1.09 (1.01-1.18)
14	XS_VLDL_CE	Weighted mode	92	0.100	0.045	2.85E-02	1.49E-01	1.11 (1.01-1.21)
15	XL_HDL_L	Inverse variance weighted	145	-0.088	0.033	7.25E-03	6.29E-02	0.92 (0.86-0.98)
15	XL_HDL_L	MR Egger	145	0.041	0.051	4.24E-01	5.11E-01	1.04 (0.94-1.15)
15	XL_HDL_L	Simple mode	145	-0.059	0.074	4.22E-01	6.84E-01	0.94 (0.82-1.09)
15	XL_HDL_L	Weighted median	145	-0.044	0.034	2.00E-01	3.80E-01	0.96 (0.90-1.02)
15	XL_HDL_L	Weighted mode	145	-0.032	0.030	2.94E-01	4.19E-01	0.97 (0.91-1.03)
16	XL_HDL_C	Inverse variance weighted	157	-0.065	0.033	5.18E-02	1.41E-01	0.94 (0.88-1.00)
16	XL_HDL_C	MR Egger	157	0.056	0.051	2.71E-01	3.74E-01	1.06 (0.96-1.17)
16	XL_HDL_C	Simple mode	157	-0.102	0.073	1.64E-01	4.29E-01	0.90 (0.78-1.04)
16	XL_HDL_C	Weighted median	157	-0.036	0.034	2.79E-01	4.30E-01	0.96 (0.90-1.03)
16	XL_HDL_C	Weighted mode	157	-0.011	0.031	7.30E-01	8.17E-01	0.99 (0.93-1.05)
17	XL_HDL_FC	Inverse variance weighted	126	-0.041	0.033	2.16E-01	4.24E-01	0.96 (0.90-1.02)
17	XL_HDL_FC	MR Egger	126	0.054	0.049	2.68E-01	3.74E-01	1.06 (0.96-1.16)
17	XL_HDL_FC	Simple mode	126	-0.046	0.065	4.82E-01	6.86E-01	0.96 (0.84-1.08)
17	XL_HDL_FC	Weighted median	126	-0.012	0.035	7.35E-01	8.11E-01	0.99 (0.92-1.06)
17	XL_HDL_FC	Weighted mode	126	-0.028	0.028	3.27E-01	4.53E-01	0.97 (0.92-1.03)
18	M_HDL_L	Inverse variance weighted	119	-0.094	0.033	3.96E-03	6.20E-02	0.91 (0.85-0.97)
18	M_HDL_L	MR Egger	119	-0.067	0.055	2.29E-01	3.36E-01	0.94 (0.84-1.04)
18	M_HDL_L	Simple mode	119	-0.115	0.070	1.03E-01	3.44E-01	0.89 (0.78-1.02)
18	M_HDL_L	Weighted median	119	-0.045	0.036	2.11E-01	3.80E-01	0.96 (0.89-1.03)
18	M_HDL_L	Weighted mode	119	-0.069	0.036	6.01E-02	1.88E-01	0.93 (0.87-1.00)
19	M_HDL_PL	Inverse variance weighted	115	-0.085	0.033	1.03E-02	6.29E-02	0.92 (0.86-0.98)
19	M_HDL_PL	MR Egger	115	-0.072	0.057	2.09E-01	3.17E-01	0.93 (0.83-1.04)
19	M_HDL_PL	Simple mode	115	-0.075	0.070	2.82E-01	5.30E-01	0.93 (0.81-1.06)
19	M_HDL_PL	Weighted median	115	-0.051	0.036	1.53E-01	3.80E-01	0.95 (0.89-1.02)

ID	Metabolite	Method	N_{SNP}	Beta	SE	p value	FDR	OR (95% CI)
19	M_HDL_PL	Weighted mode	115	-0.052	0.035	1.36E-01	3.14E-01	0.95 (0.89-1.02)
20	S_HDL_P	Inverse variance weighted	80	-0.048	0.041	2.40E-01	4.34E-01	0.95 (0.88-1.03)
20	S_HDL_P	MR Egger	80	-0.035	0.077	6.50E-01	6.94E-01	0.97 (0.83-1.12)
20	S_HDL_P	Simple mode	80	-0.118	0.082	1.57E-01	4.29E-01	0.89 (0.76-1.04)
20	S_HDL_P	Weighted median	80	-0.019	0.044	6.67E-01	8.04E-01	0.98 (0.90-1.07)
20	S_HDL_P	Weighted mode	80	-0.031	0.046	5.01E-01	6.04E-01	0.97 (0.89-1.06)
21	S_HDL_L	Inverse variance weighted	82	-0.009	0.036	8.08E-01	9.09E-01	0.99 (0.92-1.06)
21	S_HDL_L	MR Egger	82	-0.037	0.058	5.29E-01	6.22E-01	0.96 (0.86-1.08)
21	S_HDL_L	Simple mode	82	-0.132	0.078	9.24E-02	3.34E-01	0.88 (0.75-1.02)
21	S_HDL_L	Weighted median	82	-0.006	0.039	8.69E-01	8.69E-01	0.99 (0.92-1.07)
21	S_HDL_L	Weighted mode	82	-0.026	0.036	4.64E-01	5.73E-01	0.97 (0.91-1.04)
22	S_HDL_PL	Inverse variance weighted	90	0.005	0.035	8.85E-01	9.25E-01	1.01 (0.94-1.08)
22	S_HDL_PL	MR Egger	90	-0.051	0.058	3.87E-01	4.82E-01	0.95 (0.85-1.07)
22	S_HDL_PL	Simple mode	90	-0.111	0.088	2.11E-01	4.46E-01	0.90 (0.75-1.06)
22	S_HDL_PL	Weighted median	90	-0.007	0.040	8.65E-01	8.69E-01	0.99 (0.92-1.07)
22	S_HDL_PL	Weighted mode	90	-0.022	0.037	5.55E-01	6.52E-01	0.98 (0.91-1.05)
23	S_HDL_C	Inverse variance weighted	80	-0.039	0.040	3.23E-01	5.23E-01	0.96 (0.89-1.04)
23	S_HDL_C	MR Egger	80	-0.038	0.067	5.79E-01	6.64E-01	0.96 (0.84-1.10)
23	S_HDL_C	Simple mode	80	-0.099	0.079	2.18E-01	4.46E-01	0.91 (0.78-1.06)
23	S_HDL_C	Weighted median	80	0.013	0.041	7.42E-01	8.11E-01	1.01 (0.94-1.10)
23	S_HDL_C	Weighted mode	80	-0.014	0.038	7.19E-01	8.17E-01	0.99 (0.91-1.06)
24	S_HDL_CE	Inverse variance weighted	73	-0.024	0.033	4.66E-01	6.64E-01	0.98 (0.92-1.04)
24	S_HDL_CE	MR Egger	73	-0.030	0.057	5.93E-01	6.64E-01	0.97 (0.87-1.08)
24	S_HDL_CE	Simple mode	73	-0.054	0.071	4.47E-01	6.86E-01	0.95 (0.82-1.09)
24	S_HDL_CE	Weighted median	73	0.018	0.039	6.43E-01	8.04E-01	1.02 (0.94-1.10)
24	S_HDL_CE	Weighted mode	73	0.001	0.036	9.69E-01	9.69E-01	1.00 (0.93-1.08)
25	XXL_VLDL_C_pct	Inverse variance weighted	37	0.096	0.047	4.30E-02	1.41E-01	1.10 (1.00-1.21)
25	XXL_VLDL_C_pct	MR Egger	37	0.190	0.069	9.29E-03	2.91E-02	1.21 (1.06-1.38)
25	XXL_VLDL_C_pct	Simple mode	37	0.206	0.086	2.17E-02	2.18E-01	1.23 (1.04-1.45)
25	XXL_VLDL_C_pct	Weighted median	37	0.156	0.048	1.21E-03	1.95E-02	1.17 (1.06-1.29)
25	XXL_VLDL_C_pct	Weighted mode	37	0.146	0.041	1.11E-03	2.16E-02	1.16 (1.07-1.25)
26	XXL_VLDL_CE_pct	Inverse variance weighted	37	0.134	0.052	1.02E-02	6.29E-02	1.14 (1.03-1.27)

ID	Metabolite	Method	N_{SNP}	Beta	SE	p value	FDR	OR (95% CI)
26	XXL_VLDL_CE_pct	MR Egger	37	0.190	0.080	2.29E-02	5.99E-02	1.21 (1.03-1.41)
26	XXL_VLDL_CE_pct	Simple mode	37	0.201	0.105	6.41E-02	2.74E-01	1.22 (0.99-1.50)
26	XXL_VLDL_CE_pct	Weighted median	37	0.178	0.049	3.18E-04	1.50E-02	1.20 (1.08-1.32)
26	XXL_VLDL_CE_pct	Weighted mode	37	0.170	0.049	1.38E-03	2.16E-02	1.18 (1.08-1.30)
27	XL_VLDL_C_pct	Inverse variance weighted	119	0.028	0.034	3.97E-01	6.02E-01	1.03 (0.96-1.10)
27	XL_VLDL_C_pct	MR Egger	119	0.172	0.050	7.08E-04	3.81E-03	1.19 (1.08-1.31)
27	XL_VLDL_C_pct	Simple mode	119	-0.034	0.089	7.07E-01	8.51E-01	0.97 (0.81-1.15)
27	XL_VLDL_C_pct	Weighted median	119	0.051	0.035	1.50E-01	3.80E-01	1.05 (0.98-1.13)
27	XL_VLDL_C_pct	Weighted mode	119	0.074	0.035	3.89E-02	1.66E-01	1.08 (1.00-1.15)
28	XL_VLDL_CE_pct	Inverse variance weighted	123	0.017	0.034	6.07E-01	7.92E-01	1.02 (0.95-1.09)
28	XL_VLDL_CE_pct	MR Egger	123	0.148	0.051	4.65E-03	1.56E-02	1.16 (1.05-1.28)
28	XL_VLDL_CE_pct	Simple mode	123	0.014	0.099	8.85E-01	9.55E-01	1.01 (0.84-1.23)
28	XL_VLDL_CE_pct	Weighted median	123	0.049	0.040	2.18E-01	3.80E-01	1.05 (0.97-1.13)
28	XL_VLDL_CE_pct	Weighted mode	123	0.075	0.042	7.75E-02	2.14E-01	1.08 (0.99-1.17)
29	XL_VLDL_TG_pct	Inverse variance weighted	100	-0.083	0.034	1.61E-02	6.31E-02	0.92 (0.86-0.98)
29	XL_VLDL_TG_pct	MR Egger	100	-0.215	0.049	3.00E-05	1.41E-03	0.81 (0.73-0.89)
29	XL_VLDL_TG_pct	Simple mode	100	0.047	0.070	5.00E-01	6.91E-01	1.05 (0.91-1.20)
29	XL_VLDL_TG_pct	Weighted median	100	-0.116	0.036	1.25E-03	1.95E-02	0.89 (0.83-0.96)
29	XL_VLDL_TG_pct	Weighted mode	100	-0.114	0.033	8.24E-04	2.16E-02	0.89 (0.84-0.95)
30	M_VLDL_PL_pct	Inverse variance weighted	112	0.087	0.036	1.61E-02	6.31E-02	1.09 (1.02-1.17)
30	M_VLDL_PL_pct	MR Egger	112	0.209	0.053	1.54E-04	2.39E-03	1.23 (1.11-1.37)
30	M_VLDL_PL_pct	Simple mode	112	0.060	0.066	3.65E-01	6.26E-01	1.06 (0.93-1.21)
30	M_VLDL_PL_pct	Weighted median	112	0.081	0.035	2.02E-02	1.47E-01	1.08 (1.01-1.16)
30	M_VLDL_PL_pct	Weighted mode	112	0.070	0.035	4.60E-02	1.80E-01	1.07 (1.00-1.15)
31	M_VLDL_C_pct	Inverse variance weighted	141	0.026	0.034	4.40E-01	6.46E-01	1.03 (0.96-1.10)
31	M_VLDL_C_pct	MR Egger	141	0.188	0.049	1.62E-04	2.39E-03	1.21 (1.10-1.33)
31	M_VLDL_C_pct	Simple mode	141	-0.049	0.080	5.38E-01	7.02E-01	0.95 (0.81-1.11)
31	M_VLDL_C_pct	Weighted median	141	0.060	0.035	8.16E-02	3.19E-01	1.06 (0.99-1.14)
31	M_VLDL_C_pct	Weighted mode	141	0.071	0.037	5.51E-02	1.85E-01	1.07 (1.00-1.15)
32	M_VLDL_CE_pct	Inverse variance weighted	145	0.011	0.032	7.23E-01	8.71E-01	1.01 (0.95-1.08)
32	M_VLDL_CE_pct	MR Egger	145	0.162	0.046	5.72E-04	3.81E-03	1.18 (1.07-1.29)
32	M_VLDL_CE_pct	Simple mode	145	-0.097	0.089	2.82E-01	5.30E-01	0.91 (0.76-1.08)

ID	Metabolite	Method	N_{SNP}	Beta	SE	p value	FDR	OR (95% CI)
32	M_VLDL_CE_pct	Weighted median	145	0.043	0.034	2.05E-01	3.80E-01	1.04 (0.98-1.11)
32	M_VLDL_CE_pct	Weighted mode	145	0.052	0.034	1.31E-01	3.14E-01	1.05 (0.98-1.13)
33	M_VLDL_FC_pct	Inverse variance weighted	128	0.049	0.036	1.75E-01	3.75E-01	1.05 (0.98-1.13)
33	M_VLDL_FC_pct	MR Egger	128	0.162	0.054	3.38E-03	1.22E-02	1.18 (1.06-1.31)
33	M_VLDL_FC_pct	Simple mode	128	-0.026	0.081	7.46E-01	8.77E-01	0.97 (0.83-1.14)
33	M_VLDL_FC_pct	Weighted median	128	0.044	0.034	2.02E-01	3.80E-01	1.04 (0.98-1.12)
33	M_VLDL_FC_pct	Weighted mode	128	0.042	0.034	2.20E-01	3.53E-01	1.04 (0.98-1.11)
34	M_VLDL_TG_pct	Inverse variance weighted	136	-0.053	0.034	1.22E-01	2.78E-01	0.95 (0.89-1.01)
34	M_VLDL_TG_pct	MR Egger	136	-0.192	0.050	2.04E-04	2.39E-03	0.83 (0.75-0.91)
34	M_VLDL_TG_pct	Simple mode	136	0.063	0.088	4.71E-01	6.86E-01	1.07 (0.90-1.27)
34	M_VLDL_TG_pct	Weighted median	136	-0.062	0.035	8.11E-02	3.19E-01	0.94 (0.88-1.01)
34	M_VLDL_TG_pct	Weighted mode	136	-0.053	0.039	1.75E-01	3.25E-01	0.95 (0.88-1.02)
35	S_VLDL_C_pct	Inverse variance weighted	120	0.082	0.036	2.20E-02	7.94E-02	1.09 (1.01-1.16)
35	S_VLDL_C_pct	MR Egger	120	0.195	0.054	5.09E-04	3.81E-03	1.22 (1.09-1.35)
35	S_VLDL_C_pct	Simple mode	120	0.001	0.088	9.90E-01	9.90E-01	1.00 (0.84-1.19)
35	S_VLDL_C_pct	Weighted median	120	0.101	0.039	9.76E-03	9.17E-02	1.11 (1.02-1.19)
35	S_VLDL_C_pct	Weighted mode	120	0.115	0.046	1.34E-02	1.05E-01	1.12 (1.03-1.23)
36	S_VLDL_CE_pct	Inverse variance weighted	103	0.104	0.033	1.63E-03	6.20E-02	1.11 (1.04-1.18)
36	S_VLDL_CE_pct	MR Egger	103	0.153	0.050	2.95E-03	1.16E-02	1.17 (1.06-1.29)
36	S_VLDL_CE_pct	Simple mode	103	0.088	0.069	2.04E-01	4.46E-01	1.09 (0.95-1.25)
36	S_VLDL_CE_pct	Weighted median	103	0.051	0.034	1.27E-01	3.80E-01	1.05 (0.99-1.12)
36	S_VLDL_CE_pct	Weighted mode	103	0.051	0.036	1.52E-01	3.24E-01	1.05 (0.98-1.13)
37	S_VLDL_FC_pct	Inverse variance weighted	131	0.037	0.035	2.87E-01	4.87E-01	1.04 (0.97-1.11)
37	S_VLDL_FC_pct	MR Egger	131	0.177	0.051	7.29E-04	3.81E-03	1.19 (1.08-1.32)
37	S_VLDL_FC_pct	Simple mode	131	-0.004	0.086	9.67E-01	9.88E-01	1.00 (0.84-1.18)
37	S_VLDL_FC_pct	Weighted median	131	0.059	0.035	9.21E-02	3.33E-01	1.06 (0.99-1.14)
37	S_VLDL_FC_pct	Weighted mode	131	0.075	0.039	5.51E-02	1.85E-01	1.08 (1.00-1.16)
38	S_VLDL_TG_pct	Inverse variance weighted	125	-0.087	0.034	1.09E-02	6.29E-02	0.92 (0.86-0.98)
38	S_VLDL_TG_pct	MR Egger	125	-0.191	0.051	2.64E-04	2.48E-03	0.83 (0.75-0.91)
38	S_VLDL_TG_pct	Simple mode	125	-0.006	0.075	9.39E-01	9.81E-01	0.99 (0.86-1.15)
38	S_VLDL_TG_pct	Weighted median	125	-0.083	0.036	2.26E-02	1.47E-01	0.92 (0.86-0.99)
38	S_VLDL_TG_pct	Weighted mode	125	-0.092	0.038	1.72E-02	1.15E-01	0.91 (0.85-0.98)

ID	Metabolite	Method	N_{SNP}	Beta	SE	p value	FDR	OR (95% CI)
39	XS_VLDL_C_pct	Inverse variance weighted	141	-0.011	0.033	7.45E-01	8.75E-01	0.99 (0.93-1.06)
39	XS_VLDL_C_pct	MR Egger	141	0.122	0.051	1.69E-02	4.69E-02	1.13 (1.02-1.25)
39	XS_VLDL_C_pct	Simple mode	141	-0.010	0.075	8.94E-01	9.55E-01	0.99 (0.86-1.15)
39	XS_VLDL_C_pct	Weighted median	141	0.041	0.032	2.11E-01	3.80E-01	1.04 (0.98-1.11)
39	XS_VLDL_C_pct	Weighted mode	141	0.038	0.031	2.32E-01	3.53E-01	1.04 (0.98-1.10)
40	XS_VLDL_CE_pct	Inverse variance weighted	148	-0.063	0.033	5.41E-02	1.41E-01	0.94 (0.88-1.00)
40	XS_VLDL_CE_pct	MR Egger	148	0.066	0.050	1.90E-01	3.08E-01	1.07 (0.97-1.18)
40	XS_VLDL_CE_pct	Simple mode	148	-0.114	0.078	1.45E-01	4.29E-01	0.89 (0.77-1.04)
40	XS_VLDL_CE_pct	Weighted median	148	-0.010	0.035	7.84E-01	8.37E-01	0.99 (0.92-1.06)
40	XS_VLDL_CE_pct	Weighted mode	148	0.005	0.034	8.85E-01	9.24E-01	1.00 (0.94-1.07)
41	XS_VLDL_TG_pct	Inverse variance weighted	145	-0.006	0.034	8.51E-01	9.09E-01	0.99 (0.93-1.06)
41	XS_VLDL_TG_pct	MR Egger	145	-0.128	0.051	1.32E-02	3.88E-02	0.88 (0.80-0.97)
41	XS_VLDL_TG_pct	Simple mode	145	0.015	0.081	8.52E-01	9.53E-01	1.02 (0.87-1.19)
41	XS_VLDL_TG_pct	Weighted median	145	-0.033	0.033	3.21E-01	4.57E-01	0.97 (0.91-1.03)
41	XS_VLDL_TG_pct	Weighted mode	145	-0.027	0.033	4.21E-01	5.35E-01	0.97 (0.91-1.04)
42	M_LDL_CE_pct	Inverse variance weighted	69	0.003	0.043	9.49E-01	9.63E-01	1.00 (0.92-1.09)
42	M_LDL_CE_pct	MR Egger	69	-0.138	0.063	3.26E-02	8.05E-02	0.87 (0.77-0.99)
42	M_LDL_CE_pct	Simple mode	69	0.262	0.116	2.74E-02	2.18E-01	1.30 (1.03-1.63)
42	M_LDL_CE_pct	Weighted median	69	-0.047	0.045	3.06E-01	4.50E-01	0.95 (0.87-1.04)
42	M_LDL_CE_pct	Weighted mode	69	-0.076	0.044	8.33E-02	2.17E-01	0.93 (0.85-1.01)
43	S_LDL_PL_pct	Inverse variance weighted	83	0.014	0.036	6.95E-01	8.71E-01	1.01 (0.95-1.09)
43	S_LDL_PL_pct	MR Egger	83	0.080	0.050	1.16E-01	2.02E-01	1.08 (0.98-1.20)
43	S_LDL_PL_pct	Simple mode	83	0.158	0.078	4.61E-02	2.18E-01	1.17 (1.01-1.37)
43	S_LDL_PL_pct	Weighted median	83	0.051	0.038	1.85E-01	3.80E-01	1.05 (0.98-1.13)
43	S_LDL_PL_pct	Weighted mode	83	0.064	0.035	6.75E-02	1.98E-01	1.07 (1.00-1.14)
44	S_LDL_CE_pct	Inverse variance weighted	84	-0.032	0.030	2.90E-01	4.87E-01	0.97 (0.91-1.03)
44	S_LDL_CE_pct	MR Egger	84	-0.074	0.043	8.79E-02	1.65E-01	0.93 (0.85-1.01)
44	S_LDL_CE_pct	Simple mode	84	-0.176	0.065	8.36E-03	1.96E-01	0.84 (0.74-0.95)
44	S_LDL_CE_pct	Weighted median	84	-0.047	0.035	1.82E-01	3.80E-01	0.95 (0.89-1.02)
44	S_LDL_CE_pct	Weighted mode	84	-0.074	0.035	3.80E-02	1.66E-01	0.93 (0.87-0.99)
45	XL_HDL_FC_pct	Inverse variance weighted	142	0.079	0.029	6.45E-03	6.29E-02	1.08 (1.02-1.14)
45	XL_HDL_FC_pct	MR Egger	142	-0.011	0.044	7.97E-01	8.32E-01	0.99 (0.91-1.08)

ID	Metabolite	Method	N_{SNP}	Beta	SE	p value	FDR	OR (95% CI)
45	XL_HDL_FC_pct	Simple mode	142	0.066	0.088	4.53E-01	6.86E-01	1.07 (0.90-1.27)
45	XL_HDL_FC_pct	Weighted median	142	0.040	0.033	2.32E-01	3.89E-01	1.04 (0.97-1.11)
45	XL_HDL_FC_pct	Weighted mode	142	0.039	0.029	1.74E-01	3.25E-01	1.04 (0.98-1.10)
46	L_HDL_FC_pct	Inverse variance weighted	120	0.002	0.035	9.63E-01	9.63E-01	1.00 (0.93-1.07)
46	L_HDL_FC_pct	MR Egger	120	0.088	0.054	1.05E-01	1.91E-01	1.09 (0.98-1.21)
46	L_HDL_FC_pct	Simple mode	120	-0.103	0.077	1.87E-01	4.41E-01	0.90 (0.78-1.05)
46	L_HDL_FC_pct	Weighted median	120	-0.014	0.036	6.98E-01	8.11E-01	0.99 (0.92-1.06)
46	L_HDL_FC_pct	Weighted mode	120	-0.029	0.033	3.82E-01	5.13E-01	0.97 (0.91-1.04)
47	S_HDL_FC_pct	Inverse variance weighted	120	0.056	0.028	4.85E-02	1.41E-01	1.06 (1.00-1.12)
47	S_HDL_FC_pct	MR Egger	120	0.081	0.043	6.54E-02	1.34E-01	1.08 (1.00-1.18)
47	S_HDL_FC_pct	Simple mode	120	0.043	0.075	5.68E-01	7.20E-01	1.04 (0.90-1.21)
47	S_HDL_FC_pct	Weighted median	120	0.023	0.036	5.36E-01	7.20E-01	1.02 (0.95-1.10)
47	S_HDL_FC_pct	Weighted mode	120	0.010	0.033	7.70E-01	8.42E-01	1.01 (0.95-1.08)

Abbreviations are in line with **Supplementary Table 3.1**. SNP, single nucleotide polymorphism. SE, standard error. FDR, false discovery rate. OR, odds ratio. CI, confidence interval.

Supplementary Table 6.7 Detailed results from the inverse variance-weighted method and Mendelian randomisation-Egger method of association of metabolites with ischaemic stroke

ID	Metabolite	Method	N_{SNP}	Beta	SE	p value	FDR	OR (95% CI)	Q	Degrees of freedom for Q	p value for Q	FDR for Q	Intercept	$SE_{Intercept}$	p value for intercept	FDR for intercept
1	HDL_P	Inverse variance weighted	113	-0.076	0.039	4.78E-02	1.41E-01	0.93 (0.86-1.00)	296.100	112	1.48E-18	4.09E-18	NA	NA	NA	NA
1	HDL_P	MR Egger	113	-0.089	0.070	2.03E-01	3.17E-01	0.91 (0.80-1.05)	295.967	111	9.39E-19	2.76E-18	0.001	0.002	8.23E-01	8.74E-01
2	VLDL_size	Inverse variance weighted	143	0.042	0.027	1.24E-01	2.78E-01	1.04 (0.99-1.10)	280.859	142	3.55E-11	5.75E-11	NA	NA	NA	NA
2	VLDL_size	MR Egger	143	-0.044	0.042	3.03E-01	4.06E-01	0.96 (0.88-1.04)	267.749	141	6.62E-10	1.07E-09	0.004	0.002	9.55E-03	2.12E-02
3	Omega_3	Inverse variance weighted	88	0.016	0.031	6.07E-01	7.92E-01	1.02 (0.96-1.08)	149.087	87	3.90E-05	4.27E-05	NA	NA	NA	NA
3	Omega_3	MR Egger	88	-0.023	0.047	6.24E-01	6.82E-01	0.98 (0.89-1.07)	147.004	86	4.65E-05	5.20E-05	0.0023	0.002	2.73E-01	3.77E-01
4	PUFA_pct	Inverse variance weighted	89	-0.014	0.039	7.11E-01	8.71E-01	0.99 (0.91-1.06)	174.613	88	1.13E-07	1.44E-07	NA	NA	NA	NA
4	PUFA_pct	MR Egger	89	0.118	0.064	6.88E-02	1.35E-01	1.13 (0.99-1.28)	162.475	87	1.70E-06	2.05E-06	-0.006	0.002	1.25E-02	2.56E-02
5	LA_pct	Inverse variance weighted	88	-0.076	0.060	2.01E-01	4.11E-01	0.93 (0.82-1.04)	283.874	87	1.03E-22	3.71E-22	NA	NA	NA	NA
5	LA_pct	MR Egger	88	0.166	0.117	1.58E-01	2.65E-01	1.18 (0.94-1.49)	266.168	86	2.70E-20	9.05E-20	-0.008	0.003	1.89E-02	3.71E-02
6	Phe	Inverse variance weighted	17	-0.126	0.075	9.26E-02	2.29E-01	0.88 (0.76-1.02)	37.078	16	2.04E-03	2.09E-03	NA	NA	NA	NA
6	Phe	MR Egger	17	-0.027	0.132	8.40E-01	8.58E-01	0.97 (0.75-1.26)	35.149	15	2.34E-03	2.39E-03	-0.005	0.006	3.79E-01	4.94E-01
7	Glucose	Inverse variance weighted	50	0.176	0.060	3.60E-03	6.20E-02	1.19 (1.06-1.34)	104.704	49	6.40E-06	7.34E-06	NA	NA	NA	NA
7	Glucose	MR Egger	50	0.252	0.131	6.10E-02	1.34E-01	1.29 (0.99-1.66)	103.777	48	5.53E-06	6.34E-06	-0.003	0.004	5.16E-01	6.22E-01
8	Creatinine	Inverse variance weighted	132	-0.007	0.036	8.38E-01	9.09E-01	0.99 (0.92-1.07)	176.346	131	5.07E-03	5.07E-03	NA	NA	NA	NA
8	Creatinine	MR Egger	132	-0.082	0.095	3.90E-01	4.82E-01	0.92 (0.76-1.11)	175.371	130	4.93E-03	4.93E-03	0.0019	0.002	3.97E-01	5.04E-01
9	Albumin	Inverse variance weighted	75	-0.039	0.045	3.86E-01	6.02E-01	0.96 (0.88-1.05)	124.915	74	1.99E-04	2.13E-04	NA	NA	NA	NA
9	Albumin	MR Egger	75	-0.010	0.077	8.96E-01	8.96E-01	0.99 (0.85-1.15)	124.569	73	1.61E-04	1.76E-04	-0.001	0.002	6.54E-01	7.32E-01
10	GlycA	Inverse variance weighted	104	0.021	0.036	5.50E-01	7.61E-01	1.02 (0.95-1.10)	206.695	103	6.12E-09	8.72E-09	NA	NA	NA	NA
10	GlycA	MR Egger	104	-0.054	0.058	3.56E-01	4.65E-01	0.95 (0.85-1.06)	201.361	102	1.70E-08	2.34E-08	0.003	0.002	1.03E-01	1.62E-01
11	XXL_VLDL_L	Inverse variance weighted	124	0.038	0.031	2.30E-01	4.33E-01	1.04 (0.98-1.10)	265.737	123	1.62E-12	2.82E-12	NA	NA	NA	NA
11	XXL_VLDL_L	MR Egger	124	-0.086	0.046	6.53E-02	1.34E-01	0.92 (0.84-1.00)	241.374	122	7.21E-10	1.13E-09	0.006	0.002	6.31E-04	4.23E-03
12	XXL_VLDL_FC	Inverse variance weighted	123	0.006	0.025	8.18E-01	9.09E-01	1.01 (0.96-1.06)	195.022	122	2.97E-05	3.33E-05	NA	NA	NA	NA
12	XXL_VLDL_FC	MR Egger	123	-0.072	0.036	4.60E-02	1.08E-01	0.93 (0.87-1.00)	182.173	121	2.74E-04	2.92E-04	0.004	0.001	4.16E-03	1.30E-02
13	XS_VLDL_C	Inverse variance weighted	103	0.080	0.033	1.50E-02	6.31E-02	1.08 (1.02-1.15)	238.410	102	6.22E-13	1.12E-12	NA	NA	NA	NA
13	XS_VLDL_C	MR Egger	103	0.160	0.050	1.68E-03	7.91E-03	1.17 (1.06-1.29)	228.157	101	7.90E-12	1.43E-11	-0.004	0.002	3.56E-02	6.69E-02

ID	Metabolite	Method	N_{SNP}	Beta	SE	p value	FDR	OR (95% CI)	Q	Degrees of freedom for Q	p value for Q	FDR for Q	Intercept	$SE_{Intercept}$	p value for intercept	FDR for intercept
14	XS_VLDL_CE	Inverse variance weighted	92	0.086	0.034	1.20E-02	6.29E-02	1.09 (1.02-1.17)	231.998	91	2.80E-14	5.48E-14	NA	NA	NA	NA
14	XS_VLDL_CE	MR Egger	92	0.163	0.052	2.11E-03	9.02E-03	1.18 (1.06-1.30)	222.340	90	3.41E-13	6.68E-13	-0.005	0.002	5.11E-02	8.89E-02
15	XL_HDL_L	Inverse variance weighted	145	-0.088	0.033	7.25E-03	6.29E-02	0.92 (0.86-0.98)	348.648	144	5.29E-19	1.55E-18	NA	NA	NA	NA
15	XL_HDL_L	MR Egger	145	0.041	0.051	4.24E-01	5.11E-01	1.04 (0.94-1.15)	324.163	143	4.33E-16	1.20E-15	-0.006	0.002	1.28E-03	5.45E-03
16	XL_HDL_C	Inverse variance weighted	157	-0.065	0.033	5.18E-02	1.41E-01	0.94 (0.88-1.00)	442.871	156	2.74E-29	2.86E-28	NA	NA	NA	NA
16	XL_HDL_C	MR Egger	157	0.056	0.051	2.71E-01	3.74E-01	1.06 (0.96-1.17)	416.557	155	7.96E-26	1.25E-24	-0.006	0.002	2.10E-03	8.06E-03
17	XL_HDL_FC	Inverse variance weighted	126	-0.041	0.033	2.16E-01	4.24E-01	0.96 (0.90-1.02)	334.448	125	5.06E-21	1.59E-20	NA	NA	NA	NA
17	XL_HDL_FC	MR Egger	126	0.054	0.049	2.68E-01	3.74E-01	1.06 (0.96-1.16)	316.924	124	7.57E-19	2.37E-18	-0.005	0.002	9.93E-03	2.12E-02
18	M_HDL_L	Inverse variance weighted	119	-0.094	0.033	3.96E-03	6.20E-02	0.91 (0.85-0.97)	271.803	118	3.73E-14	7.02E-14	NA	NA	NA	NA
18	M_HDL_L	MR Egger	119	-0.067	0.055	2.29E-01	3.36E-01	0.94 (0.84-1.04)	270.948	117	3.12E-14	7.32E-14	-0.001	0.002	5.44E-01	6.24E-01
19	M_HDL_PL	Inverse variance weighted	115	-0.085	0.033	1.03E-02	6.29E-02	0.92 (0.86-0.98)	267.880	114	2.08E-14	4.24E-14	NA	NA	NA	NA
19	M_HDL_PL	MR Egger	115	-0.072	0.057	2.09E-01	3.17E-01	0.93 (0.83-1.04)	267.680	113	1.42E-14	3.51E-14	-0.001	0.002	7.72E-01	8.44E-01
20	S_HDL_P	Inverse variance weighted	80	-0.048	0.041	2.40E-01	4.34E-01	0.95 (0.88-1.03)	184.668	79	1.92E-10	2.81E-10	NA	NA	NA	NA
20	S_HDL_P	MR Egger	80	-0.035	0.077	6.50E-01	6.94E-01	0.97 (0.83-1.12)	184.567	78	1.27E-10	2.13E-10	-0.0006	0.003	8.37E-01	8.74E-01
21	S_HDL_L	Inverse variance weighted	82	-0.009	0.036	8.08E-01	9.09E-01	0.99 (0.92-1.06)	172.714	81	1.34E-08	1.86E-08	NA	NA	NA	NA
21	S_HDL_L	MR Egger	82	-0.037	0.058	5.29E-01	6.22E-01	0.96 (0.86-1.08)	171.909	80	1.13E-08	1.61E-08	0.00141	0.002	5.42E-01	6.24E-01
22	S_HDL_PL	Inverse variance weighted	90	0.005	0.035	8.85E-01	9.25E-01	1.01 (0.94-1.08)	183.257	89	1.68E-08	2.25E-08	NA	NA	NA	NA
22	S_HDL_PL	MR Egger	90	-0.051	0.058	3.87E-01	4.82E-01	0.95 (0.85-1.07)	180.281	88	2.55E-08	3.43E-08	0.003	0.002	2.31E-01	3.29E-01
23	S_HDL_C	Inverse variance weighted	80	-0.039	0.040	3.23E-01	5.23E-01	0.96 (0.89-1.04)	184.855	79	1.81E-10	2.75E-10	NA	NA	NA	NA
23	S_HDL_C	MR Egger	80	-0.038	0.067	5.79E-01	6.64E-01	0.96 (0.84-1.10)	184.852	78	1.16E-10	2.03E-10	-0.0001	0.003	9.73E-01	9.73E-01
24	S_HDL_CE	Inverse variance weighted	73	-0.024	0.033	4.66E-01	6.64E-01	0.98 (0.92-1.04)	117.217	72	6.07E-04	6.34E-04	NA	NA	NA	NA
24	S_HDL_CE	MR Egger	73	-0.030	0.057	5.93E-01	6.64E-01	0.97 (0.87-1.08)	117.183	71	4.66E-04	4.86E-04	0.000	0.002	8.87E-01	9.06E-01
25	XXL_VLDL_C_pct	Inverse variance weighted	37	0.096	0.047	4.30E-02	1.41E-01	1.10 (1.00-1.21)	92.694	36	6.82E-07	8.01E-07	NA	NA	NA	NA
25	XXL_VLDL_C_pct	MR Egger	37	0.190	0.069	9.29E-03	2.91E-02	1.21 (1.06-1.38)	84.630	35	5.40E-06	6.34E-06	-0.007	0.004	7.64E-02	1.24E-01
26	XXL_VLDL_CE_pct	Inverse variance weighted	37	0.134	0.052	1.02E-02	6.29E-02	1.14 (1.03-1.27)	96.640	36	1.89E-07	2.27E-07	NA	NA	NA	NA
26	XXL_VLDL_CE_pct	MR Egger	37	0.190	0.080	2.29E-02	5.99E-02	1.21 (1.03-1.41)	94.316	35	2.40E-07	2.97E-07	-0.004	0.004	3.59E-01	4.83E-01
27	XL_VLDL_C_pct	Inverse variance weighted	119	0.028	0.034	3.97E-01	6.02E-01	1.03 (0.96-1.10)	294.912	118	3.85E-17	1.01E-16	NA	NA	NA	NA
27	XL_VLDL_C_pct	MR Egger	119	0.172	0.050	7.08E-04	3.81E-03	1.19 (1.08-1.31)	262.635	117	3.39E-13	6.68E-13	-0.007	0.002	2.38E-04	3.73E-03

ID	Metabolite	Method	N_{SNP}	Beta	SE	p value	FDR	OR (95% CI)	Q	Degrees of freedom for Q	p value for Q	FDR for Q	Intercept	$SE_{Intercept}$	p value for intercept	FDR for intercept
28	XL_VLDL_CE_pct	Inverse variance weighted	123	0.017	0.034	6.07E-01	7.92E-01	1.02 (0.95-1.09)	295.344	122	2.13E-16	5.26E-16	NA	NA	NA	NA
28	XL_VLDL_CE_pct	MR Egger	123	0.148	0.051	4.65E-03	1.56E-02	1.16 (1.05-1.28)	270.948	121	1.68E-13	3.70E-13	-0.006	0.002	1.27E-03	5.45E-03
29	XL_VLDL_TG_pct	Inverse variance weighted	100	-0.083	0.034	1.61E-02	6.31E-02	0.92 (0.86-0.98)	253.235	99	1.78E-15	3.98E-15	NA	NA	NA	NA
29	XL_VLDL_TG_pct	MR Egger	100	-0.215	0.049	3.00E-05	1.41E-03	0.81 (0.73-0.89)	223.786	98	7.81E-12	1.43E-11	0.007	0.002	5.17E-04	4.05E-03
30	M_VLDL_PL_pct	Inverse variance weighted	112	0.087	0.036	1.61E-02	6.31E-02	1.09 (1.02-1.17)	367.431	111	3.38E-29	2.86E-28	NA	NA	NA	NA
30	M_VLDL_PL_pct	MR Egger	112	0.209	0.053	1.54E-04	2.39E-03	1.23 (1.11-1.37)	339.190	110	3.43E-25	4.03E-24	-0.007	0.002	3.08E-03	1.04E-02
31	M_VLDL_C_pct	Inverse variance weighted	141	0.026	0.034	4.40E-01	6.46E-01	1.03 (0.96-1.10)	433.515	140	9.49E-32	2.23E-30	NA	NA	NA	NA
31	M_VLDL_C_pct	MR Egger	141	0.188	0.049	1.62E-04	2.39E-03	1.21 (1.10-1.33)	379.862	139	3.00E-24	2.01E-23	-0.009	0.002	1.89E-05	5.66E-04
32	M_VLDL_CE_pct	Inverse variance weighted	145	0.011	0.032	7.23E-01	8.71E-01	1.01 (0.95-1.08)	408.929	144	3.24E-27	1.69E-26	NA	NA	NA	NA
32	M_VLDL_CE_pct	MR Egger	145	0.162	0.046	5.72E-04	3.81E-03	1.18 (1.07-1.29)	360.826	143	8.40E-21	3.29E-20	-0.008	0.002	2.41E-05	5.66E-04
33	M_VLDL_FC_pct	Inverse variance weighted	128	0.049	0.036	1.75E-01	3.75E-01	1.05 (0.98-1.13)	407.489	127	3.92E-31	6.15E-30	NA	NA	NA	NA
33	M_VLDL_FC_pct	MR Egger	128	0.162	0.054	3.38E-03	1.22E-02	1.18 (1.06-1.31)	384.559	126	5.86E-28	1.38E-26	-0.006	0.002	7.02E-03	1.94E-02
34	M_VLDL_TG_pct	Inverse variance weighted	136	-0.053	0.034	1.22E-01	2.78E-01	0.95 (0.89-1.01)	392.999	135	5.01E-27	2.35E-26	NA	NA	NA	NA
34	M_VLDL_TG_pct	MR Egger	136	-0.192	0.050	2.04E-04	2.39E-03	0.83 (0.75-0.91)	357.597	134	2.93E-22	1.53E-21	0.007	0.002	3.85E-04	4.05E-03
35	S_VLDL_C_pct	Inverse variance weighted	120	0.082	0.036	2.20E-02	7.94E-02	1.09 (1.01-1.16)	346.285	119	4.71E-24	1.85E-23	NA	NA	NA	NA
35	S_VLDL_C_pct	MR Egger	120	0.195	0.054	5.09E-04	3.81E-03	1.22 (1.09-1.35)	326.215	118	2.01E-21	8.61E-21	-0.006	0.002	8.08E-03	2.00E-02
36	S_VLDL_CE_pct	Inverse variance weighted	103	0.104	0.033	1.63E-03	6.20E-02	1.11 (1.04-1.18)	259.866	102	9.58E-16	2.25E-15	NA	NA	NA	NA
36	S_VLDL_CE_pct	MR Egger	103	0.153	0.050	2.95E-03	1.16E-02	1.17 (1.06-1.29)	255.647	101	2.19E-15	5.73E-15	-0.003	0.002	2.00E-01	2.93E-01
37	S_VLDL_FC_pct	Inverse variance weighted	131	0.037	0.035	2.87E-01	4.87E-01	1.04 (0.97-1.11)	399.076	130	4.04E-29	2.86E-28	NA	NA	NA	NA
37	S_VLDL_FC_pct	MR Egger	131	0.177	0.051	7.29E-04	3.81E-03	1.19 (1.08-1.32)	362.416	129	4.81E-24	2.83E-23	-0.007	0.002	4.33E-04	4.05E-03
38	S_VLDL_TG_pct	Inverse variance weighted	125	-0.087	0.034	1.09E-02	6.29E-02	0.92 (0.86-0.98)	355.717	124	3.07E-24	1.31E-23	NA	NA	NA	NA
38	S_VLDL_TG_pct	MR Egger	125	-0.191	0.051	2.64E-04	2.48E-03	0.83 (0.75-0.91)	335.653	123	1.25E-21	5.89E-21	0.006	0.002	7.66E-03	2.00E-02
39	XS_VLDL_C_pct	Inverse variance weighted	141	-0.011	0.033	7.45E-01	8.75E-01	0.99 (0.93-1.06)	415.477	140	4.26E-29	2.86E-28	NA	NA	NA	NA
39	XS_VLDL_C_pct	MR Egger	141	0.122	0.051	1.69E-02	4.69E-02	1.13 (1.02-1.25)	383.209	139	1.02E-24	9.60E-24	-0.007	0.002	8.19E-04	4.81E-03
40	XS_VLDL_CE_pct	Inverse variance weighted	148	-0.063	0.033	5.41E-02	1.41E-01	0.94 (0.88-1.00)	420.482	147	3.59E-28	2.11E-27	NA	NA	NA	NA
40	XS_VLDL_CE_pct	MR Egger	148	0.066	0.050	1.90E-01	3.08E-01	1.07 (0.97-1.18)	391.539	146	2.48E-24	1.94E-23	-0.006	0.002	1.28E-03	5.45E-03
41	XS_VLDL_TG_pct	Inverse variance weighted	145	-0.006	0.034	8.51E-01	9.09E-01	0.99 (0.93-1.06)	442.309	144	4.62E-32	2.17E-30	NA	NA	NA	NA
41	XS_VLDL_TG_pct	MR Egger	145	-0.128	0.051	1.32E-02	3.88E-02	0.88 (0.80-0.97)	414.222	143	3.32E-28	1.38E-26	0.006	0.002	2.23E-03	8.06E-03

ID	Metabolite	Method	N_{SNP}	Beta	SE	p value	FDR	OR (95% CI)	Q	Degrees of freedom for Q	p value for Q	FDR for Q	Intercept	$SE_{Intercept}$	p value for intercept	FDR for intercept
42	M_LDL_CE_pct	Inverse variance weighted	69	0.003	0.043	9.49E-01	9.63E-01	1.00 (0.92-1.09)	179.154	68	5.97E-12	1.00E-11	NA	NA	NA	NA
42	M_LDL_CE_pct	MR Egger	69	-0.138	0.063	3.26E-02	8.05E-02	0.87 (0.77-0.99)	159.138	67	1.83E-09	2.69E-09	0.008	0.003	5.00E-03	1.47E-02
43	S_LDL_PL_pct	Inverse variance weighted	83	0.014	0.036	6.95E-01	8.71E-01	1.01 (0.95-1.09)	219.371	82	1.79E-14	3.82E-14	NA	NA	NA	NA
43	S_LDL_PL_pct	MR Egger	83	0.080	0.050	1.16E-01	2.02E-01	1.08 (0.98-1.20)	210.650	81	1.73E-13	3.70E-13	-0.004	0.002	7.07E-02	1.19E-01
44	S_LDL_CE_pct	Inverse variance weighted	84	-0.032	0.030	2.90E-01	4.87E-01	0.97 (0.91-1.03)	172.758	83	2.89E-08	3.77E-08	NA	NA	NA	NA
44	S_LDL_CE_pct	MR Egger	84	-0.074	0.043	8.79E-02	1.65E-01	0.93 (0.85-1.01)	168.864	82	5.64E-08	7.36E-08	0.003	0.002	1.73E-01	2.62E-01
45	XL_HDL_FC_pct	Inverse variance weighted	142	0.079	0.029	6.45E-03	6.29E-02	1.08 (1.02-1.14)	276.457	141	7.62E-11	1.19E-10	NA	NA	NA	NA
45	XL_HDL_FC_pct	MR Egger	142	-0.011	0.044	7.97E-01	8.32E-01	0.99 (0.91-1.08)	263.313	140	1.40E-09	2.12E-09	0.004	0.002	9.14E-03	2.12E-02
46	L_HDL_FC_pct	Inverse variance weighted	120	0.002	0.035	9.63E-01	9.63E-01	1.00 (0.93-1.07)	330.024	119	9.84E-22	3.30E-21	NA	NA	NA	NA
46	L_HDL_FC_pct	MR Egger	120	0.088	0.054	1.05E-01	1.91E-01	1.09 (0.98-1.21)	318.204	118	2.65E-20	9.05E-20	-0.005	0.002	3.84E-02	6.95E-02
47	S_HDL_FC_pct	Inverse variance weighted	120	0.056	0.028	4.85E-02	1.41E-01	1.06 (1.00-1.12)	216.416	119	1.21E-07	1.50E-07	NA	NA	NA	NA
47	S_HDL_FC_pct	MR Egger	120	0.081	0.043	6.54E-02	1.34E-01	1.08 (1.00-1.18)	215.363	118	1.13E-07	1.44E-07	-0.001	0.002	4.49E-01	5.55E-01

SNP, single nucleotide polymorphism. SE, standard error. FDR, false discovery rate. OR, odds ratio. CI, confidence interval. Abbreviations are in line with **Supplementary Table 3.1**.

Supplementary Table 6.8 Results from Mendelian randomisation pleiotropy residual sum and outlier (MR-PRESSO) global test for association of metabolites with ischaemic stroke

ID	Metabolite	Observed residual sum of squares	p value	False discovery rate
1	HDL_P	300.251	<2e-04	<2e-04
2	VLDL_size	284.701	<2e-04	<2e-04
3	Omega_3	151.578	<2e-04	<2e-04
4	PUFA_pct	179.049	<2e-04	<2e-04
5	LA_pct	289.360	<2e-04	<2e-04
6	Phe	39.807	0.009	0.009
7	Glucose	109.340	<2e-04	<2e-04
8	Creatinine	178.998	0.0052	0.0065
9	Albumin	127.910	0.0004	0.000666667
10	GlycA	210.396	<2e-04	<2e-04
11	XXL_VLDL_L	270.062	<2e-04	<2e-04
12	XXL_VLDL_FC	199.350	<2e-04	<2e-04
13	XS_VLDL_C	246.015	<2e-04	<2e-04
14	XS_VLDL_CE	240.671	<2e-04	<2e-04
15	XL_HDL_L	352.060	<2e-04	<2e-04
16	XL_HDL_C	447.133	<2e-04	<2e-04
17	XL_HDL_FC	339.590	<2e-04	<2e-04
18	M_HDL_L	274.894	<2e-04	<2e-04
19	M_HDL_PL	270.899	<2e-04	<2e-04
20	S_HDL_P	188.219	<2e-04	<2e-04
21	S_HDL_L	177.423	<2e-04	<2e-04
22	S_HDL_PL	187.329	<2e-04	<2e-04
23	S_HDL_C	189.174	<2e-04	<2e-04
24	S_HDL_CE	119.921	0.0002	0.0005
25	XXL_VLDL_C_pct	97.406	0.0002	0.0005
26	XXL_VLDL_CE_pct	102.278	<2e-04	<2e-04
27	XL_VLDL_C_pct	301.278	<2e-04	<2e-04
28	XL_VLDL_CE_pct	302.131	<2e-04	<2e-04
29	XL_VLDL_TG_pct	260.256	<2e-04	<2e-04
30	M_VLDL_PL_pct	375.869	<2e-04	<2e-04

ID	Metabolite	Observed residual sum of squares	p value	False discovery rate
31	M_VLDL_C_pct	441.641	<2e-04	<2e-04
32	M_VLDL_CE_pct	416.171	<2e-04	<2e-04
33	M_VLDL_FC_pct	416.151	<2e-04	<2e-04
34	M_VLDL_TG_pct	401.265	<2e-04	<2e-04
35	S_VLDL_C_pct	355.769	<2e-04	<2e-04
36	S_VLDL_CE_pct	267.477	<2e-04	<2e-04
37	S_VLDL_FC_pct	407.441	<2e-04	<2e-04
38	S_VLDL_TG_pct	363.749	<2e-04	<2e-04
39	XS_VLDL_C_pct	422.181	<2e-04	<2e-04
40	XS_VLDL_CE_pct	425.940	<2e-04	<2e-04
41	XS_VLDL_TG_pct	449.800	<2e-04	<2e-04
42	M_LDL_CE_pct	184.458	<2e-04	<2e-04
43	S_LDL_PL_pct	225.176	<2e-04	<2e-04
44	S_LDL_CE_pct	177.343	<2e-04	<2e-04
45	XL_HDL_FC_pct	279.398	<2e-04	<2e-04
46	L_HDL_FC_pct	337.489	<2e-04	<2e-04
47	S_HDL_FC_pct	221.823	<2e-04	<2e-04

Abbreviations are in line with **Supplementary Table 3.1**.

Supplementary Table 6.9 Results from Mendelian randomisation pleiotropy residual sum and outlier (MR-PRESSO) distortion test for association of metabolites with ischaemic stroke

ID	Metabolite	Outlier indices	Coefficient	p value
1	HDL_P	81	-0.6	0.9868
1	HDL_P	84	-0.6	0.9868
2	VLDL_size	61	-6.8	0.9092
2	VLDL_size	122	-6.8	0.9092
3	Omega_3	No significant outliers	NA	NA
4	PUFA_pct	No significant outliers	NA	NA
5	LA_pct	5	-46.9	0.3148
5	LA_pct	26	-46.9	0.3148
5	LA_pct	56	-46.9	0.3148
5	LA_pct	73	-46.9	0.3148
6	Phe	16	-74.8	0.0326
6	Phe	17	-74.8	0.0326
7	Glucose	33	24.5	0.3922
8	Creatinine	No significant outliers	NA	NA
9	Albumin	72	-56.3	0.4554
10	GlycA	3	52.1	0.635
10	GlycA	26	52.1	0.635
10	GlycA	65	52.1	0.635
10	GlycA	90	52.1	0.635
11	XXL_VLDL_L	86	-11.1	0.891
11	XXL_VLDL_L	98	-11.1	0.891
12	XXL_VLDL_FC	No significant outliers	NA	NA
13	XS_VLDL_C	29	32.1	0.2594
13	XS_VLDL_C	75	32.1	0.2594
14	XS_VLDL_CE	26	26.7	0.3826
14	XS_VLDL_CE	64	26.7	0.3826
15	XL_HDL_L	5	-20.7	0.4298
15	XL_HDL_L	33	-20.7	0.4298
15	XL_HDL_L	52	-20.7	0.4298
15	XL_HDL_L	103	-20.7	0.4298
15	XL_HDL_L	117	-20.7	0.4298
16	XL_HDL_C	36	-15.5	0.6816
16	XL_HDL_C	55	-15.5	0.6816
16	XL_HDL_C	61	-15.5	0.6816
16	XL_HDL_C	87	-15.5	0.6816

16	XL_HDL_C	115	-15.5	0.6816
16	XL_HDL_C	128	-15.5	0.6816
17	XL_HDL_FC	4	2.9	0.9622
17	XL_HDL_FC	8	2.9	0.9622
17	XL_HDL_FC	79	2.9	0.9622
17	XL_HDL_FC	90	2.9	0.9622
18	M_HDL_L	41	-22.1	0.376
18	M_HDL_L	92	-22.1	0.376
19	M_HDL_PL	36	-26.2	0.3314
19	M_HDL_PL	84	-26.2	0.3314
20	S_HDL_P	60	-26.8	0.6474
20	S_HDL_P	73	-26.8	0.6474
21	S_HDL_L	62	73.8	0.6606
22	S_HDL_PL	67	130.4	0.238
23	S_HDL_C	59	-77.6	0.2574
24	S_HDL_CE	No significant outliers	NA	NA
25	XXL_VLDL_C_pct	17	9.8	0.8114
25	XXL_VLDL_C_pct	19	9.8	0.8114
26	XXL_VLDL_CE_pct	16	28.9	0.3376
27	XL_VLDL_C_pct	39	321.4	0.1396
27	XL_VLDL_C_pct	51	321.4	0.1396
28	XL_VLDL_CE_pct	45	177.2	0.2636
28	XL_VLDL_CE_pct	57	177.2	0.2636
29	XL_VLDL_TG_pct	38	-126.0	0.048
29	XL_VLDL_TG_pct	43	-126.0	0.048
30	M_VLDL_PL_pct	31	50.7	0.1022
30	M_VLDL_PL_pct	42	50.7	0.1022
30	M_VLDL_PL_pct	45	50.7	0.1022
30	M_VLDL_PL_pct	77	50.7	0.1022
31	M_VLDL_C_pct	50	5070.6	0.0094
31	M_VLDL_C_pct	60	5070.6	0.0094
31	M_VLDL_C_pct	94	5070.6	0.0094
32	M_VLDL_CE_pct	49	176.9	0.2288
32	M_VLDL_CE_pct	60	176.9	0.2288
32	M_VLDL_CE_pct	102	176.9	0.2288
33	M_VLDL_FC_pct	36	201.9	0.0888
33	M_VLDL_FC_pct	49	201.9	0.0888
33	M_VLDL_FC_pct	84	201.9	0.0888

34	M_VLDL_TG_pct	43	-129.2	0.1268
34	M_VLDL_TG_pct	54	-129.2	0.1268
34	M_VLDL_TG_pct	83	-129.2	0.1268
35	S_VLDL_C_pct	36	79.5	0.1046
35	S_VLDL_C_pct	53	79.5	0.1046
35	S_VLDL_C_pct	80	79.5	0.1046
35	S_VLDL_C_pct	87	79.5	0.1046
36	S_VLDL_CE_pct	5	57.5	0.0302
36	S_VLDL_CE_pct	24	57.5	0.0302
36	S_VLDL_CE_pct	38	57.5	0.0302
36	S_VLDL_CE_pct	39	57.5	0.0302
37	S_VLDL_FC_pct	41	204.7	0.167
37	S_VLDL_FC_pct	54	204.7	0.167
37	S_VLDL_FC_pct	88	204.7	0.167
38	S_VLDL_TG_pct	32	-48.6	0.1106
38	S_VLDL_TG_pct	50	-48.6	0.1106
38	S_VLDL_TG_pct	60	-48.6	0.1106
38	S_VLDL_TG_pct	87	-48.6	0.1106
38	S_VLDL_TG_pct	92	-48.6	0.1106
39	XS_VLDL_C_pct	1	63.8	0.664
39	XS_VLDL_C_pct	75	63.8	0.664
39	XS_VLDL_C_pct	89	63.8	0.664
39	XS_VLDL_C_pct	104	63.8	0.664
40	XS_VLDL_CE_pct	76	-4.4	0.9234
40	XS_VLDL_CE_pct	91	-4.4	0.9234
40	XS_VLDL_CE_pct	107	-4.4	0.9234
41	XS_VLDL_TG_pct	2	-140.6	0.2312
41	XS_VLDL_TG_pct	96	-140.6	0.2312
41	XS_VLDL_TG_pct	102	-140.6	0.2312
42	M_LDL_CE_pct	45	120.6	0.1476
42	M_LDL_CE_pct	65	120.6	0.1476
43	S_LDL_PL_pct	35	-58.5	0.7038
43	S_LDL_PL_pct	62	-58.5	0.7038
44	S_LDL_CE_pct	33	12.7	0.887
45	XL_HDL_FC_pct	90	14.1	0.6336
45	XL_HDL_FC_pct	119	14.1	0.6336
46	L_HDL_FC_pct	79	110.4	0.221
46	L_HDL_FC_pct	80	110.4	0.221

46	L_HDL_FC_pct	90	110.4	0.221
46	L_HDL_FC_pct	103	110.4	0.221
46	L_HDL_FC_pct	105	110.4	0.221
47	S_HDL_FC_pct	97	66.2	0.1698

Supplementary Table 6.10 Mendelian randomisation pleiotropy residual sum and outlier (MR-PRESSO) outlier-corrected results for association of metabolites with ischaemic stroke

ID	Metabolite	Method	Beta	SE	t statistic	p value	FDR	OR (95% CI)
1	HDL_P	Raw	-0.076	0.039	-1.980	0.050	0.146	0.93 (0.86, 1)
1	HDL_P	Outlier-corrected	-0.076	0.031	-2.463	0.015	0.092	0.93 (0.87, 0.98)
2	VLDL_size	Raw	0.042	0.027	1.537	0.127	0.283	1.04 (0.99, 1.1)
2	VLDL_size	Outlier-corrected	0.045	0.024	1.860	0.065	0.171	1.05 (1, 1.1)
3	Omega_3	Raw	0.016	0.031	0.515	0.608	0.794	1.02 (0.96, 1.08)
3	Omega_3	Outlier-corrected	NA	NA	NA	NA	NA	NA
4	PUFA_pct	Raw	-0.014	0.039	-0.371	0.712	0.872	0.99 (0.91, 1.06)
4	PUFA_pct	Outlier-corrected	NA	NA	NA	NA	NA	NA
5	LA_pct	Raw	-0.076	0.060	-1.278	0.205	0.418	0.93 (0.82, 1.04)
5	LA_pct	Outlier-corrected	-0.052	0.044	-1.177	0.242	0.407	0.95 (0.87, 1.04)
6	Phe	Raw	-0.126	0.075	-1.682	0.112	0.277	0.88 (0.76, 1.02)
6	Phe	Outlier-corrected	-0.072	0.052	-1.370	0.192	0.404	0.93 (0.84, 1.03)
7	Glucose	Raw	0.176	0.060	2.911	0.005	0.070	1.19 (1.06, 1.34)
7	Glucose	Outlier-corrected	0.141	0.050	2.809	0.007	0.092	1.15 (1.04, 1.27)
8	Creatinine	Raw	-0.007	0.036	-0.204	0.838	0.910	0.99 (0.92, 1.07)
8	Creatinine	Outlier-corrected	NA	NA	NA	NA	NA	NA
9	Albumin	Raw	-0.039	0.045	-0.866	0.389	0.605	0.96 (0.88, 1.05)
9	Albumin	Outlier-corrected	-0.025	0.041	-0.595	0.553	0.680	0.98 (0.9, 1.06)
10	GlycA	Raw	0.021	0.036	0.597	0.552	0.763	1.02 (0.95, 1.1)
10	GlycA	Outlier-corrected	0.014	0.032	0.446	0.657	0.699	1.01 (0.95, 1.08)
11	XXL_VLDL_L	Raw	0.038	0.031	1.199	0.233	0.437	1.04 (0.98, 1.1)
11	XXL_VLDL_L	Outlier-corrected	0.042	0.027	1.554	0.123	0.288	1.04 (0.99, 1.1)
12	XXL_VLDL_FC	Raw	0.006	0.025	0.230	0.819	0.910	1.01 (0.96, 1.06)
12	XXL_VLDL_FC	Outlier-corrected	NA	NA	NA	NA	NA	NA
13	XS_VLDL_C	Raw	0.080	0.033	2.432	0.017	0.070	1.08 (1.02, 1.15)
13	XS_VLDL_C	Outlier-corrected	0.060	0.025	2.368	0.020	0.094	1.06 (1.01, 1.12)
14	XS_VLDL_CE	Raw	0.086	0.034	2.511	0.014	0.070	1.09 (1.02, 1.17)
14	XS_VLDL_CE	Outlier-corrected	0.068	0.027	2.568	0.012	0.092	1.07 (1.02, 1.13)
15	XL_HDL_L	Raw	-0.088	0.033	-2.685	0.008	0.070	0.92 (0.86, 0.98)
15	XL_HDL_L	Outlier-corrected	-0.073	0.028	-2.650	0.009	0.092	0.93 (0.88, 0.98)

ID	Metabolite	Method	Beta	SE	t statistic	p value	FDR	OR (95% CI)
16	XL_HDL_C	Raw	-0.065	0.033	-1.945	0.054	0.146	0.94 (0.88, 1)
16	XL_HDL_C	Outlier-corrected	-0.056	0.026	-2.145	0.034	0.108	0.95 (0.9, 1)
17	XL_HDL_FC	Raw	-0.041	0.033	-1.236	0.219	0.428	0.96 (0.9, 1.02)
17	XL_HDL_FC	Outlier-corrected	-0.042	0.028	-1.521	0.131	0.289	0.96 (0.91, 1.01)
18	M_HDL_L	Raw	-0.094	0.033	-2.882	0.005	0.070	0.91 (0.85, 0.97)
18	M_HDL_L	Outlier-corrected	-0.077	0.027	-2.863	0.005	0.092	0.93 (0.88, 0.98)
19	M_HDL_PL	Raw	-0.085	0.033	-2.566	0.012	0.070	0.92 (0.86, 0.98)
19	M_HDL_PL	Outlier-corrected	-0.067	0.027	-2.478	0.015	0.092	0.94 (0.89, 0.99)
20	S_HDL_P	Raw	-0.048	0.041	-1.174	0.244	0.441	0.95 (0.88, 1.03)
20	S_HDL_P	Outlier-corrected	-0.038	0.032	-1.200	0.234	0.407	0.96 (0.9, 1.02)
21	S_HDL_L	Raw	-0.009	0.036	-0.243	0.809	0.910	0.99 (0.92, 1.06)
21	S_HDL_L	Outlier-corrected	-0.033	0.030	-1.086	0.281	0.421	0.97 (0.91, 1.03)
22	S_HDL_PL	Raw	0.005	0.035	0.144	0.886	0.925	1.01 (0.94, 1.08)
22	S_HDL_PL	Outlier-corrected	-0.017	0.031	-0.549	0.584	0.680	0.98 (0.93, 1.04)
23	S_HDL_C	Raw	-0.039	0.040	-0.988	0.326	0.528	0.96 (0.89, 1.04)
23	S_HDL_C	Outlier-corrected	-0.022	0.033	-0.677	0.500	0.678	0.98 (0.92, 1.04)
24	S_HDL_CE	Raw	-0.024	0.033	-0.729	0.469	0.667	0.98 (0.92, 1.04)
24	S_HDL_CE	Outlier-corrected	NA	NA	NA	NA	NA	NA
25	XXL_VLDL_C_pct	Raw	0.096	0.047	2.023	0.050	0.146	1.1 (1, 1.21)
25	XXL_VLDL_C_pct	Outlier-corrected	0.088	0.039	2.240	0.032	0.108	1.09 (1.01, 1.18)
26	XXL_VLDL_CE_pct	Raw	0.134	0.052	2.570	0.014	0.070	1.14 (1.03, 1.27)
26	XXL_VLDL_CE_pct	Outlier-corrected	0.104	0.045	2.293	0.028	0.108	1.11 (1.02, 1.21)
27	XL_VLDL_C_pct	Raw	0.028	0.034	0.847	0.399	0.605	1.03 (0.96, 1.1)
27	XL_VLDL_C_pct	Outlier-corrected	-0.013	0.030	-0.434	0.665	0.699	0.99 (0.93, 1.05)
28	XL_VLDL_CE_pct	Raw	0.017	0.034	0.515	0.608	0.794	1.02 (0.95, 1.09)
28	XL_VLDL_CE_pct	Outlier-corrected	-0.023	0.030	-0.748	0.456	0.638	0.98 (0.92, 1.04)
29	XL_VLDL_TG_pct	Raw	-0.083	0.034	-2.406	0.018	0.070	0.92 (0.86, 0.98)
29	XL_VLDL_TG_pct	Outlier-corrected	-0.037	0.031	-1.195	0.235	0.407	0.96 (0.91, 1.02)
30	M_VLDL_PL_pct	Raw	0.087	0.036	2.407	0.018	0.070	1.09 (1.02, 1.17)
30	M_VLDL_PL_pct	Outlier-corrected	0.058	0.029	2.005	0.047	0.133	1.06 (1, 1.12)
31	M_VLDL_C_pct	Raw	0.026	0.034	0.772	0.441	0.648	1.03 (0.96, 1.1)
31	M_VLDL_C_pct	Outlier-corrected	-0.001	0.028	-0.019	0.985	0.985	1 (0.95, 1.06)

ID	Metabolite	Method	Beta	SE	t statistic	p value	FDR	OR (95% CI)
32	M_VLDL_CE_pct	Raw	0.011	0.032	0.355	0.723	0.872	1.01 (0.95, 1.08)
32	M_VLDL_CE_pct	Outlier-corrected	-0.015	0.027	-0.555	0.579	0.680	0.99 (0.93, 1.04)
33	M_VLDL_FC_pct	Raw	0.049	0.036	1.355	0.178	0.380	1.05 (0.98, 1.13)
33	M_VLDL_FC_pct	Outlier-corrected	0.016	0.030	0.543	0.588	0.680	1.02 (0.96, 1.08)
34	M_VLDL_TG_pct	Raw	-0.053	0.034	-1.546	0.124	0.283	0.95 (0.89, 1.01)
34	M_VLDL_TG_pct	Outlier-corrected	-0.023	0.027	-0.834	0.406	0.587	0.98 (0.93, 1.03)
35	S_VLDL_C_pct	Raw	0.082	0.036	2.291	0.024	0.086	1.09 (1.01, 1.16)
35	S_VLDL_C_pct	Outlier-corrected	0.046	0.029	1.551	0.124	0.288	1.05 (0.99, 1.11)
36	S_VLDL_CE_pct	Raw	0.104	0.033	3.150	0.002	0.070	1.11 (1.04, 1.18)
36	S_VLDL_CE_pct	Outlier-corrected	0.066	0.028	2.361	0.020	0.094	1.07 (1.01, 1.13)
37	S_VLDL_FC_pct	Raw	0.037	0.035	1.064	0.289	0.492	1.04 (0.97, 1.11)
37	S_VLDL_FC_pct	Outlier-corrected	0.012	0.028	0.433	0.666	0.699	1.01 (0.96, 1.07)
38	S_VLDL_TG_pct	Raw	-0.087	0.034	-2.545	0.012	0.070	0.92 (0.86, 0.98)
38	S_VLDL_TG_pct	Outlier-corrected	-0.058	0.028	-2.115	0.037	0.110	0.94 (0.89, 1)
39	XS_VLDL_C_pct	Raw	-0.011	0.033	-0.325	0.746	0.876	0.99 (0.93, 1.06)
39	XS_VLDL_C_pct	Outlier-corrected	-0.030	0.027	-1.104	0.271	0.421	0.97 (0.92, 1.02)
40	XS_VLDL_CE_pct	Raw	-0.063	0.033	-1.926	0.056	0.146	0.94 (0.88, 1)
40	XS_VLDL_CE_pct	Outlier-corrected	-0.060	0.028	-2.165	0.032	0.108	0.94 (0.89, 0.99)
41	XS_VLDL_TG_pct	Raw	-0.006	0.034	-0.187	0.852	0.910	0.99 (0.93, 1.06)
41	XS_VLDL_TG_pct	Outlier-corrected	0.016	0.028	0.556	0.579	0.680	1.02 (0.96, 1.07)
42	M_LDL_CE_pct	Raw	0.003	0.043	0.064	0.949	0.963	1 (0.92, 1.09)
42	M_LDL_CE_pct	Outlier-corrected	-0.013	0.036	-0.366	0.715	0.733	0.99 (0.92, 1.06)
43	S_LDL_PL_pct	Raw	0.014	0.036	0.392	0.696	0.872	1.01 (0.95, 1.09)
43	S_LDL_PL_pct	Outlier-corrected	0.034	0.031	1.096	0.276	0.421	1.03 (0.97, 1.1)
44	S_LDL_CE_pct	Raw	-0.032	0.030	-1.058	0.293	0.492	0.97 (0.91, 1.03)
44	S_LDL_CE_pct	Outlier-corrected	-0.037	0.029	-1.275	0.206	0.407	0.96 (0.91, 1.02)
45	XL_HDL_FC_pct	Raw	0.079	0.029	2.724	0.007	0.070	1.08 (1.02, 1.14)
45	XL_HDL_FC_pct	Outlier-corrected	0.069	0.026	2.632	0.009	0.092	1.07 (1.02, 1.13)
46	L_HDL_FC_pct	Raw	0.002	0.035	0.046	0.963	0.963	1 (0.93, 1.07)
46	L_HDL_FC_pct	Outlier-corrected	-0.016	0.030	-0.528	0.599	0.680	0.98 (0.93, 1.04)
47	S_HDL_FC_pct	Raw	0.056	0.028	1.973	0.051	0.146	1.06 (1, 1.12)
47	S_HDL_FC_pct	Outlier-corrected	0.034	0.027	1.224	0.223	0.407	1.03 (0.98, 1.09)

SD, standard deviation. FDR, false discovery rate. OR, odds ratio. CI, confidence interval. Abbreviations are in line with **Supplementary Table 3.1**.

Supplementary Table 6.11 Results of posterior inclusion probabilities estimated via the iterative Sum of Single Effects (SuSiE) algorithm

ID	Metabolite	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
1	HDL_P	0.2503	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
2	VLDL_size	0.0002	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
3	Omega_3	0.0004	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
4	PUFA_pct	0.0002	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
5	LA_pct	0.0002	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
6	Phe	0.0002	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
7	Glucose	0.0001	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
8	Creatinine	0.0001	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
9	Albumin	0.0003	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
10	GlycA	0.0001	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
11	XXL_VLDL_L	0.0001	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
12	XXL_VLDL_FC	0.0002	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
13	XS_VLDL_C	0.0003	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
14	XS_VLDL_CE	0.0002	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
15	XL_HDL_L	0.0021	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
16	XL_HDL_C	0.0018	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
17	XL_HDL_FC	0.0005	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
18	M_HDL_L	0.3614	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
19	M_HDL_PL	0.2593	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
20	S_HDL_P	0.0002	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
21	S_HDL_L	0.0003	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
22	S_HDL_PL	0.0006	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
23	S_HDL_C	0.0004	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213

ID	Metabolite	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
24	S_HDL_CE	0.0003	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
25	XXL_VLDL_C_pct	0.0098	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
26	XXL_VLDL_CE_pct	0.0682	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
27	XL_VLDL_C_pct	0.0003	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
28	XL_VLDL_CE_pct	0.0003	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
29	XL_VLDL_TG_pct	0.0006	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
30	M_VLDL_PL_pct	0.0003	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
31	M_VLDL_C_pct	0.0001	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
32	M_VLDL_CE_pct	0.0001	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
33	M_VLDL_FC_pct	0.0002	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
34	M_VLDL_TG_pct	0.0002	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
35	S_VLDL_C_pct	0.0005	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
36	S_VLDL_CE_pct	0.0077	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
37	S_VLDL_FC_pct	0.0001	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
38	S_VLDL_TG_pct	0.0002	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
39	XS_VLDL_C_pct	0.0002	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
40	XS_VLDL_CE_pct	0.0002	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
41	XS_VLDL_TG_pct	0.0001	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
42	M_LDL_CE_pct	0.0002	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
43	S_LDL_PL_pct	0.0002	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
44	S_LDL_CE_pct	0.0002	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
45	XL_HDL_FC_pct	0.0299	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
46	L_HDL_FC_pct	0.0003	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213
47	S_HDL_FC_pct	0.0002	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213	0.0213

Abbreviations are in line with **Supplementary Table 3.1**.

Supplementary Table 6.12 The results from multivariable Mendelian randomisation using inverse-variance weighted estimation

Metabolite	Beta	Standard error	p value

XXL_VLDL_CE_pct	Cholesteryl esters to total lipids ratio in chylomicrons and extremely large VLDL	0.063	0.015	3.42E-05
M_HDL_PL	Phospholipids in medium HDL	0.084	0.129	5.16E-01
HDL_P	Concentration of HDL particles	0.020	0.047	6.71E-01
M_HDL_L	Total lipids in medium HDL	-0.102	0.167	5.44E-01
XL_HDL_FC_pct	Free cholesterol to total lipids ratio in very large HDL	0.026	0.017	1.22E-01

Abbreviations are in line with **Supplementary Table 3.1**.

Supplementary Table 6.13 All results from reverse Mendelian randomisation analyses of association of ischaemic stroke with metabolites

ID	Metabolite	Method	N_{SNP}	Beta	SE	p value	FDR	OR (95% CI)
1	HDL_P	Inverse variance weighted	23	0.016	0.037	0.662	0.889	1.02 (0.94-1.09)
1	HDL_P	MR Egger	23	-0.007	0.185	0.972	0.979	0.99 (0.69-1.43)
1	HDL_P	Simple mode	23	0.036	0.033	0.289	0.902	1.04 (0.97-1.11)
1	HDL_P	Weighted median	23	0.038	0.020	0.056	0.658	1.04 (1.00-1.08)
1	HDL_P	Weighted mode	23	0.043	0.030	0.165	0.918	1.04 (0.98-1.11)
2	VLDL_size	Inverse variance weighted	23	-0.043	0.024	0.070	0.889	0.96 (0.91-1.00)
2	VLDL_size	MR Egger	23	-0.003	0.118	0.979	0.979	1.00 (0.79-1.26)
2	VLDL_size	Simple mode	23	0.031	0.035	0.385	0.902	1.03 (0.96-1.10)
2	VLDL_size	Weighted median	23	-0.015	0.020	0.442	0.832	0.98 (0.95-1.02)
2	VLDL_size	Weighted mode	23	0.025	0.028	0.395	0.918	1.02 (0.97-1.08)
3	Omega_3	Inverse variance weighted	23	0.025	0.034	0.451	0.889	1.03 (0.96-1.10)
3	Omega_3	MR Egger	23	0.125	0.165	0.459	0.929	1.13 (0.82-1.57)
3	Omega_3	Simple mode	23	-0.020	0.045	0.651	0.902	0.98 (0.90-1.07)
3	Omega_3	Weighted median	23	-0.014	0.021	0.504	0.832	0.99 (0.95-1.03)
3	Omega_3	Weighted mode	23	-0.034	0.042	0.431	0.918	0.97 (0.89-1.05)
4	PUFA_pct	Inverse variance weighted	23	0.012	0.025	0.637	0.889	1.01 (0.96-1.06)
4	PUFA_pct	MR Egger	23	-0.056	0.124	0.658	0.929	0.95 (0.74-1.21)
4	PUFA_pct	Simple mode	23	-0.010	0.046	0.823	0.962	0.99 (0.90-1.08)
4	PUFA_pct	Weighted median	23	-0.003	0.021	0.889	0.954	1.00 (0.96-1.04)
4	PUFA_pct	Weighted mode	23	-0.006	0.037	0.872	0.997	0.99 (0.92-1.07)
5	LA_pct	Inverse variance weighted	23	0.009	0.039	0.812	0.948	1.01 (0.94-1.09)
5	LA_pct	MR Egger	23	0.064	0.192	0.743	0.929	1.07 (0.73-1.55)
5	LA_pct	Simple mode	23	0.000	0.032	0.993	0.993	1.00 (0.94-1.06)
5	LA_pct	Weighted median	23	-0.009	0.020	0.649	0.832	0.99 (0.95-1.03)
5	LA_pct	Weighted mode	23	-0.004	0.031	0.903	0.997	1.00 (0.94-1.06)
6	Phe	Inverse variance weighted	23	0.017	0.016	0.293	0.889	1.02 (0.99-1.05)
6	Phe	MR Egger	23	0.010	0.082	0.903	0.964	1.01 (0.86-1.19)
6	Phe	Simple mode	23	-0.034	0.045	0.455	0.902	0.97 (0.88-1.06)
6	Phe	Weighted median	23	0.015	0.020	0.462	0.832	1.02 (0.98-1.06)
6	Phe	Weighted mode	23	-0.023	0.043	0.601	0.918	0.98 (0.90-1.06)
7	Glucose	Inverse variance weighted	23	-0.017	0.038	0.655	0.889	0.98 (0.91-1.06)
7	Glucose	MR Egger	23	-0.076	0.189	0.694	0.929	0.93 (0.64-1.34)
7	Glucose	Simple mode	23	0.052	0.040	0.215	0.902	1.05 (0.97-1.14)
7	Glucose	Weighted median	23	0.027	0.022	0.220	0.776	1.03 (0.98-1.07)
7	Glucose	Weighted mode	23	0.049	0.034	0.167	0.918	1.05 (0.98-1.12)
8	Creatinine	Inverse variance weighted	23	-0.006	0.018	0.719	0.925	0.99 (0.96-1.03)
8	Creatinine	MR Egger	23	-0.175	0.081	0.042	0.929	0.84 (0.72-0.98)
8	Creatinine	Simple mode	23	-0.002	0.039	0.966	0.993	1.00 (0.92-1.08)
8	Creatinine	Weighted median	23	-0.011	0.018	0.531	0.832	0.99 (0.96-1.02)
8	Creatinine	Weighted mode	23	-0.002	0.033	0.959	0.997	1.00 (0.94-1.06)
9	Albumin	Inverse variance weighted	23	-0.052	0.030	0.080	0.889	0.95 (0.90-1.01)
9	Albumin	MR Egger	23	-0.150	0.146	0.316	0.929	0.86 (0.65-1.15)
9	Albumin	Simple mode	23	-0.019	0.049	0.700	0.940	0.98 (0.89-1.08)
9	Albumin	Weighted median	23	-0.031	0.024	0.194	0.761	0.97 (0.93-1.02)
9	Albumin	Weighted mode	23	-0.021	0.049	0.675	0.961	0.98 (0.89-1.08)
10	GlycA	Inverse variance weighted	23	0.005	0.031	0.876	0.977	1.00 (0.95-1.07)

ID	Metabolite	Method	N_{SNP}	Beta	SE	p value	FDR	OR (95% CI)
10	GlycA	MR Egger	23	0.027	0.155	0.862	0.942	1.03 (0.76-1.39)
10	GlycA	Simple mode	23	0.050	0.058	0.402	0.902	1.05 (0.94-1.18)
10	GlycA	Weighted median	23	0.003	0.024	0.893	0.954	1.00 (0.96-1.05)
10	GlycA	Weighted mode	23	0.039	0.055	0.488	0.918	1.04 (0.93-1.16)
11	XXL_VLDL_L	Inverse variance weighted	23	-0.035	0.025	0.174	0.889	0.97 (0.92-1.02)
11	XXL_VLDL_L	MR Egger	23	0.056	0.125	0.660	0.929	1.06 (0.83-1.35)
11	XXL_VLDL_L	Simple mode	23	0.025	0.035	0.479	0.902	1.03 (0.96-1.10)
11	XXL_VLDL_L	Weighted median	23	-0.011	0.020	0.585	0.832	0.99 (0.95-1.03)
11	XXL_VLDL_L	Weighted mode	23	0.025	0.033	0.456	0.918	1.03 (0.96-1.09)
12	XXL_VLDL_FC	Inverse variance weighted	23	-0.023	0.021	0.286	0.889	0.98 (0.94-1.02)
12	XXL_VLDL_FC	MR Egger	23	0.072	0.103	0.494	0.929	1.07 (0.88-1.32)
12	XXL_VLDL_FC	Simple mode	23	0.030	0.037	0.421	0.902	1.03 (0.96-1.11)
12	XXL_VLDL_FC	Weighted median	23	-0.005	0.020	0.805	0.929	0.99 (0.96-1.04)
12	XXL_VLDL_FC	Weighted mode	23	0.030	0.037	0.423	0.918	1.03 (0.96-1.11)
13	XS_VLDL_C	Inverse variance weighted	23	0.081	0.105	0.440	0.889	1.08 (0.88-1.33)
13	XS_VLDL_C	MR Egger	23	0.463	0.514	0.378	0.929	1.59 (0.58-4.35)
13	XS_VLDL_C	Simple mode	23	0.019	0.034	0.588	0.902	1.02 (0.95-1.09)
13	XS_VLDL_C	Weighted median	23	-0.009	0.022	0.675	0.835	0.99 (0.95-1.03)
13	XS_VLDL_C	Weighted mode	23	0.001	0.033	0.965	0.997	1.00 (0.94-1.07)
14	XS_VLDL_CE	Inverse variance weighted	23	0.079	0.106	0.453	0.889	1.08 (0.88-1.33)
14	XS_VLDL_CE	MR Egger	23	0.447	0.519	0.399	0.929	1.56 (0.57-4.32)
14	XS_VLDL_CE	Simple mode	23	0.017	0.036	0.633	0.902	1.02 (0.95-1.09)
14	XS_VLDL_CE	Weighted median	23	-0.013	0.022	0.543	0.832	0.99 (0.95-1.03)
14	XS_VLDL_CE	Weighted mode	23	-0.001	0.034	0.988	0.997	1.00 (0.93-1.07)
15	XL_HDL_L	Inverse variance weighted	23	-0.005	0.045	0.915	0.977	1.00 (0.91-1.09)
15	XL_HDL_L	MR Egger	23	0.041	0.224	0.856	0.942	1.04 (0.67-1.62)
15	XL_HDL_L	Simple mode	23	0.001	0.031	0.975	0.993	1.00 (0.94-1.06)
15	XL_HDL_L	Weighted median	23	0.002	0.019	0.931	0.959	1.00 (0.97-1.04)
15	XL_HDL_L	Weighted mode	23	0.001	0.030	0.974	0.997	1.00 (0.94-1.06)
16	XL_HDL_C	Inverse variance weighted	23	-0.001	0.050	0.989	0.989	1.00 (0.91-1.10)
16	XL_HDL_C	MR Egger	23	0.060	0.247	0.809	0.942	1.06 (0.65-1.72)
16	XL_HDL_C	Simple mode	23	0.017	0.031	0.596	0.902	1.02 (0.96-1.08)
16	XL_HDL_C	Weighted median	23	0.011	0.019	0.554	0.832	1.01 (0.97-1.05)
16	XL_HDL_C	Weighted mode	23	0.017	0.029	0.568	0.918	1.02 (0.96-1.08)
17	XL_HDL_FC	Inverse variance weighted	23	-0.002	0.059	0.970	0.989	1.00 (0.89-1.12)
17	XL_HDL_FC	MR Egger	23	0.107	0.290	0.715	0.929	1.11 (0.63-1.97)
17	XL_HDL_FC	Simple mode	23	-0.008	0.033	0.818	0.962	0.99 (0.93-1.06)
17	XL_HDL_FC	Weighted median	23	-0.001	0.021	0.963	0.963	1.00 (0.96-1.04)
17	XL_HDL_FC	Weighted mode	23	0.000	0.032	0.996	0.997	1.00 (0.94-1.07)
18	M_HDL_L	Inverse variance weighted	23	-0.008	0.035	0.827	0.948	0.99 (0.93-1.06)
18	M_HDL_L	MR Egger	23	-0.128	0.172	0.464	0.929	0.88 (0.63-1.23)
18	M_HDL_L	Simple mode	23	0.053	0.031	0.107	0.902	1.05 (0.99-1.12)
18	M_HDL_L	Weighted median	23	0.040	0.020	0.046	0.658	1.04 (1.00-1.08)
18	M_HDL_L	Weighted mode	23	0.055	0.028	0.066	0.918	1.06 (1.00-1.12)
19	M_HDL_PL	Inverse variance weighted	23	-0.012	0.037	0.748	0.925	0.99 (0.92-1.06)
19	M_HDL_PL	MR Egger	23	-0.136	0.181	0.459	0.929	0.87 (0.61-1.24)
19	M_HDL_PL	Simple mode	23	0.054	0.033	0.111	0.902	1.06 (0.99-1.13)
19	M_HDL_PL	Weighted median	23	0.048	0.020	0.019	0.658	1.05 (1.01-1.09)
19	M_HDL_PL	Weighted mode	23	0.054	0.030	0.085	0.918	1.06 (1.00-1.12)
20	S_HDL_P	Inverse variance weighted	23	0.037	0.052	0.479	0.889	1.04 (0.94-1.15)

ID	Metabolite	Method	N_{SNP}	Beta	SE	p value	FDR	OR (95% CI)
20	S_HDL_P	MR Egger	23	0.106	0.259	0.686	0.929	1.11 (0.67-1.85)
20	S_HDL_P	Simple mode	23	0.026	0.033	0.439	0.902	1.03 (0.96-1.09)
20	S_HDL_P	Weighted median	23	0.021	0.022	0.339	0.797	1.02 (0.98-1.07)
20	S_HDL_P	Weighted mode	23	0.023	0.031	0.475	0.918	1.02 (0.96-1.09)
21	S_HDL_L	Inverse variance weighted	23	0.023	0.046	0.611	0.889	1.02 (0.94-1.12)
21	S_HDL_L	MR Egger	23	0.067	0.227	0.771	0.929	1.07 (0.68-1.67)
21	S_HDL_L	Simple mode	23	0.024	0.035	0.512	0.902	1.02 (0.96-1.10)
21	S_HDL_L	Weighted median	23	0.026	0.022	0.237	0.776	1.03 (0.98-1.07)
21	S_HDL_L	Weighted mode	23	0.018	0.034	0.601	0.918	1.02 (0.95-1.09)
22	S_HDL_PL	Inverse variance weighted	23	0.015	0.044	0.738	0.925	1.01 (0.93-1.11)
22	S_HDL_PL	MR Egger	23	0.021	0.220	0.923	0.965	1.02 (0.66-1.57)
22	S_HDL_PL	Simple mode	23	0.025	0.035	0.481	0.902	1.03 (0.96-1.10)
22	S_HDL_PL	Weighted median	23	0.030	0.022	0.175	0.761	1.03 (0.99-1.08)
22	S_HDL_PL	Weighted mode	23	0.019	0.037	0.606	0.918	1.02 (0.95-1.10)
23	S_HDL_C	Inverse variance weighted	23	0.034	0.051	0.509	0.889	1.03 (0.94-1.14)
23	S_HDL_C	MR Egger	23	0.079	0.253	0.758	0.929	1.08 (0.66-1.77)
23	S_HDL_C	Simple mode	23	0.032	0.034	0.348	0.902	1.03 (0.97-1.10)
23	S_HDL_C	Weighted median	23	0.030	0.021	0.160	0.761	1.03 (0.99-1.07)
23	S_HDL_C	Weighted mode	23	0.032	0.030	0.294	0.918	1.03 (0.97-1.10)
24	S_HDL_CE	Inverse variance weighted	23	0.030	0.051	0.548	0.889	1.03 (0.93-1.14)
24	S_HDL_CE	MR Egger	23	0.048	0.251	0.851	0.942	1.05 (0.64-1.72)
24	S_HDL_CE	Simple mode	23	0.020	0.035	0.572	0.902	1.02 (0.95-1.09)
24	S_HDL_CE	Weighted median	23	0.007	0.021	0.729	0.878	1.01 (0.97-1.05)
24	S_HDL_CE	Weighted mode	23	0.020	0.031	0.530	0.918	1.02 (0.96-1.08)
25	XXL_VLDL_C_pct	Inverse variance weighted	23	0.092	0.049	0.063	0.889	1.10 (1.00-1.21)
25	XXL_VLDL_C_pct	MR Egger	23	0.175	0.243	0.480	0.929	1.19 (0.74-1.92)
25	XXL_VLDL_C_pct	Simple mode	23	0.007	0.031	0.832	0.962	1.01 (0.95-1.07)
25	XXL_VLDL_C_pct	Weighted median	23	0.016	0.019	0.413	0.832	1.02 (0.98-1.05)
25	XXL_VLDL_C_pct	Weighted mode	23	0.007	0.026	0.804	0.997	1.01 (0.96-1.06)
26	XXL_VLDL_CE_pct	Inverse variance weighted	23	0.094	0.053	0.077	0.889	1.10 (0.99-1.22)
26	XXL_VLDL_CE_pct	MR Egger	23	0.253	0.262	0.345	0.929	1.29 (0.77-2.15)
26	XXL_VLDL_CE_pct	Simple mode	23	-0.008	0.034	0.812	0.962	0.99 (0.93-1.06)
26	XXL_VLDL_CE_pct	Weighted median	23	0.020	0.020	0.317	0.783	1.02 (0.98-1.06)
26	XXL_VLDL_CE_pct	Weighted mode	23	-0.006	0.032	0.858	0.997	0.99 (0.93-1.06)
27	XL_VLDL_C_pct	Inverse variance weighted	23	0.078	0.070	0.263	0.889	1.08 (0.94-1.24)
27	XL_VLDL_C_pct	MR Egger	23	0.280	0.342	0.423	0.929	1.32 (0.68-2.59)
27	XL_VLDL_C_pct	Simple mode	23	0.017	0.029	0.571	0.902	1.02 (0.96-1.08)
27	XL_VLDL_C_pct	Weighted median	23	0.010	0.020	0.625	0.832	1.01 (0.97-1.05)
27	XL_VLDL_C_pct	Weighted mode	23	0.020	0.024	0.410	0.918	1.02 (0.97-1.07)
28	XL_VLDL_CE_pct	Inverse variance weighted	23	0.078	0.066	0.237	0.889	1.08 (0.95-1.23)
28	XL_VLDL_CE_pct	MR Egger	23	0.266	0.325	0.421	0.929	1.31 (0.69-2.47)
28	XL_VLDL_CE_pct	Simple mode	23	0.018	0.030	0.554	0.902	1.02 (0.96-1.08)
28	XL_VLDL_CE_pct	Weighted median	23	0.009	0.019	0.655	0.832	1.01 (0.97-1.05)
28	XL_VLDL_CE_pct	Weighted mode	23	0.018	0.026	0.488	0.918	1.02 (0.97-1.07)
29	XL_VLDL_TG_pct	Inverse variance weighted	23	-0.086	0.083	0.296	0.889	0.92 (0.78-1.08)
29	XL_VLDL_TG_pct	MR Egger	23	-0.350	0.406	0.397	0.929	0.70 (0.32-1.56)
29	XL_VLDL_TG_pct	Simple mode	23	0.019	0.029	0.515	0.902	1.02 (0.96-1.08)
29	XL_VLDL_TG_pct	Weighted median	23	0.010	0.020	0.619	0.832	1.01 (0.97-1.05)
29	XL_VLDL_TG_pct	Weighted mode	23	0.008	0.026	0.778	0.997	1.01 (0.96-1.06)
30	M_VLDL_PL_pct	Inverse variance weighted	23	0.088	0.102	0.391	0.889	1.09 (0.89-1.33)

ID	Metabolite	Method	N_{SNP}	Beta	SE	p value	FDR	OR (95% CI)
30	M_VLDL_PL_pct	MR Egger	23	0.434	0.502	0.397	0.929	1.54 (0.58-4.13)
30	M_VLDL_PL_pct	Simple mode	23	-0.042	0.037	0.261	0.902	0.96 (0.89-1.03)
30	M_VLDL_PL_pct	Weighted median	23	-0.017	0.022	0.424	0.832	0.98 (0.94-1.03)
30	M_VLDL_PL_pct	Weighted mode	23	-0.025	0.029	0.400	0.918	0.98 (0.92-1.03)
31	M_VLDL_C_pct	Inverse variance weighted	23	0.073	0.082	0.378	0.889	1.08 (0.91-1.26)
31	M_VLDL_C_pct	MR Egger	23	0.290	0.406	0.482	0.929	1.34 (0.60-2.96)
31	M_VLDL_C_pct	Simple mode	23	-0.006	0.027	0.821	0.962	0.99 (0.94-1.05)
31	M_VLDL_C_pct	Weighted median	23	-0.004	0.019	0.830	0.929	1.00 (0.96-1.03)
31	M_VLDL_C_pct	Weighted mode	23	-0.002	0.024	0.944	0.997	1.00 (0.95-1.05)
32	M_VLDL_CE_pct	Inverse variance weighted	23	0.068	0.077	0.377	0.889	1.07 (0.92-1.25)
32	M_VLDL_CE_pct	MR Egger	23	0.258	0.380	0.505	0.929	1.29 (0.61-2.72)
32	M_VLDL_CE_pct	Simple mode	23	0.003	0.027	0.921	0.984	1.00 (0.95-1.06)
32	M_VLDL_CE_pct	Weighted median	23	-0.001	0.019	0.939	0.959	1.00 (0.96-1.04)
32	M_VLDL_CE_pct	Weighted mode	23	0.007	0.024	0.777	0.997	1.01 (0.96-1.05)
33	M_VLDL_FC_pct	Inverse variance weighted	23	0.082	0.096	0.395	0.889	1.09 (0.90-1.31)
33	M_VLDL_FC_pct	MR Egger	23	0.361	0.471	0.452	0.929	1.44 (0.57-3.62)
33	M_VLDL_FC_pct	Simple mode	23	-0.028	0.030	0.358	0.902	0.97 (0.92-1.03)
33	M_VLDL_FC_pct	Weighted median	23	-0.024	0.020	0.248	0.776	0.98 (0.94-1.02)
33	M_VLDL_FC_pct	Weighted mode	23	-0.017	0.027	0.525	0.918	0.98 (0.93-1.04)
34	M_VLDL_TG_pct	Inverse variance weighted	23	-0.077	0.088	0.385	0.889	0.93 (0.78-1.10)
34	M_VLDL_TG_pct	MR Egger	23	-0.323	0.434	0.464	0.929	0.72 (0.31-1.69)
34	M_VLDL_TG_pct	Simple mode	23	0.020	0.028	0.490	0.902	1.02 (0.97-1.08)
34	M_VLDL_TG_pct	Weighted median	23	0.018	0.020	0.365	0.818	1.02 (0.98-1.06)
34	M_VLDL_TG_pct	Weighted mode	23	0.015	0.027	0.586	0.918	1.02 (0.96-1.07)
35	S_VLDL_C_pct	Inverse variance weighted	23	0.089	0.100	0.375	0.889	1.09 (0.90-1.33)
35	S_VLDL_C_pct	MR Egger	23	0.381	0.492	0.448	0.929	1.46 (0.56-3.84)
35	S_VLDL_C_pct	Simple mode	23	-0.036	0.026	0.180	0.902	0.96 (0.92-1.02)
35	S_VLDL_C_pct	Weighted median	23	-0.021	0.020	0.301	0.783	0.98 (0.94-1.02)
35	S_VLDL_C_pct	Weighted mode	23	-0.026	0.024	0.291	0.918	0.97 (0.93-1.02)
36	S_VLDL_CE_pct	Inverse variance weighted	23	0.101	0.100	0.313	0.889	1.11 (0.91-1.35)
36	S_VLDL_CE_pct	MR Egger	23	0.420	0.492	0.403	0.929	1.52 (0.58-3.99)
36	S_VLDL_CE_pct	Simple mode	23	-0.042	0.031	0.186	0.902	0.96 (0.90-1.02)
36	S_VLDL_CE_pct	Weighted median	23	-0.014	0.020	0.491	0.832	0.99 (0.95-1.03)
36	S_VLDL_CE_pct	Weighted mode	23	-0.021	0.027	0.451	0.918	0.98 (0.93-1.03)
37	S_VLDL_FC_pct	Inverse variance weighted	23	0.058	0.086	0.499	0.889	1.06 (0.90-1.25)
37	S_VLDL_FC_pct	MR Egger	23	0.271	0.423	0.528	0.929	1.31 (0.57-3.00)
37	S_VLDL_FC_pct	Simple mode	23	-0.038	0.030	0.219	0.902	0.96 (0.91-1.02)
37	S_VLDL_FC_pct	Weighted median	23	-0.021	0.020	0.277	0.783	0.98 (0.94-1.02)
37	S_VLDL_FC_pct	Weighted mode	23	-0.023	0.026	0.378	0.918	0.98 (0.93-1.03)
38	S_VLDL_TG_pct	Inverse variance weighted	23	-0.080	0.097	0.410	0.889	0.92 (0.76-1.12)
38	S_VLDL_TG_pct	MR Egger	23	-0.352	0.478	0.469	0.929	0.70 (0.28-1.79)
38	S_VLDL_TG_pct	Simple mode	23	0.038	0.028	0.182	0.902	1.04 (0.98-1.10)
38	S_VLDL_TG_pct	Weighted median	23	0.028	0.020	0.166	0.761	1.03 (0.99-1.07)
38	S_VLDL_TG_pct	Weighted mode	23	0.028	0.027	0.306	0.918	1.03 (0.98-1.08)
39	XS_VLDL_C_pct	Inverse variance weighted	23	0.036	0.074	0.623	0.889	1.04 (0.90-1.20)
39	XS_VLDL_C_pct	MR Egger	23	0.173	0.366	0.641	0.929	1.19 (0.58-2.43)
39	XS_VLDL_C_pct	Simple mode	23	-0.034	0.027	0.223	0.902	0.97 (0.92-1.02)
39	XS_VLDL_C_pct	Weighted median	23	-0.032	0.019	0.094	0.761	0.97 (0.93-1.01)
39	XS_VLDL_C_pct	Weighted mode	23	-0.030	0.027	0.272	0.918	0.97 (0.92-1.02)
40	XS_VLDL_CE_pct	Inverse variance weighted	23	0.031	0.069	0.656	0.889	1.03 (0.90-1.18)

ID	Metabolite	Method	N_{SNP}	Beta	SE	p value	FDR	OR (95% CI)
40	XS_VLDL_CE_pct	MR Egger	23	0.151	0.343	0.664	0.929	1.16 (0.59-2.28)
40	XS_VLDL_CE_pct	Simple mode	23	-0.044	0.028	0.130	0.902	0.96 (0.91-1.01)
40	XS_VLDL_CE_pct	Weighted median	23	-0.036	0.019	0.052	0.658	0.96 (0.93-1.00)
40	XS_VLDL_CE_pct	Weighted mode	23	-0.036	0.027	0.194	0.918	0.97 (0.92-1.02)
41	XS_VLDL_TG_pct	Inverse variance weighted	23	-0.044	0.080	0.584	0.889	0.96 (0.82-1.12)
41	XS_VLDL_TG_pct	MR Egger	23	-0.211	0.394	0.599	0.929	0.81 (0.37-1.75)
41	XS_VLDL_TG_pct	Simple mode	23	0.033	0.031	0.292	0.902	1.03 (0.97-1.10)
41	XS_VLDL_TG_pct	Weighted median	23	0.029	0.020	0.146	0.761	1.03 (0.99-1.07)
41	XS_VLDL_TG_pct	Weighted mode	23	0.024	0.026	0.374	0.918	1.02 (0.97-1.08)
42	M_LDL_CE_pct	Inverse variance weighted	23	-0.002	0.023	0.948	0.989	1.00 (0.95-1.05)
42	M_LDL_CE_pct	MR Egger	23	0.153	0.111	0.182	0.929	1.17 (0.94-1.45)
42	M_LDL_CE_pct	Simple mode	23	-0.020	0.041	0.630	0.902	0.98 (0.91-1.06)
42	M_LDL_CE_pct	Weighted median	23	-0.023	0.021	0.288	0.783	0.98 (0.94-1.02)
42	M_LDL_CE_pct	Weighted mode	23	-0.031	0.035	0.392	0.918	0.97 (0.91-1.04)
43	S_LDL_PL_pct	Inverse variance weighted	23	-0.024	0.036	0.504	0.889	0.98 (0.91-1.05)
43	S_LDL_PL_pct	MR Egger	23	-0.134	0.176	0.455	0.929	0.87 (0.62-1.23)
43	S_LDL_PL_pct	Simple mode	23	-0.007	0.044	0.875	0.979	0.99 (0.91-1.08)
43	S_LDL_PL_pct	Weighted median	23	0.005	0.021	0.817	0.929	1.00 (0.96-1.05)
43	S_LDL_PL_pct	Weighted mode	23	0.002	0.043	0.960	0.997	1.00 (0.92-1.09)
44	S_LDL_CE_pct	Inverse variance weighted	23	0.023	0.038	0.547	0.889	1.02 (0.95-1.10)
44	S_LDL_CE_pct	MR Egger	23	0.165	0.188	0.390	0.929	1.18 (0.82-1.71)
44	S_LDL_CE_pct	Simple mode	23	-0.008	0.037	0.839	0.962	0.99 (0.92-1.07)
44	S_LDL_CE_pct	Weighted median	23	-0.014	0.021	0.517	0.832	0.99 (0.95-1.03)
44	S_LDL_CE_pct	Weighted mode	23	-0.017	0.032	0.605	0.918	0.98 (0.92-1.05)
45	XL_HDL_FC_pct	Inverse variance weighted	23	0.006	0.026	0.804	0.948	1.01 (0.96-1.06)
45	XL_HDL_FC_pct	MR Egger	23	0.059	0.126	0.643	0.929	1.06 (0.83-1.36)
45	XL_HDL_FC_pct	Simple mode	23	-0.013	0.029	0.653	0.902	0.99 (0.93-1.04)
45	XL_HDL_FC_pct	Weighted median	23	-0.022	0.017	0.193	0.761	0.98 (0.95-1.01)
45	XL_HDL_FC_pct	Weighted mode	23	-0.013	0.027	0.630	0.925	0.99 (0.94-1.04)
46	L_HDL_FC_pct	Inverse variance weighted	23	0.010	0.076	0.898	0.977	1.01 (0.87-1.17)
46	L_HDL_FC_pct	MR Egger	23	0.192	0.374	0.612	0.929	1.21 (0.58-2.52)
46	L_HDL_FC_pct	Simple mode	23	-0.025	0.030	0.414	0.902	0.98 (0.92-1.03)
46	L_HDL_FC_pct	Weighted median	23	-0.032	0.020	0.098	0.761	0.97 (0.93-1.01)
46	L_HDL_FC_pct	Weighted mode	23	-0.025	0.029	0.391	0.918	0.98 (0.92-1.03)
47	S_HDL_FC_pct	Inverse variance weighted	23	0.041	0.073	0.574	0.889	1.04 (0.90-1.20)
47	S_HDL_FC_pct	MR Egger	23	0.291	0.356	0.424	0.929	1.34 (0.67-2.69)
47	S_HDL_FC_pct	Simple mode	23	-0.004	0.036	0.902	0.984	1.00 (0.93-1.07)
47	S_HDL_FC_pct	Weighted median	23	0.009	0.020	0.647	0.832	1.01 (0.97-1.05)
47	S_HDL_FC_pct	Weighted mode	23	0.000	0.032	0.997	0.997	1.00 (0.94-1.06)

Abbreviations are in line with **Supplementary Table 3.1**. SNP, single nucleotide polymorphism. SE, standard error. FDR, false discovery rate. OR, odds ratio. CI, confidence interval.

Supplementary Table 6.14 Reverse Mendelian randomisation pleiotropy residual sum and outlier (MR-PRESSO) outlier-corrected results for association of ischaemic stroke with metabolites

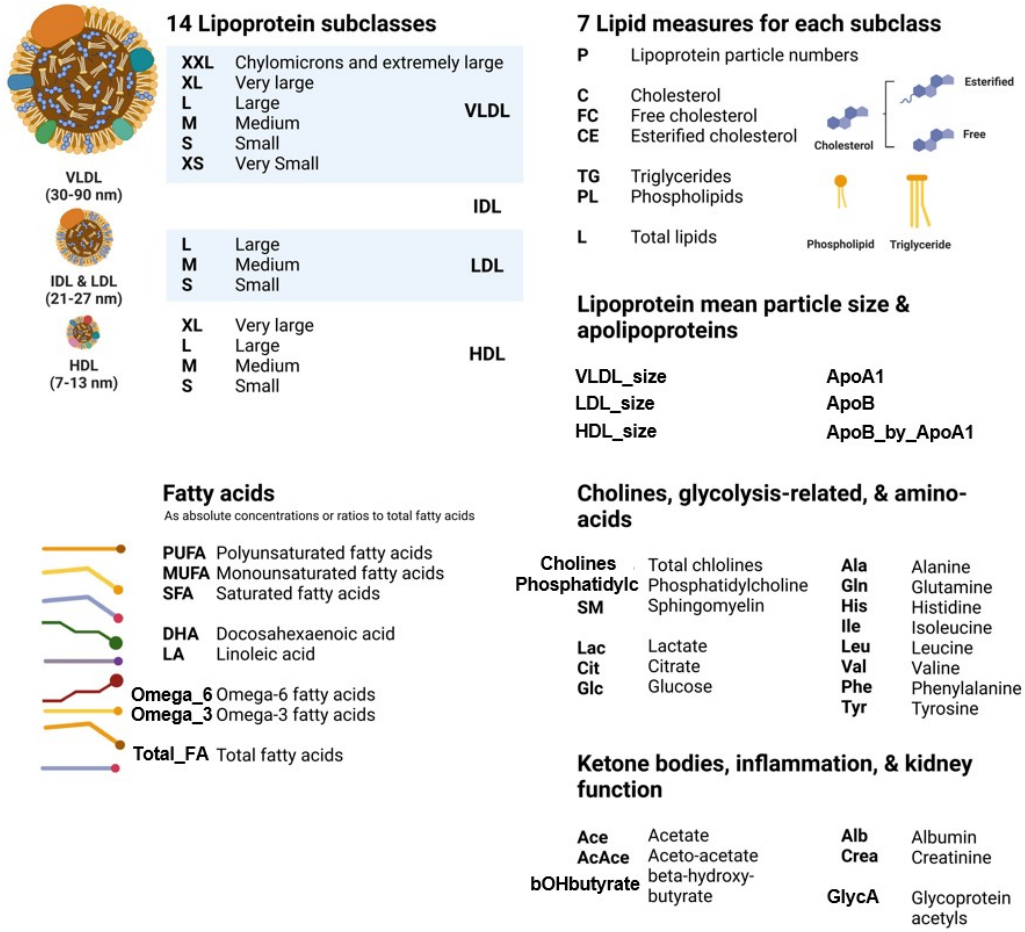
ID	Metabolite	Method	Beta	SE	t statistic	p value	FDR	OR (95% CI)
1	HDL_P	Raw	0.016	0.037	0.438	0.666	0.894	1.02 (0.94, 1.09)
1	HDL_P	Outlier-corrected	0.028	0.018	1.508	0.147	0.701	1.03 (0.99, 1.07)
2	VLDL_size	Raw	-0.043	0.024	-1.809	0.084	0.894	0.96 (0.91, 1)
2	VLDL_size	Outlier-corrected	-0.028	0.020	-1.404	0.175	0.701	0.97 (0.94, 1.01)
3	Omega_3	Raw	0.025	0.034	0.754	0.459	0.894	1.03 (0.96, 1.1)
3	Omega_3	Outlier-corrected	0.007	0.019	0.344	0.734	0.889	1.01 (0.97, 1.05)
4	PUFA_pct	Raw	0.012	0.025	0.472	0.641	0.894	1.01 (0.96, 1.06)
4	PUFA_pct	Outlier-corrected	-0.001	0.020	-0.033	0.974	0.997	1 (0.96, 1.04)
5	LA_pct	Raw	0.009	0.039	0.238	0.814	0.950	1.01 (0.94, 1.09)
5	LA_pct	Outlier-corrected	-0.015	0.016	-0.956	0.351	0.831	0.99 (0.96, 1.02)
6	Phe	Raw	0.017	0.016	1.052	0.304	0.894	1.02 (0.99, 1.05)
6	Phe	Outlier-corrected	NA	NA	NA	NA	NA	NA
7	Glucose	Raw	-0.017	0.038	-0.446	0.660	0.894	0.98 (0.91, 1.06)
7	Glucose	Outlier-corrected	-0.011	0.019	-0.607	0.551	0.831	0.99 (0.95, 1.03)
8	Creatinine	Raw	-0.006	0.018	-0.360	0.722	0.929	0.99 (0.96, 1.03)
8	Creatinine	Outlier-corrected	-0.007	0.016	-0.460	0.650	0.831	0.99 (0.96, 1.02)
9	Albumin	Raw	-0.052	0.030	-1.752	0.094	0.894	0.95 (0.9, 1.01)
9	Albumin	Outlier-corrected	-0.040	0.020	-2.060	0.053	0.701	0.96 (0.92, 1)
10	GlycA	Raw	0.005	0.031	0.156	0.877	0.978	1 (0.95, 1.07)
10	GlycA	Outlier-corrected	-0.016	0.022	-0.760	0.457	0.831	0.98 (0.94, 1.03)
11	XXL_VLDL_L	Raw	-0.035	0.025	-1.361	0.187	0.894	0.97 (0.92, 1.02)
11	XXL_VLDL_L	Outlier-corrected	0.001	0.018	0.075	0.941	0.997	1 (0.97, 1.04)
12	XXL_VLDL_FC	Raw	-0.023	0.021	-1.066	0.298	0.894	0.98 (0.94, 1.02)
12	XXL_VLDL_FC	Outlier-corrected	0.005	0.017	0.296	0.770	0.908	1.01 (0.97, 1.04)
13	XS_VLDL_C	Raw	0.081	0.105	0.772	0.448	0.894	1.08 (0.88, 1.33)
13	XS_VLDL_C	Outlier-corrected	0.038	0.018	2.113	0.053	0.701	1.04 (1, 1.08)
14	XS_VLDL_CE	Raw	0.079	0.106	0.750	0.461	0.894	1.08 (0.88, 1.33)

ID	Metabolite	Method	Beta	SE	t statistic	p value	FDR	OR (95% CI)
14	XS_VLDL_CE	Outlier-corrected	0.036	0.018	2.024	0.062	0.701	1.04 (1, 1.07)
15	XL_HDL_L	Raw	-0.005	0.045	-0.107	0.916	0.978	1 (0.91, 1.09)
15	XL_HDL_L	Outlier-corrected	0.000	0.017	-0.004	0.997	0.997	1 (0.97, 1.03)
16	XL_HDL_C	Raw	-0.001	0.050	-0.013	0.989	0.989	1 (0.91, 1.1)
16	XL_HDL_C	Outlier-corrected	-0.001	0.017	-0.045	0.965	0.997	1 (0.97, 1.03)
17	XL_HDL_FC	Raw	-0.002	0.059	-0.038	0.970	0.989	1 (0.89, 1.12)
17	XL_HDL_FC	Outlier-corrected	-0.009	0.018	-0.491	0.629	0.831	0.99 (0.96, 1.03)
18	M_HDL_L	Raw	-0.008	0.035	-0.218	0.829	0.950	0.99 (0.93, 1.06)
18	M_HDL_L	Outlier-corrected	0.024	0.015	1.565	0.135	0.701	1.02 (0.99, 1.05)
19	M_HDL_PL	Raw	-0.012	0.037	-0.321	0.751	0.929	0.99 (0.92, 1.06)
19	M_HDL_PL	Outlier-corrected	0.021	0.016	1.325	0.202	0.701	1.02 (0.99, 1.05)
20	S_HDL_P	Raw	0.037	0.052	0.708	0.486	0.894	1.04 (0.94, 1.15)
20	S_HDL_P	Outlier-corrected	0.033	0.020	1.632	0.121	0.701	1.03 (0.99, 1.07)
21	S_HDL_L	Raw	0.023	0.046	0.508	0.616	0.894	1.02 (0.94, 1.12)
21	S_HDL_L	Outlier-corrected	0.020	0.018	1.070	0.302	0.808	1.02 (0.98, 1.06)
22	S_HDL_PL	Raw	0.015	0.044	0.334	0.741	0.929	1.01 (0.93, 1.11)
22	S_HDL_PL	Outlier-corrected	0.018	0.021	0.866	0.399	0.831	1.02 (0.98, 1.06)
23	S_HDL_C	Raw	0.034	0.051	0.660	0.516	0.894	1.03 (0.94, 1.14)
23	S_HDL_C	Outlier-corrected	0.022	0.020	1.146	0.267	0.767	1.02 (0.98, 1.06)
24	S_HDL_CE	Raw	0.030	0.051	0.601	0.554	0.894	1.03 (0.93, 1.14)
24	S_HDL_CE	Outlier-corrected	0.016	0.019	0.841	0.411	0.831	1.02 (0.98, 1.05)
25	XXL_VLDL_C_pct	Raw	0.092	0.049	1.862	0.076	0.894	1.1 (1, 1.21)
25	XXL_VLDL_C_pct	Outlier-corrected	0.025	0.015	1.622	0.121	0.701	1.03 (0.99, 1.06)
26	XXL_VLDL_CE_pct	Raw	0.094	0.053	1.771	0.090	0.894	1.1 (0.99, 1.22)
26	XXL_VLDL_CE_pct	Outlier-corrected	0.021	0.015	1.348	0.193	0.701	1.02 (0.99, 1.05)
27	XL_VLDL_C_pct	Raw	0.078	0.070	1.118	0.276	0.894	1.08 (0.94, 1.24)
27	XL_VLDL_C_pct	Outlier-corrected	0.023	0.015	1.526	0.146	0.701	1.02 (0.99, 1.05)
28	XL_VLDL_CE_pct	Raw	0.078	0.066	1.181	0.250	0.894	1.08 (0.95, 1.23)
28	XL_VLDL_CE_pct	Outlier-corrected	0.019	0.015	1.293	0.213	0.701	1.02 (0.99, 1.05)

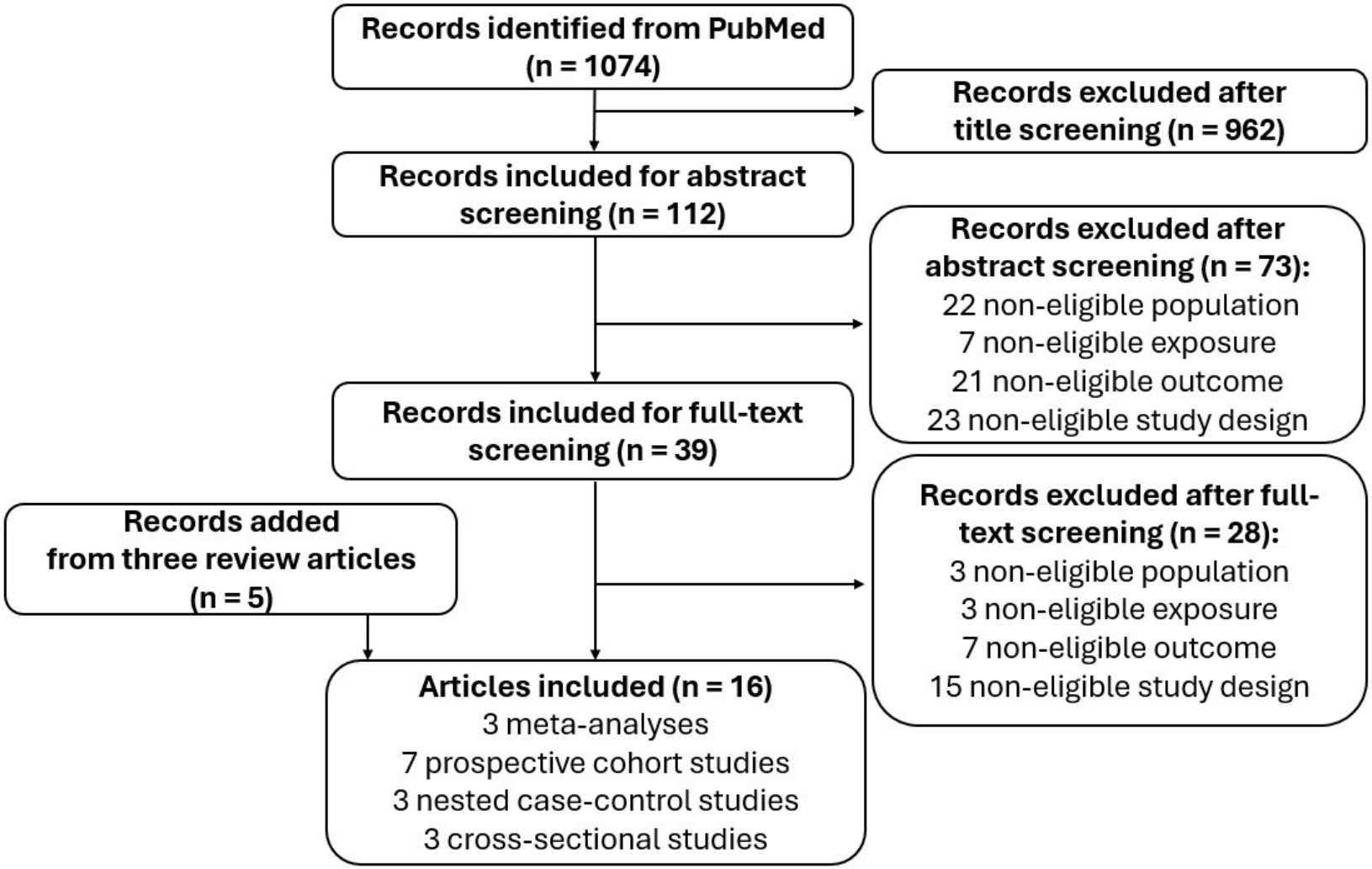
ID	Metabolite	Method	Beta	SE	t statistic	p value	FDR	OR (95% CI)
29	XL_VLDL_TG_pct	Raw	-0.086	0.083	-1.044	0.308	0.894	0.92 (0.78, 1.08)
29	XL_VLDL_TG_pct	Outlier-corrected	-0.019	0.018	-1.035	0.316	0.808	0.98 (0.95, 1.02)
30	M_VLDL_PL_pct	Raw	0.088	0.102	0.857	0.400	0.894	1.09 (0.89, 1.33)
30	M_VLDL_PL_pct	Outlier-corrected	0.017	0.021	0.807	0.432	0.831	1.02 (0.98, 1.06)
31	M_VLDL_C_pct	Raw	0.073	0.082	0.881	0.388	0.894	1.08 (0.91, 1.26)
31	M_VLDL_C_pct	Outlier-corrected	0.010	0.019	0.500	0.623	0.831	1.01 (0.97, 1.05)
32	M_VLDL_CE_pct	Raw	0.068	0.077	0.884	0.386	0.894	1.07 (0.92, 1.25)
32	M_VLDL_CE_pct	Outlier-corrected	0.010	0.018	0.529	0.603	0.831	1.01 (0.97, 1.05)
33	M_VLDL_FC_pct	Raw	0.082	0.096	0.851	0.404	0.894	1.09 (0.9, 1.31)
33	M_VLDL_FC_pct	Outlier-corrected	0.024	0.020	1.209	0.246	0.753	1.02 (0.99, 1.06)
34	M_VLDL_TG_pct	Raw	-0.077	0.088	-0.869	0.394	0.894	0.93 (0.78, 1.1)
34	M_VLDL_TG_pct	Outlier-corrected	-0.012	0.019	-0.626	0.540	0.831	0.99 (0.95, 1.03)
35	S_VLDL_C_pct	Raw	0.089	0.100	0.888	0.384	0.894	1.09 (0.9, 1.33)
35	S_VLDL_C_pct	Outlier-corrected	0.011	0.019	0.557	0.585	0.831	1.01 (0.97, 1.05)
36	S_VLDL_CE_pct	Raw	0.101	0.100	1.008	0.324	0.894	1.11 (0.91, 1.35)
36	S_VLDL_CE_pct	Outlier-corrected	0.014	0.021	0.697	0.496	0.831	1.01 (0.97, 1.06)
37	S_VLDL_FC_pct	Raw	0.058	0.086	0.675	0.506	0.894	1.06 (0.9, 1.25)
37	S_VLDL_FC_pct	Outlier-corrected	0.006	0.018	0.354	0.728	0.889	1.01 (0.97, 1.04)
38	S_VLDL_TG_pct	Raw	-0.080	0.097	-0.824	0.419	0.894	0.92 (0.76, 1.12)
38	S_VLDL_TG_pct	Outlier-corrected	-0.011	0.019	-0.594	0.561	0.831	0.99 (0.95, 1.03)
39	XS_VLDL_C_pct	Raw	0.036	0.074	0.492	0.627	0.894	1.04 (0.9, 1.2)
39	XS_VLDL_C_pct	Outlier-corrected	-0.009	0.018	-0.505	0.620	0.831	0.99 (0.96, 1.03)
40	XS_VLDL_CE_pct	Raw	0.031	0.069	0.446	0.660	0.894	1.03 (0.9, 1.18)
40	XS_VLDL_CE_pct	Outlier-corrected	-0.009	0.017	-0.557	0.584	0.831	0.99 (0.96, 1.02)
41	XS_VLDL_TG_pct	Raw	-0.044	0.080	-0.547	0.590	0.894	0.96 (0.82, 1.12)
41	XS_VLDL_TG_pct	Outlier-corrected	0.000	0.019	0.021	0.984	0.997	1 (0.96, 1.04)
42	M_LDL_CE_pct	Raw	-0.002	0.023	-0.065	0.949	0.989	1 (0.95, 1.05)
42	M_LDL_CE_pct	Outlier-corrected	-0.025	0.018	-1.373	0.185	0.701	0.98 (0.94, 1.01)
43	S_LDL_PL_pct	Raw	-0.024	0.036	-0.668	0.511	0.894	0.98 (0.91, 1.05)

ID	Metabolite	Method	Beta	SE	t statistic	p value	FDR	OR (95% CI)
43	S_LDL_PL_pct	Outlier-corrected	0.000	0.019	-0.014	0.989	0.997	1 (0.96, 1.04)
44	S_LDL_CE_pct	Raw	0.023	0.038	0.603	0.553	0.894	1.02 (0.95, 1.1)
44	S_LDL_CE_pct	Outlier-corrected	-0.008	0.017	-0.483	0.635	0.831	0.99 (0.96, 1.03)
45	XL_HDL_FC_pct	Raw	0.006	0.026	0.248	0.806	0.950	1.01 (0.96, 1.06)
45	XL_HDL_FC_pct	Outlier-corrected	-0.027	0.013	-1.991	0.061	0.701	0.97 (0.95, 1)
46	L_HDL_FC_pct	Raw	0.010	0.076	0.128	0.899	0.978	1.01 (0.87, 1.17)
46	L_HDL_FC_pct	Outlier-corrected	0.003	0.014	0.199	0.845	0.972	1 (0.98, 1.03)
47	S_HDL_FC_pct	Raw	0.041	0.073	0.563	0.579	0.894	1.04 (0.9, 1.2)
47	S_HDL_FC_pct	Outlier-corrected	0.008	0.016	0.529	0.603	0.831	1.01 (0.98, 1.04)

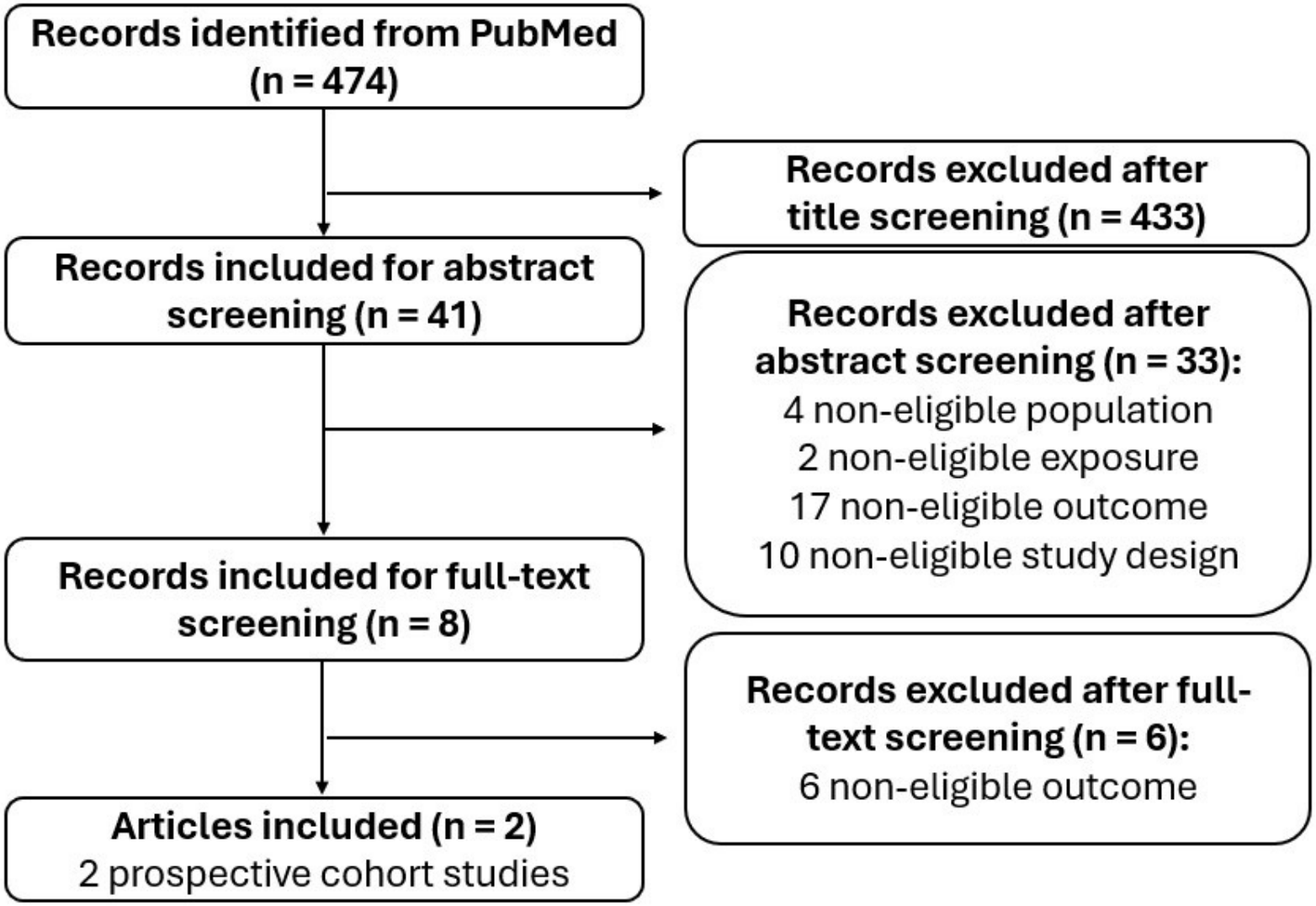
SE, standard error. FDR, false discovery rate. OR, odds ratio. CI, confidence interval. Abbreviations are in line with **Supplementary Table 3.1**.



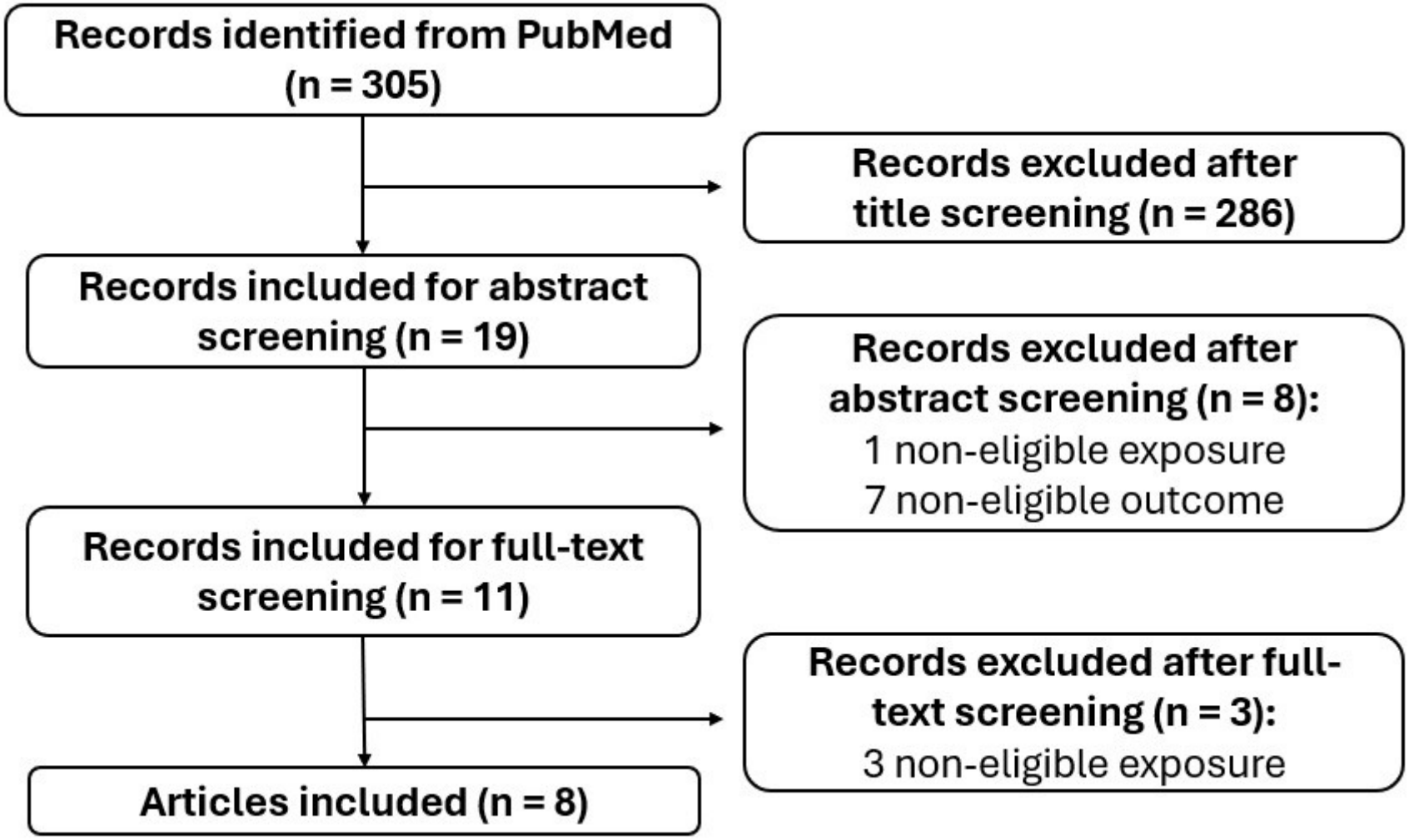
Supplementary Figure 1.1 Metabolic measures quantified by nuclear magnetic resonance spectroscopy
HDL, high-density lipoprotein. IDL, intermediate-density lipoprotein. LDL, low-density lipoprotein. VLDL, very low-density lipoprotein. Other abbreviations are in line with **Supplementary Table 3.1**. Modified from: Aguilar-Ramirez et al. (2022) *Communications Medicine*.



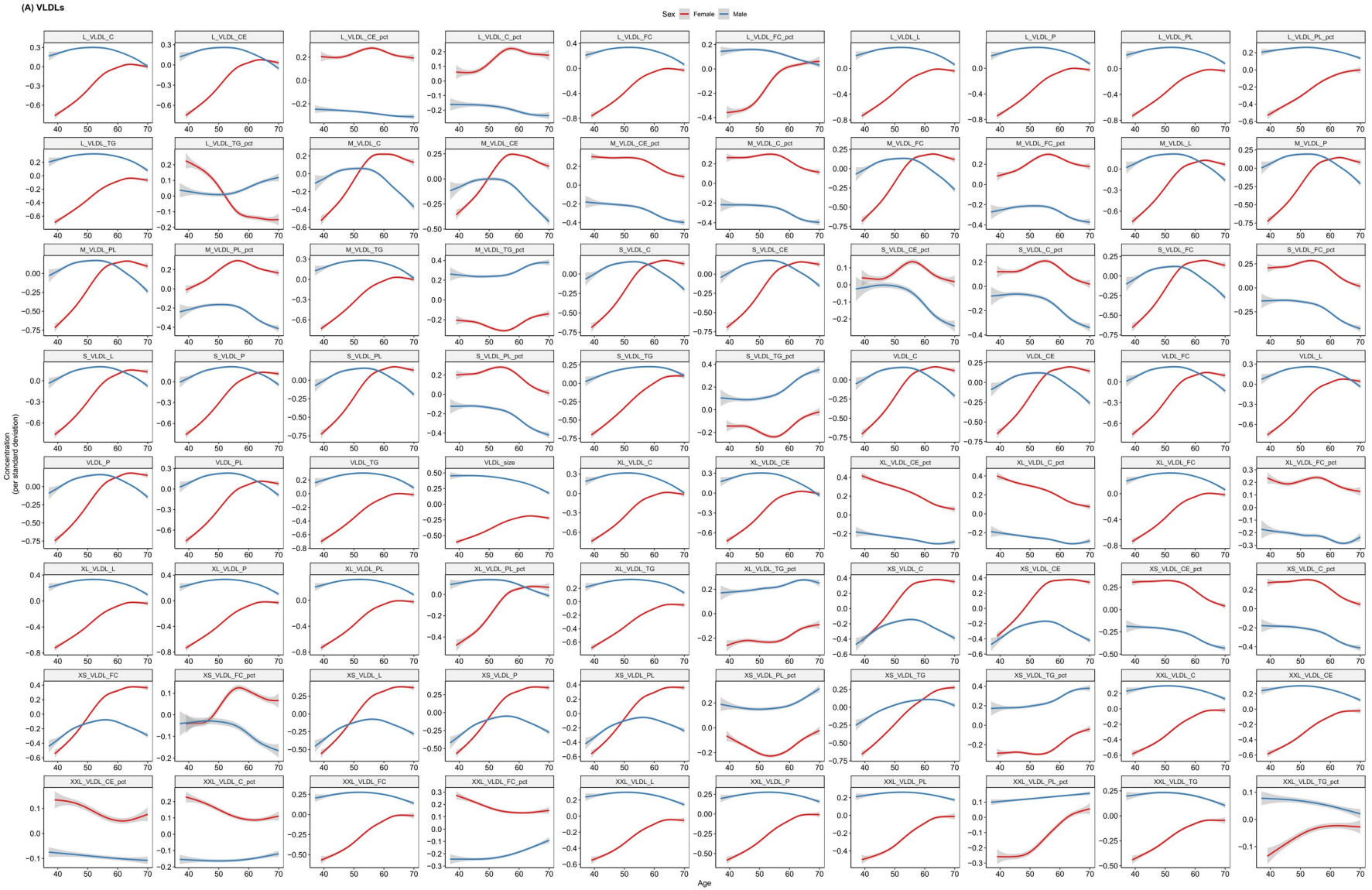
Supplementary Figure 2.1 Flowchart of article selection and inclusion in literature review for the first objective of the thesis

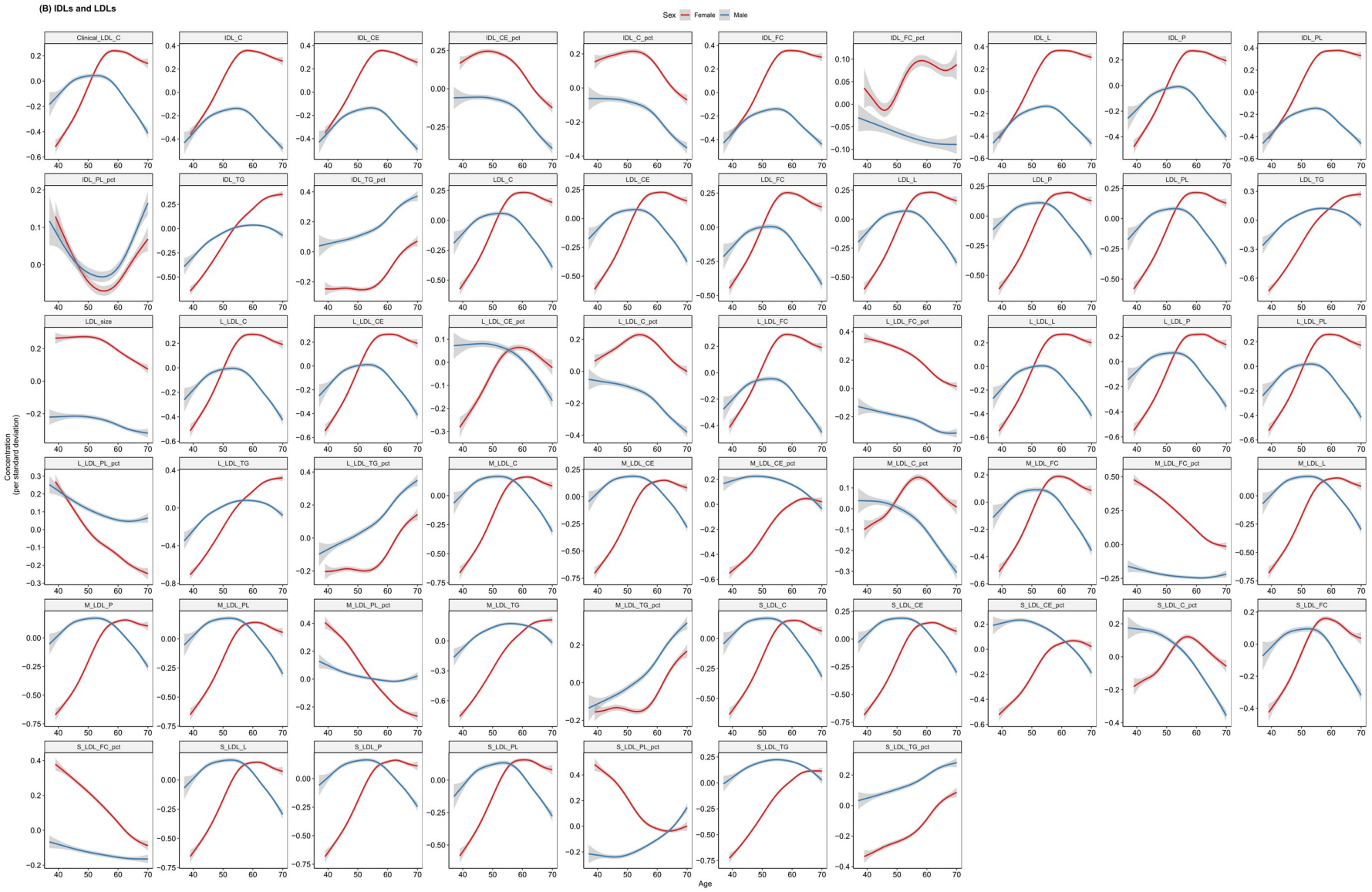


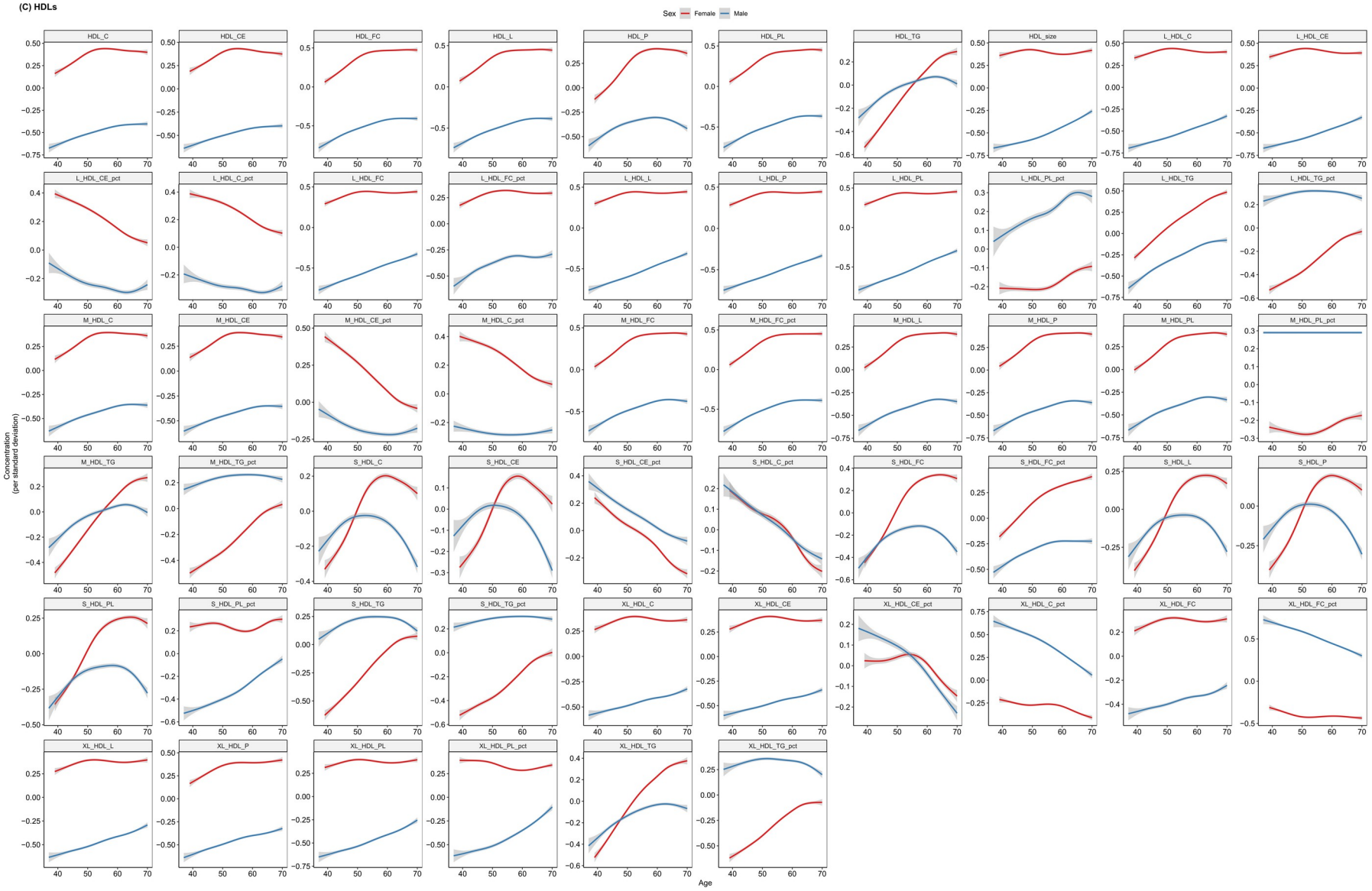
Supplementary Figure 2.2 Flowchart of article selection and inclusion in literature review for the second objective of the thesis

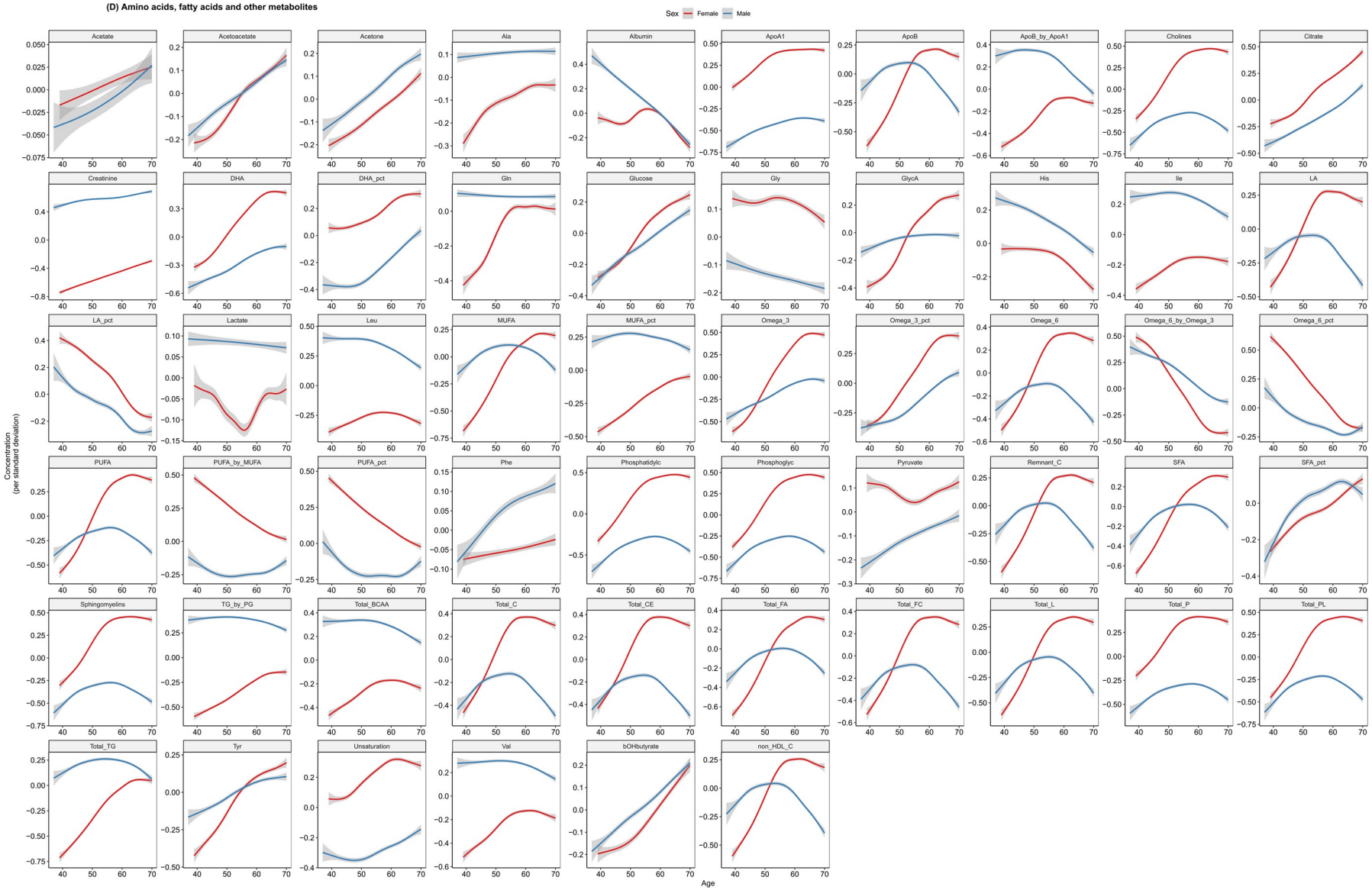


Supplementary Figure 2.3 Flowchart of article selection and inclusion in literature review for the third objective of the thesis


(A) VLDLs
Sex Female Male
L_VLDL_C
L_VLDL_CE
L_VLDL_CE_pct
L_VLDL_C_pct
L_VLDL_FC
L_VLDL_FC_pct
L_VLDL_L
L_VLDL_P
L_VLDL_PL
L_VLDL_PL_pct
L_VLDL_TG
L_VLDL_TG_pct
M_VLDL_C
M_VLDL_CE
M_VLDL_CE_pct
M_VLDL_C_pct
M_VLDL_FC
M_VLDL_FC_pct
M_VLDL_L
M_VLDL_P
M_VLDL_PL
M_VLDL_PL_pct
M_VLDL_TG
M_VLDL_TG_pct
S_VLDL_C
S_VLDL_CE
S_VLDL_CE_pct
S_VLDL_C_pct
S_VLDL_FC
S_VLDL_FC_pct
S_VLDL_L
S_VLDL_P
S_VLDL_PL
S_VLDL_PL_pct
S_VLDL_TG
S_VLDL_TG_pct
VLDL_C
VLDL_CE
VLDL_FC
VLDL_L
VLDL_P
VLDL_PL
VLDL_TG
VLDL_size
XL_VLDL_C
XL_VLDL_CE
XL_VLDL_CE_pct
XL_VLDL_C_pct
XL_VLDL_FC
XL_VLDL_FC_pct
XL_VLDL_L
XL_VLDL_P
XL_VLDL_PL
XL_VLDL_PL_pct
XL_VLDL_TG
XL_VLDL_TG_pct
XS_VLDL_C
XS_VLDL_CE
XS_VLDL_CE_pct
XS_VLDL_C_pct
XS_VLDL_FC
XS_VLDL_FC_pct
XS_VLDL_L
XS_VLDL_P
XS_VLDL_PL
XS_VLDL_PL_pct
XS_VLDL_TG
XS_VLDL_TG_pct
XXL_VLDL_C
XXL_VLDL_CE
XXL_VLDL_CE_pct
XXL_VLDL_C_pct
XXL_VLDL_FC
XXL_VLDL_FC_pct
XXL_VLDL_L
XXL_VLDL_P
XXL_VLDL_PL
XXL_VLDL_PL_pct
XXL_VLDL_TG
XXL_VLDL_TG_pct
Concentration (per standard deviation)
Age


(B) IDLs and LDLs
Sex Female Male
Clinical_LDL_C
IDL_C
IDL_CE
IDL_CE_pct
IDL_C_pct
IDL_FC
IDL_FC_pct
IDL_L
IDL_P
IDL_PL
IDL_PL_pct
IDL_TG
IDL_TG_pct
LDL_C
LDL_CE
LDL_FC
LDL_L
LDL_P
LDL_PL
LDL_TG
LDL_size
L_LDL_C
L_LDL_CE
L_LDL_CE_pct
L_LDL_C_pct
L_LDL_FC
L_LDL_FC_pct
L_LDL_L
L_LDL_P
L_LDL_PL
L_LDL_PL_pct
L_LDL_TG
L_LDL_TG_pct
M_LDL_C
M_LDL_CE
M_LDL_CE_pct
M_LDL_C_pct
M_LDL_FC
M_LDL_FC_pct
M_LDL_L
M_LDL_P
M_LDL_PL
M_LDL_PL_pct
M_LDL_TG
M_LDL_TG_pct
S_LDL_C
S_LDL_CE
S_LDL_CE_pct
S_LDL_C_pct
S_LDL_FC
S_LDL_FC_pct
S_LDL_L
S_LDL_P
S_LDL_PL
S_LDL_PL_pct
S_LDL_TG
S_LDL_TG_pct
Concentration (per standard deviation)
Age

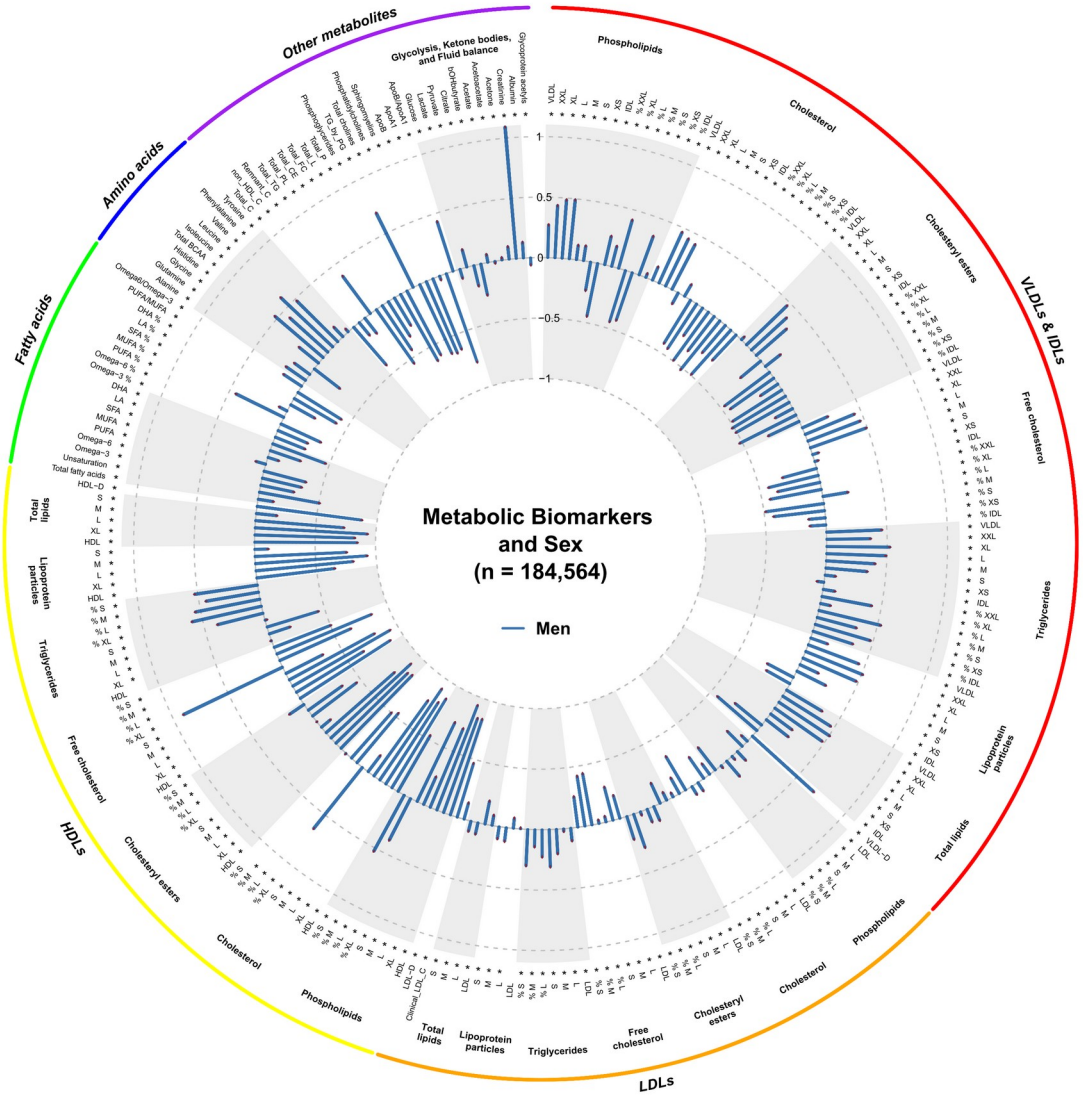

(C) HDLs
Sex Female Male
HDL_C
HDL_CE
HDL_FC
HDL_L
HDL_P
HDL_PL
HDL_TG
HDL_size
L_HDL_C
L_HDL_CE
L_HDL_CE_pct
L_HDL_C_pct
L_HDL_FC
L_HDL_FC_pct
L_HDL_L
L_HDL_P
L_HDL_PL
L_HDL_PL_pct
L_HDL_TG
L_HDL_TG_pct
M_HDL_C
M_HDL_CE
M_HDL_CE_pct
M_HDL_C_pct
M_HDL_FC
M_HDL_FC_pct
M_HDL_L
M_HDL_P
M_HDL_PL
M_HDL_PL_pct
M_HDL_TG
M_HDL_TG_pct
S_HDL_C
S_HDL_CE
S_HDL_CE_pct
S_HDL_C_pct
S_HDL_FC
S_HDL_FC_pct
S_HDL_L
S_HDL_P
S_HDL_PL
S_HDL_PL_pct
S_HDL_TG
S_HDL_TG_pct
XL_HDL_C
XL_HDL_CE
XL_HDL_CE_pct
XL_HDL_C_pct
XL_HDL_FC
XL_HDL_FC_pct
XL_HDL_L
XL_HDL_P
XL_HDL_PL
XL_HDL_PL_pct
XL_HDL_TG
XL_HDL_TG_pct
Concentration (per standard deviation)
Age


(D) Amino acids, fatty acids and other metabolites
Sex Female Male
Acetate
Acetoacetate
Acetone
Ala
Albumin
ApoA1
ApoB
ApoB_by_ApoA1
Cholines
Citrate
Creatinine
DHA
DHA_pct
Gln
Glucose
Gly
GlycA
His
Ile
LA
LA_pct
Lactate
Leu
MUFA
MUFA_pct
Omega_3
Omega_3_pct
Omega_6
Omega_6_by_Omega_3
Omega_6_pct
PUFA
PUFA_by_MUFA
PUFA_pct
Phe
Phosphatidylc
Phosphoglyc
Pyruvate
Remnant_C
SFA
SFA_pct
Sphingomyelins
TG_by_PG
Total_BCAA
Total_C
Total_CE
Total_FA
Total_FC
Total_L
Total_P
Total_PL
Total_TG
Tyr
Unsaturation
Val
bOHbutyrate
non_HDL_C
Concentration (per standard deviation)
Age

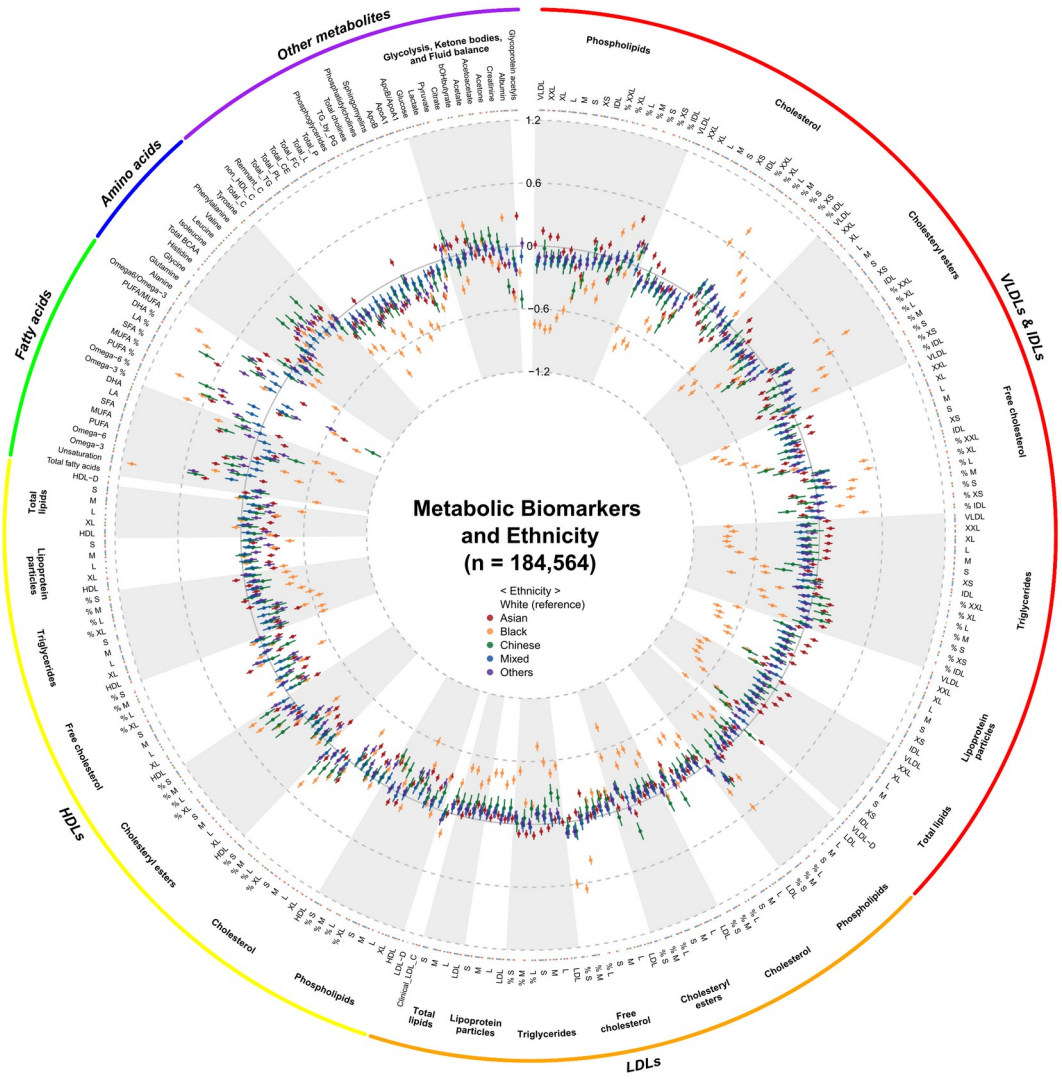
Supplementary Figure 3.1 The distribution of NMR metabolites by age and sex (n = 184,564)

Generalized additive models were used to smooth the lines in the figure. The ranges in the y−axis of each panel correspond to +/− 1 standard deviations and 95% confidence intervals from the study population mean of the relevant NMR metabolites. Blue: male; Red, female. (A) VLDLs, (B) IDLs and LDLs, (C) HDLs, (D) Amino acids, fatty acids, and other metabolites. Abbreviations are in line with **Supplementary Table 3.1**.

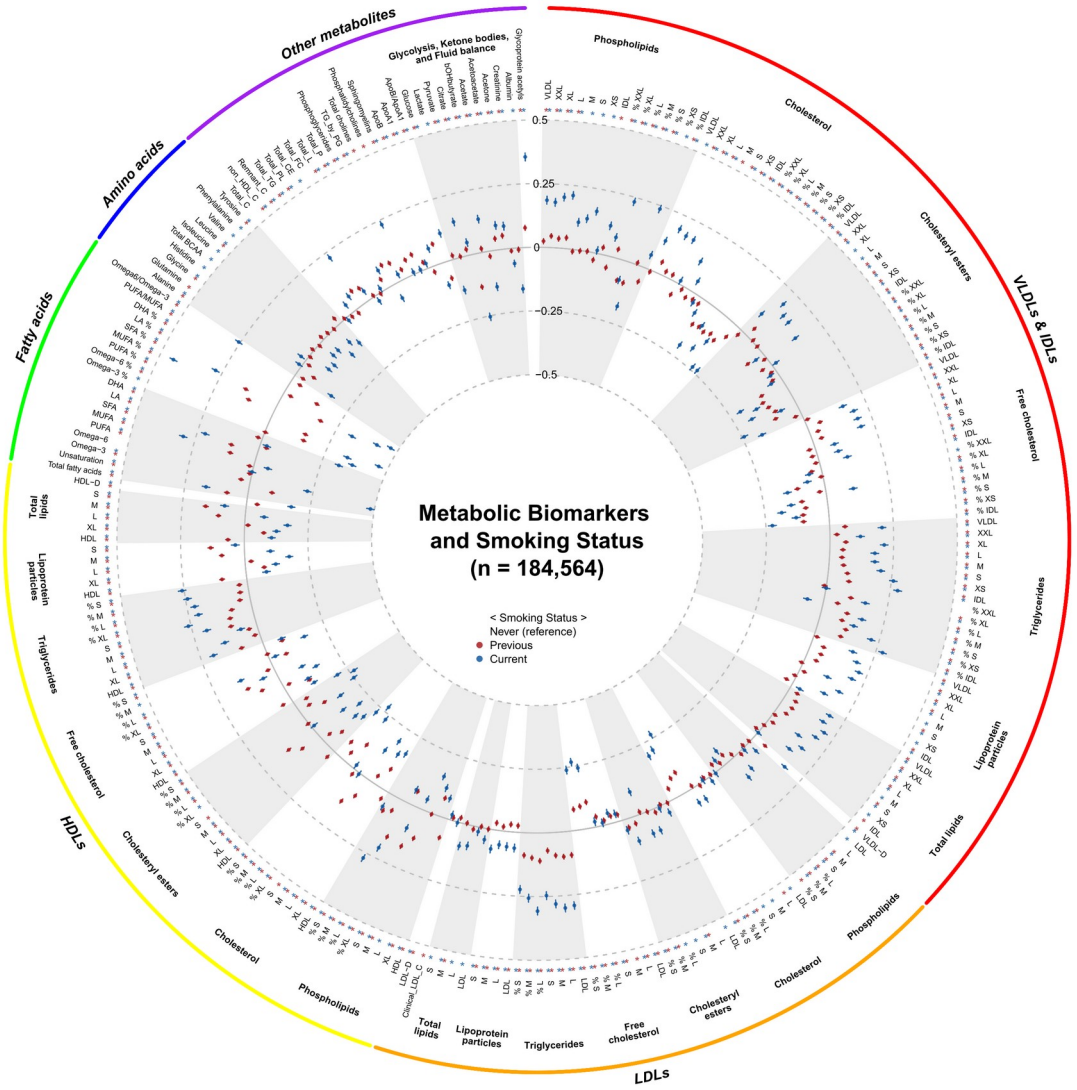
(A)


Metabolic Biomarkers
and Sex
(n = 184,564)
Men
VLDLs & IDLs
LDLs
HDLs
Fatty acids
Amino acids
Other metabolites
Phospholipids
Cholesterol
Cholesteryl esters
Free cholesterol
Triglycerides
Lipoprotein particles
Total lipids
Glycolysis, Ketone bodies, and Fluid balance

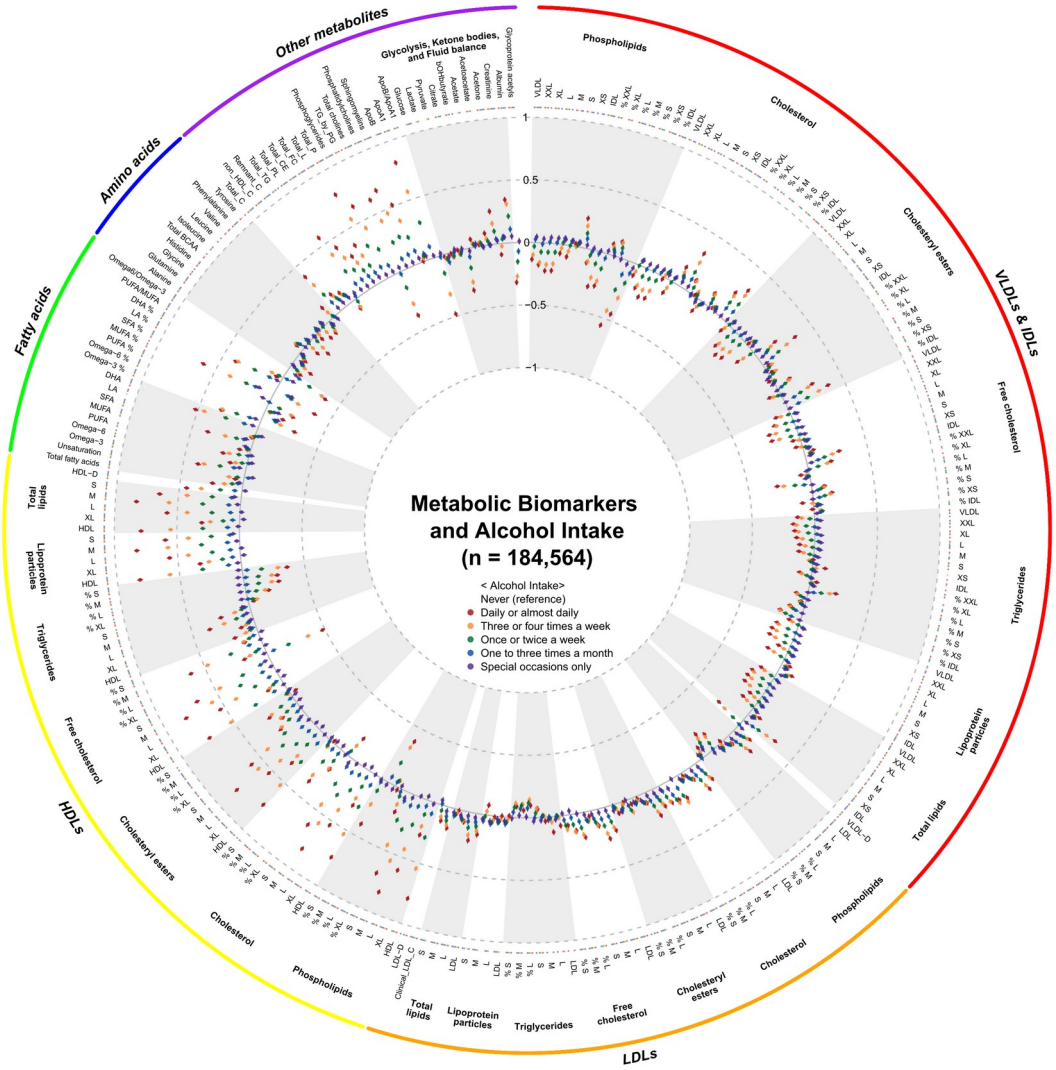
(B)


Metabolic Biomarkers
and Ethnicity
(n = 184,564)
< Ethnicity >
White (reference)
Asian
Black
Chinese
Mixed
Others
VLDLs & IDLs
Free cholesterol
Triglycerides
Lipoprotein particles
Total lipids
Phospholipids
Cholesterol
Cholesteryl esters
LDLs
HDLs
Fatty acids
Amino acids
Other metabolites
Glycolysis, Ketone bodies, and Fluid balance

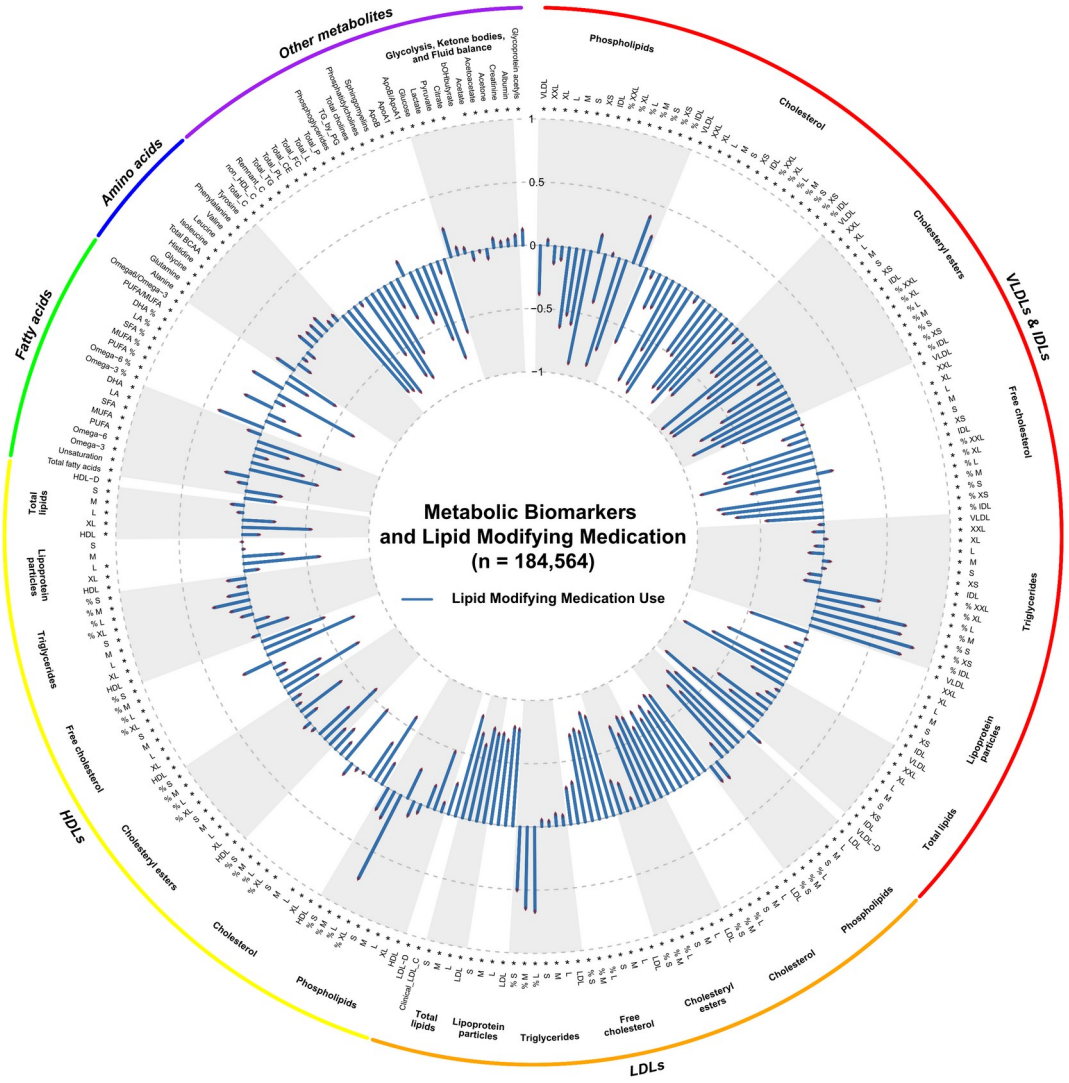
(C)


Metabolic Biomarkers
and Smoking Status
(n = 184,564)
< Smoking Status >
Never (reference)
Previous
Current
VLDLs & IDLs
LDLs
HDLs
Fatty acids
Amino acids
Other metabolites
Phospholipids
Cholesterol
Cholesteryl esters
Free cholesterol
Triglycerides
Lipoprotein particles
Total lipids
Glycolysis, Ketone bodies, and Fluid balance

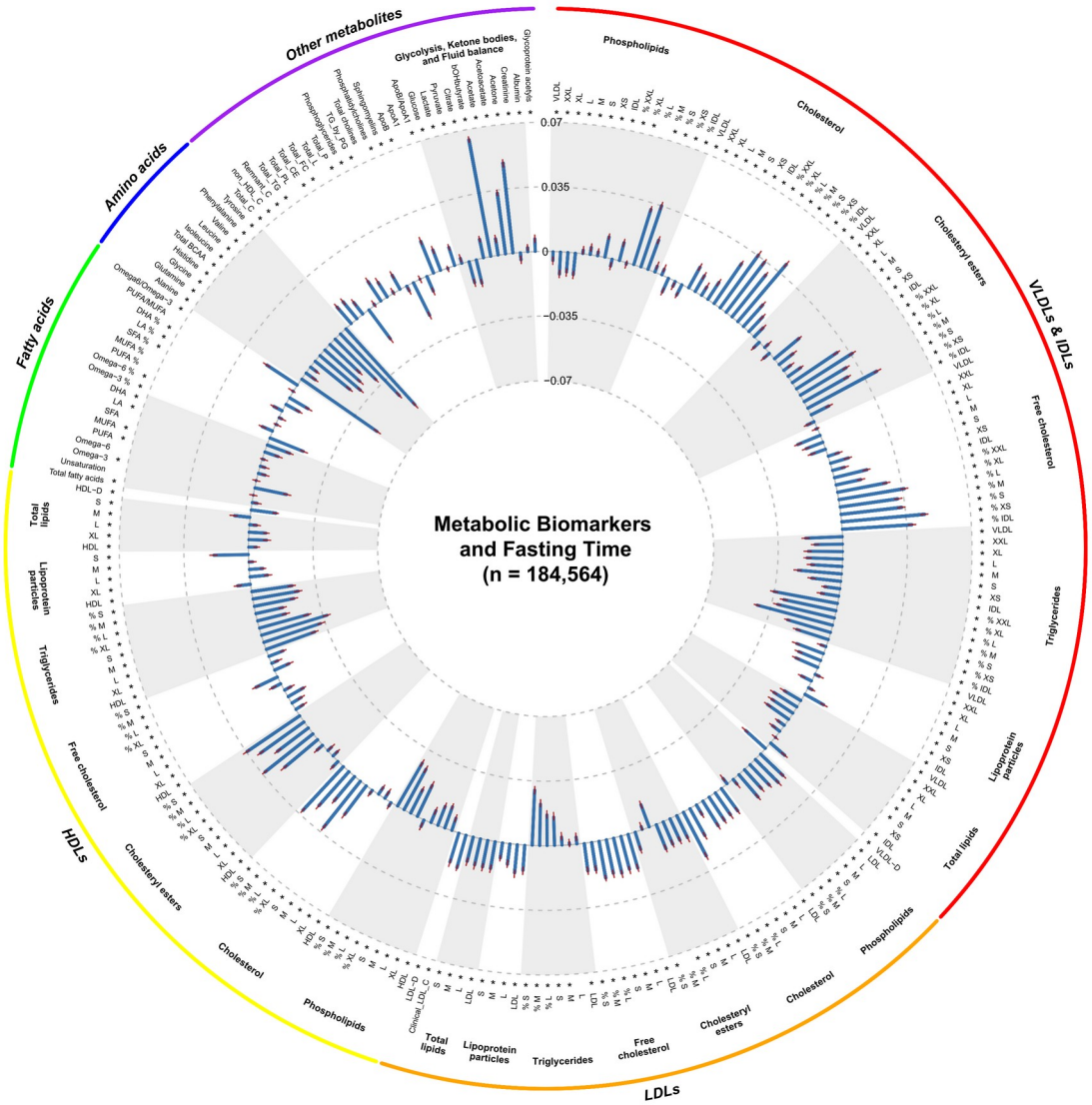
(D)


Metabolic Biomarkers
and Alcohol Intake
(n = 184,564)
< Alcohol Intake>
Never (reference)
Daily or almost daily
Three or four times a week
Once or twice a week
One to three times a month
Special occasions only
Other metabolites
Amino acids
Fatty acids
VLDLs & IDLs
HDLs
LDLs
Glycolysis, Ketone bodies, and Fluid balance
Phospholipids
Cholesterol
Cholesteryl esters
Free cholesterol
Triglycerides
Lipoprotein particles
Total lipids

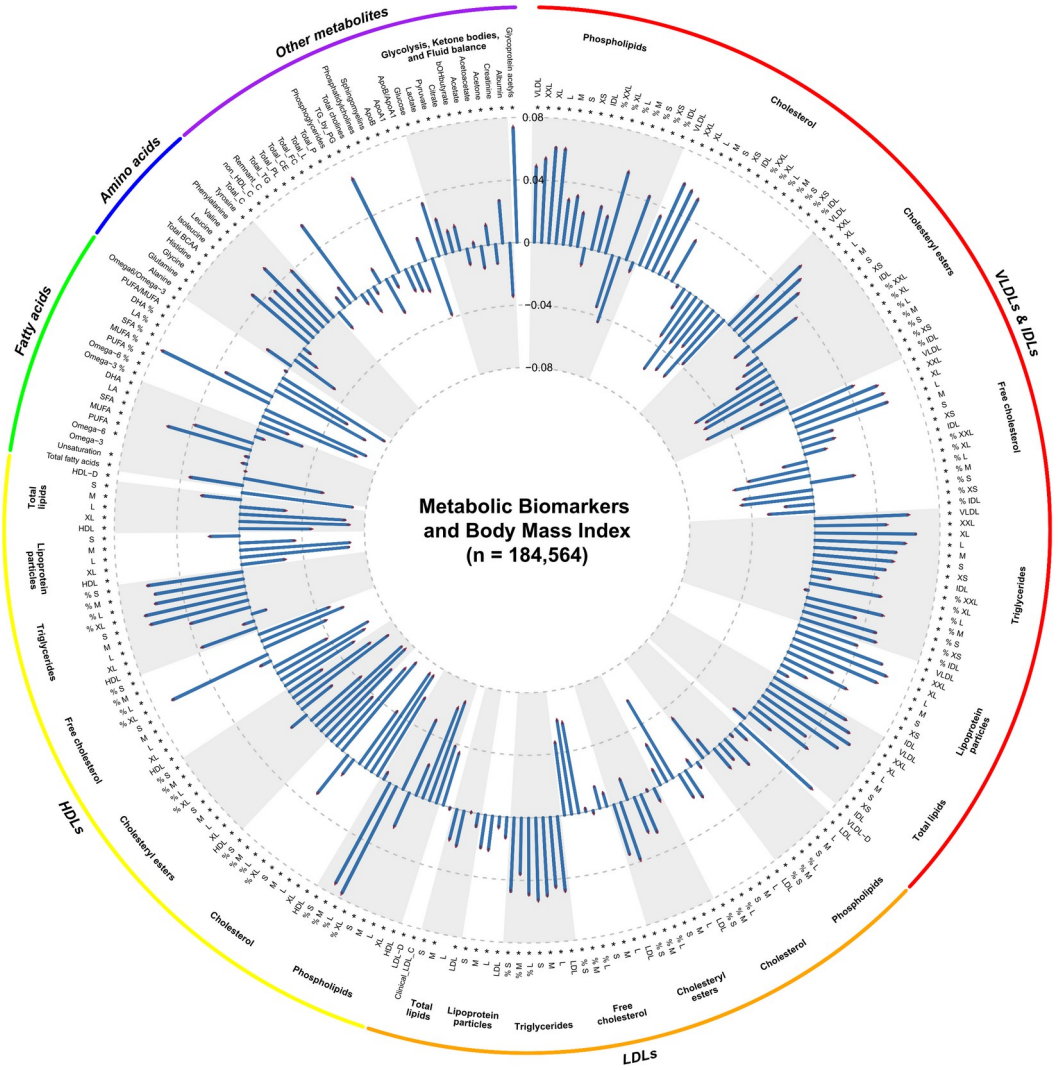
(E)


Metabolic Biomarkers
and Lipid Modifying Medication
(n = 184,564)
Lipid Modifying Medication Use
VLDLs & IDLs
LDLs
HDLs
Fatty acids
Amino acids
Other metabolites
Phospholipids
Cholesterol
Cholesteryl esters
Free cholesterol
Triglycerides
Lipoprotein particles
Total lipids
Glycolysis, Ketone bodies, and Fluid balance

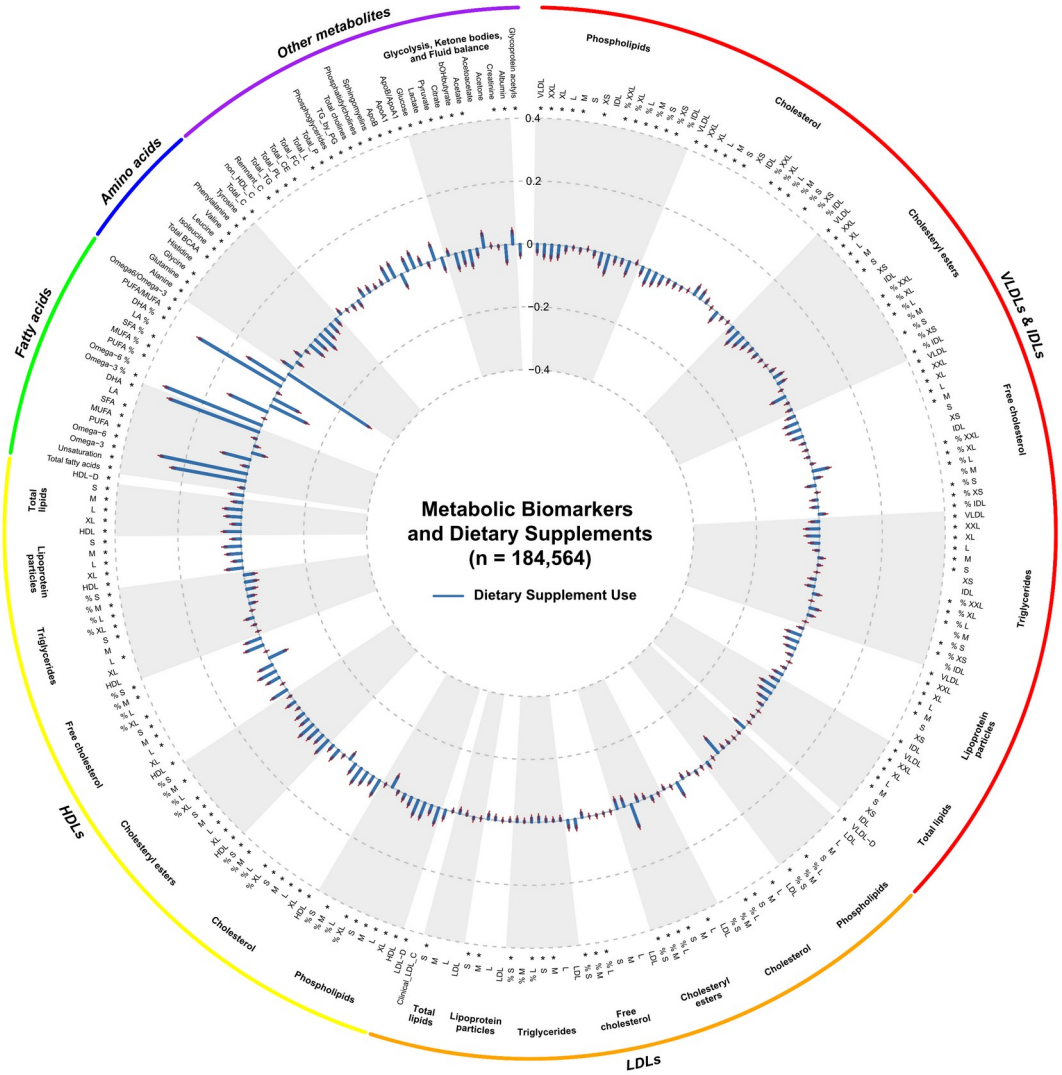
(F)


Metabolic Biomarkers
and Fasting Time
(n = 184,564)
VLDLs & IDLs
LDLs
HDLs
Fatty acids
Amino acids
Other metabolites
Phospholipids
Cholesterol
Cholesteryl esters
Free cholesterol
Triglycerides
Lipoprotein particles
Total lipids
Glycolysis, Ketone bodies, and Fluid balance
0.07
0.035
0
−0.035
−0.07

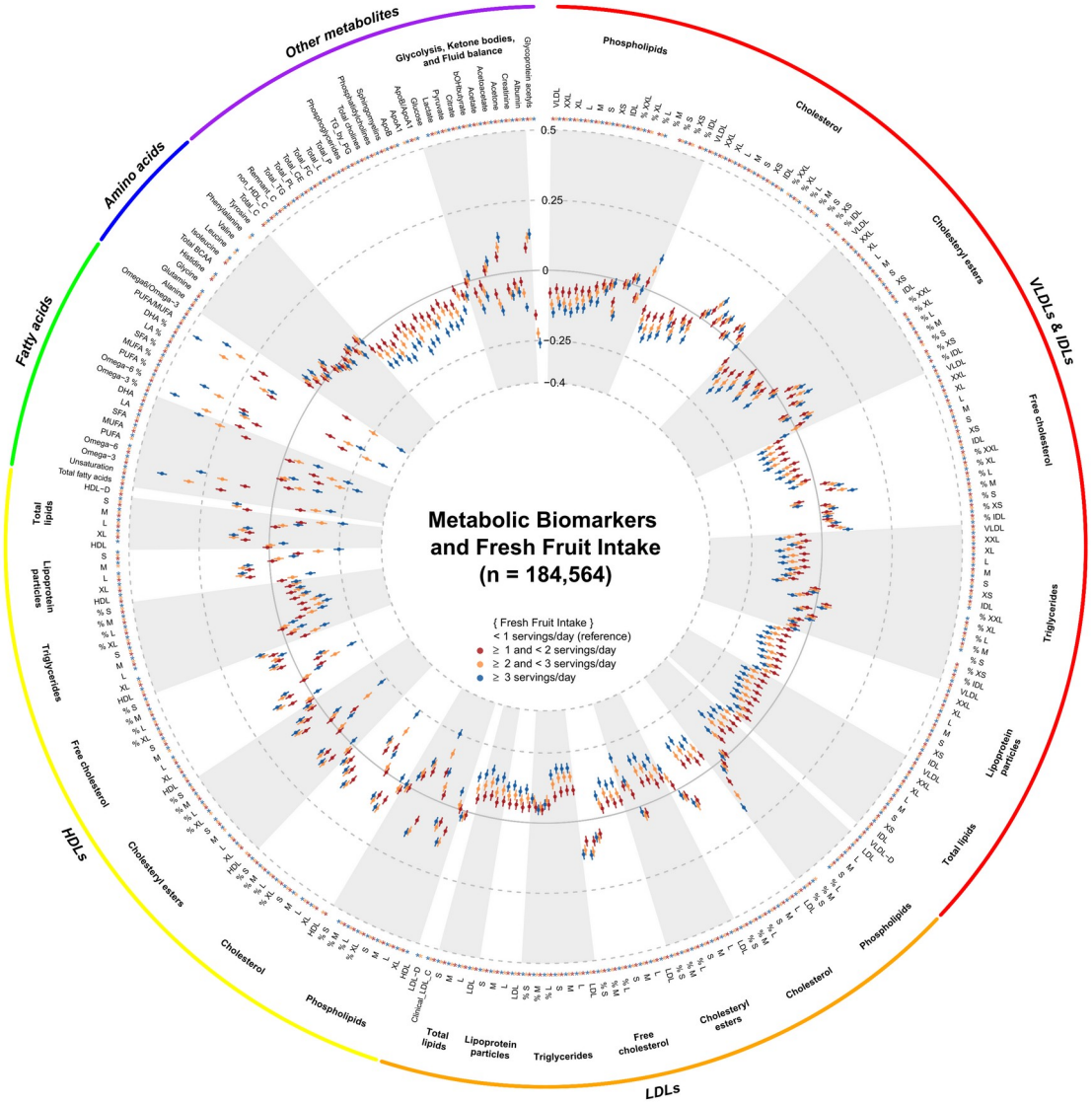
(G)


Metabolic Biomarkers
and Body Mass Index
(n = 184,564)
VLDLs & IDLs
LDLs
HDLs
Fatty acids
Amino acids
Other metabolites
Glycolysis, Ketone bodies, and Fluid balance
Phospholipids
Cholesterol
Cholesteryl esters
Free cholesterol
Triglycerides
Lipoprotein particles
Total lipids
0.08
0.04
0
−0.04
−0.08

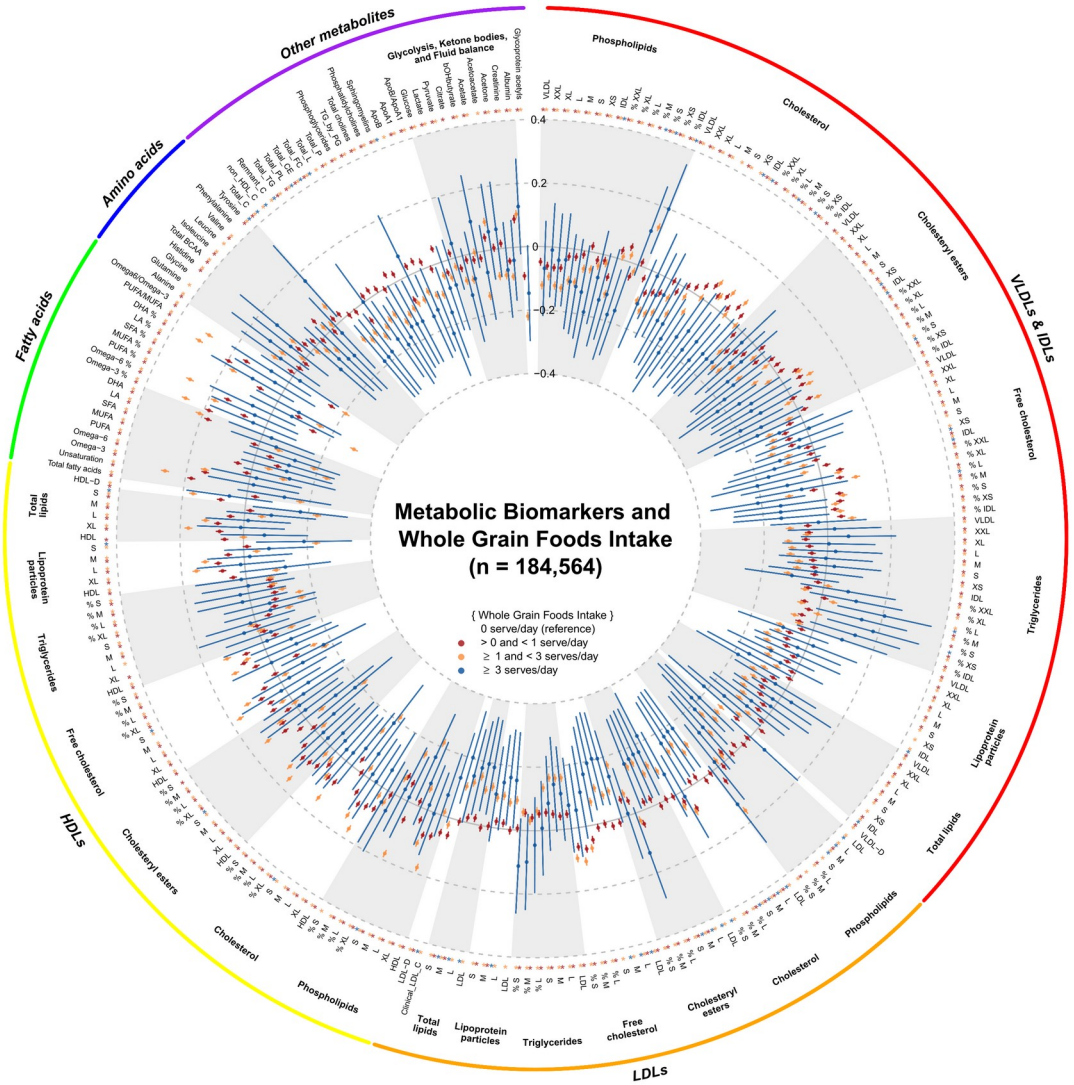
(H)


Metabolic Biomarkers
and Dietary Supplements
(n = 184,564)
Dietary Supplement Use
VLDLs & IDLs
LDLs
HDLs
Fatty acids
Amino acids
Other metabolites
Glycolysis, Ketone bodies, and Fluid balance
Phospholipids
Cholesterol
Cholesteryl esters
Free cholesterol
Triglycerides
Lipoprotein particles
Total lipids
0.4
0.2
0
−0.2
−0.4

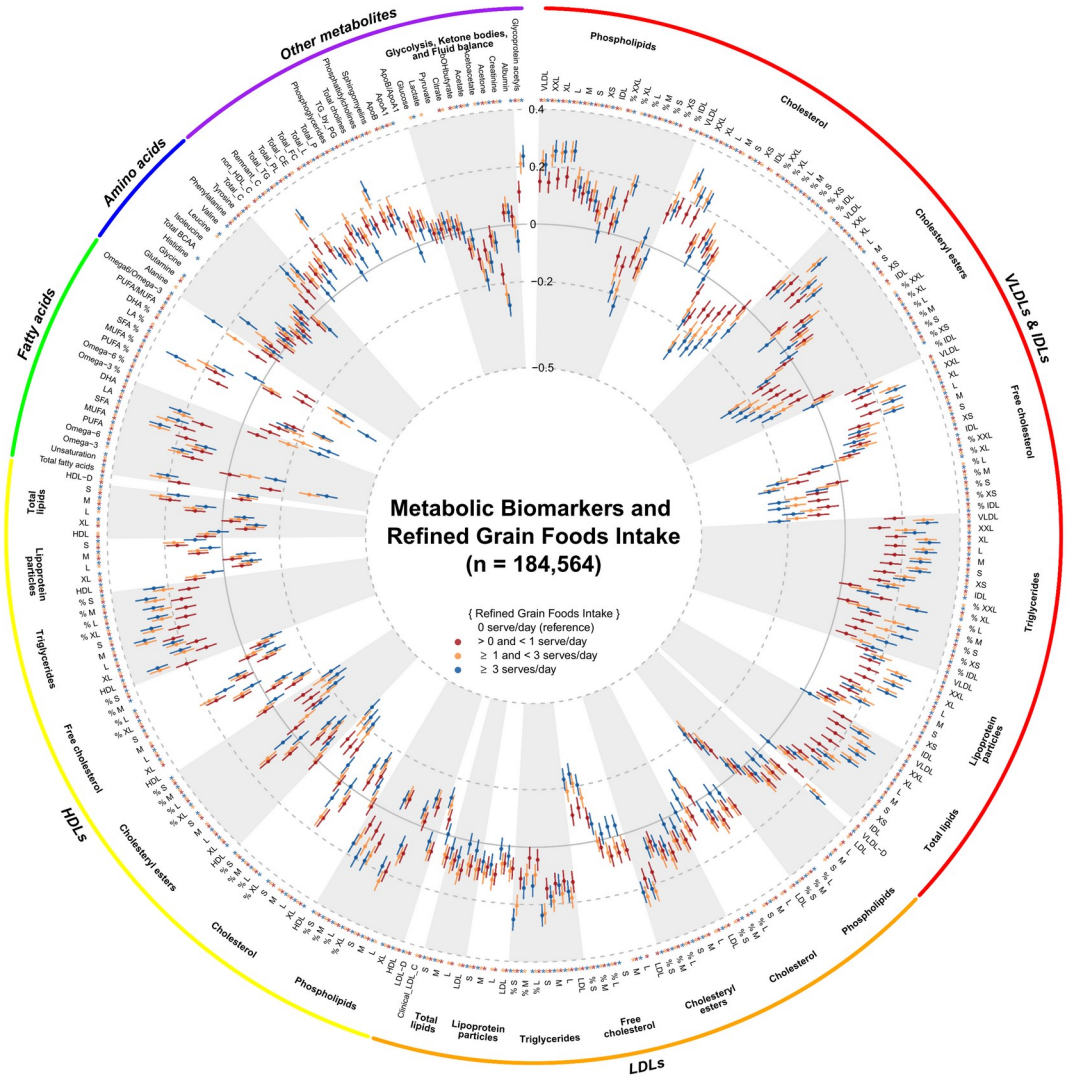
(I)


Metabolic Biomarkers
and Fresh Fruit Intake
(n = 184,564)
{ Fresh Fruit Intake }
< 1 servings/day (reference)
≥ 1 and < 2 servings/day
≥ 2 and < 3 servings/day
≥ 3 servings/day
0.5
0.25
0
−0.25
−0.4
VLDLs & IDLs
LDLs
HDLs
Fatty acids
Amino acids
Other metabolites
Phospholipids
Cholesterol
Cholesteryl esters
Free cholesterol
Triglycerides
Lipoprotein particles
Total lipids
Glycolysis, Ketone bodies, and Fluid balance

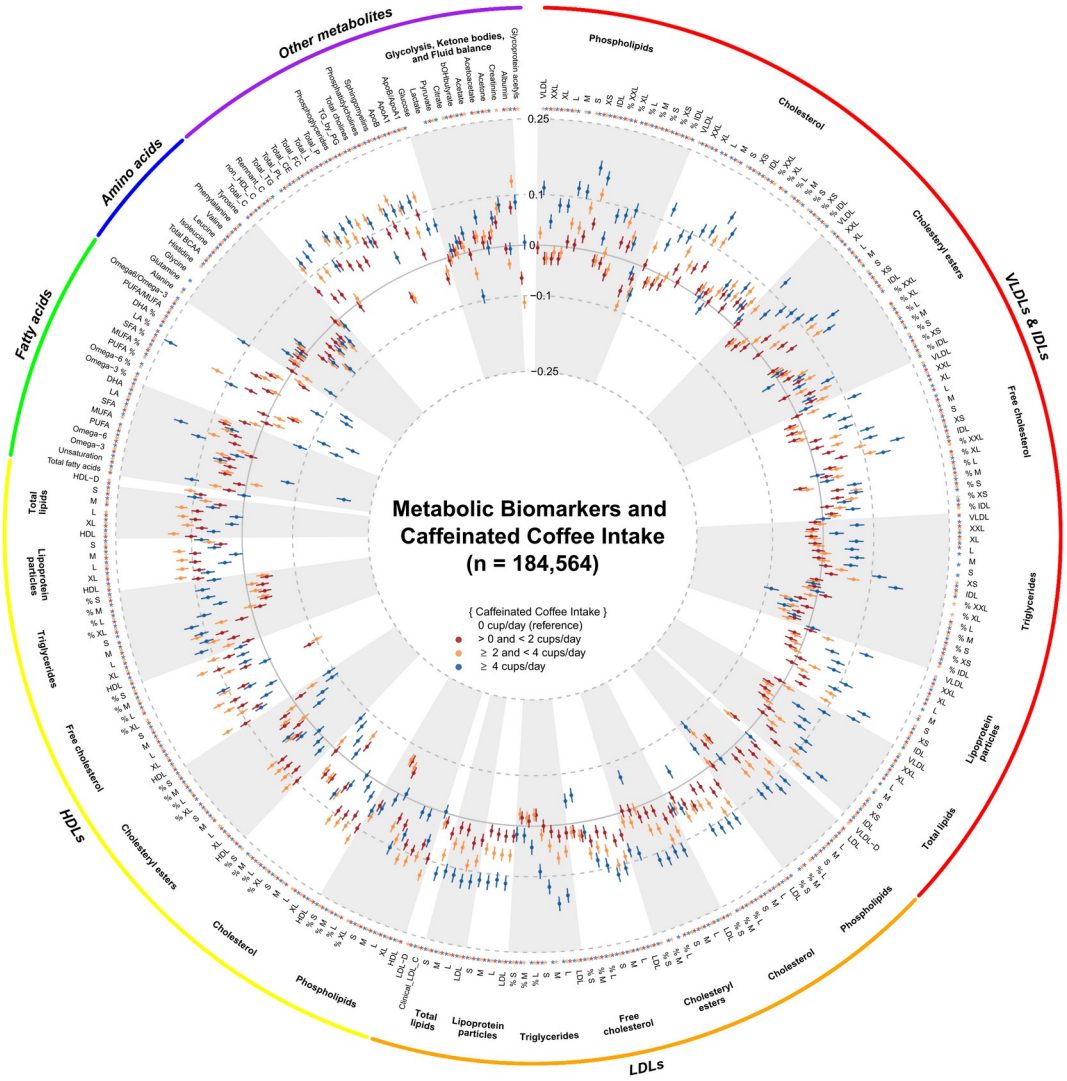
(J)


Metabolic Biomarkers and Whole Grain Foods Intake (n = 184,564)
{ Whole Grain Foods Intake }
0 serve/day (reference)
> 0 and < 1 serve/day
≥ 1 and < 3 serves/day
≥ 3 serves/day
VLDLs & IDLs
LDLs
HDLs
Fatty acids
Amino acids
Other metabolites
Phospholipids
Cholesterol
Cholesteryl esters
Free cholesterol
Triglycerides
Lipoprotein particles
Total lipids
Glycolysis, Ketone bodies, and Fluid balance
0.4
0.2
0
−0.2
−0.4

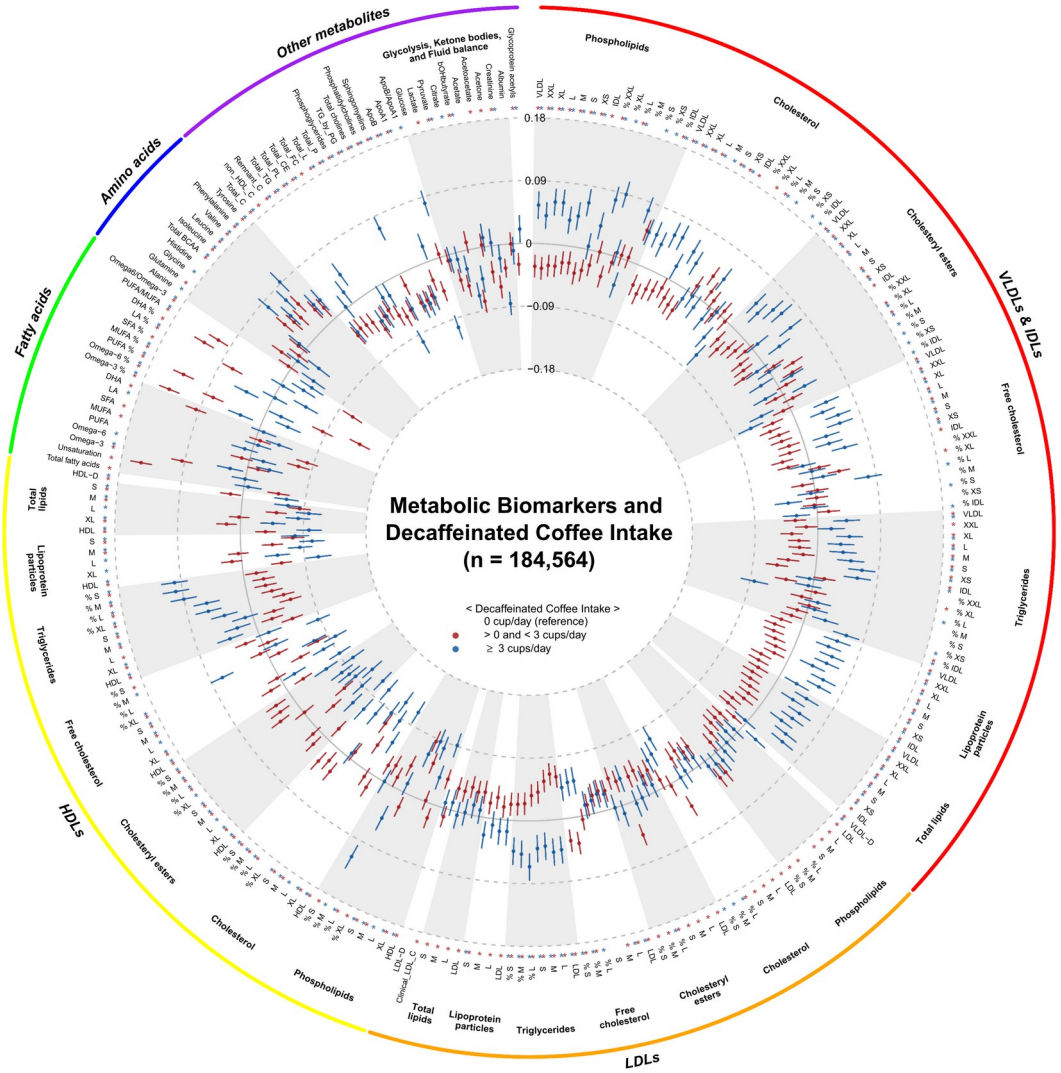
(K)


Metabolic Biomarkers and
Refined Grain Foods Intake
(n = 184,564)
{ Refined Grain Foods Intake }
0 serve/day (reference)
> 0 and < 1 serve/day
≥ 1 and < 3 serves/day
≥ 3 serves/day
Other metabolites
Glycolysis, Ketone bodies, and Fluid balance
Amino acids
Fatty acids
Phospholipids
Cholesterol
Cholesteryl esters
Free cholesterol
Triglycerides
Lipoprotein particles
Total lipids
VLDLs & IDLs
HDLs
LDLs

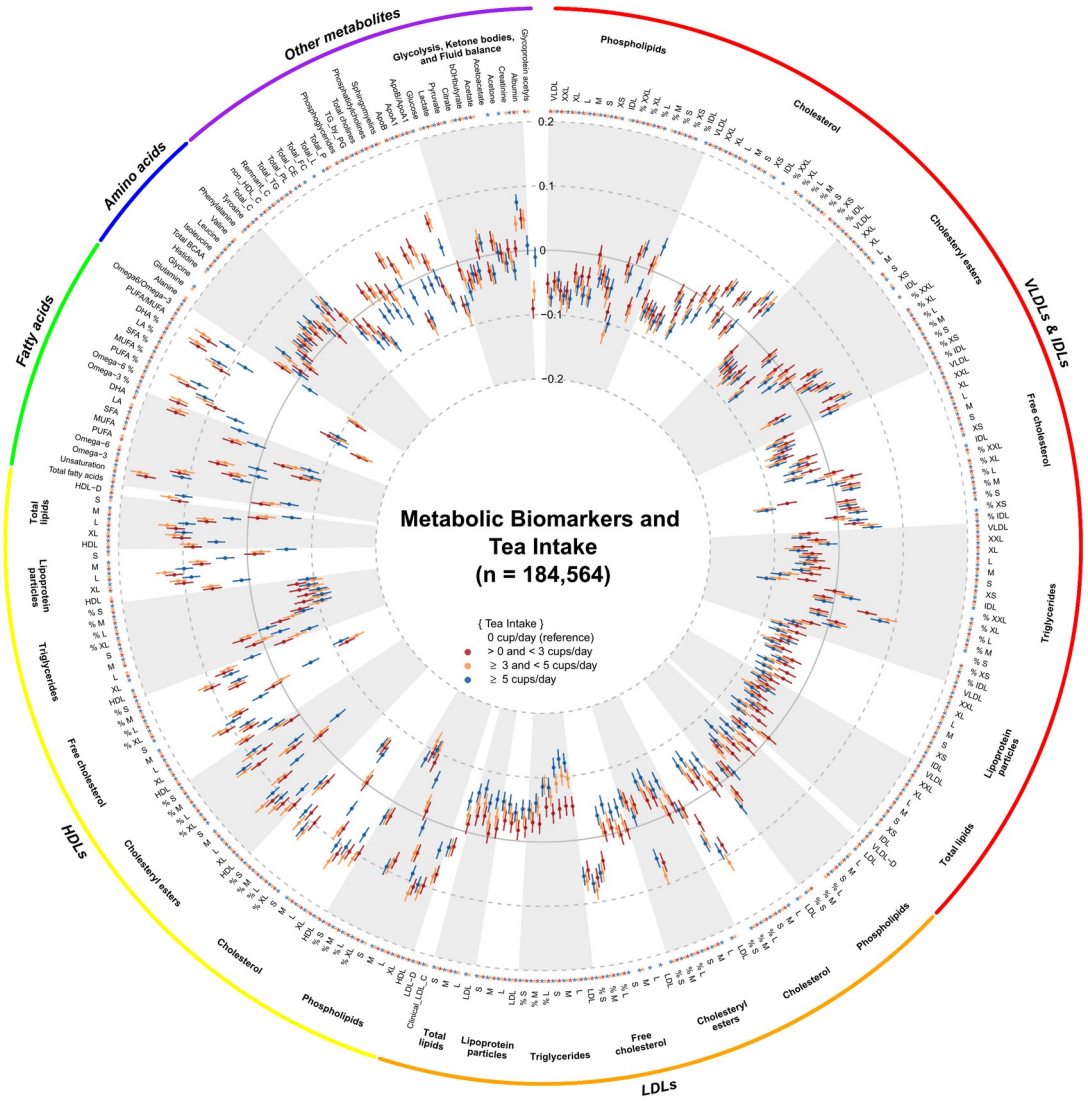
(L)


Metabolic Biomarkers and Caffeinated Coffee Intake (n = 184,564)
{ Caffeinated Coffee Intake }
0 cup/day (reference)
> 0 and < 2 cups/day
≥ 2 and < 4 cups/day
≥ 4 cups/day
VLDLs & IDLs
LDLs
HDLs
Fatty acids
Amino acids
Other metabolites
Phospholipids
Cholesterol
Cholesteryl esters
Free cholesterol
Triglycerides
Lipoprotein particles
Total lipids
Glycolysis, Ketone bodies, and Fluid balance

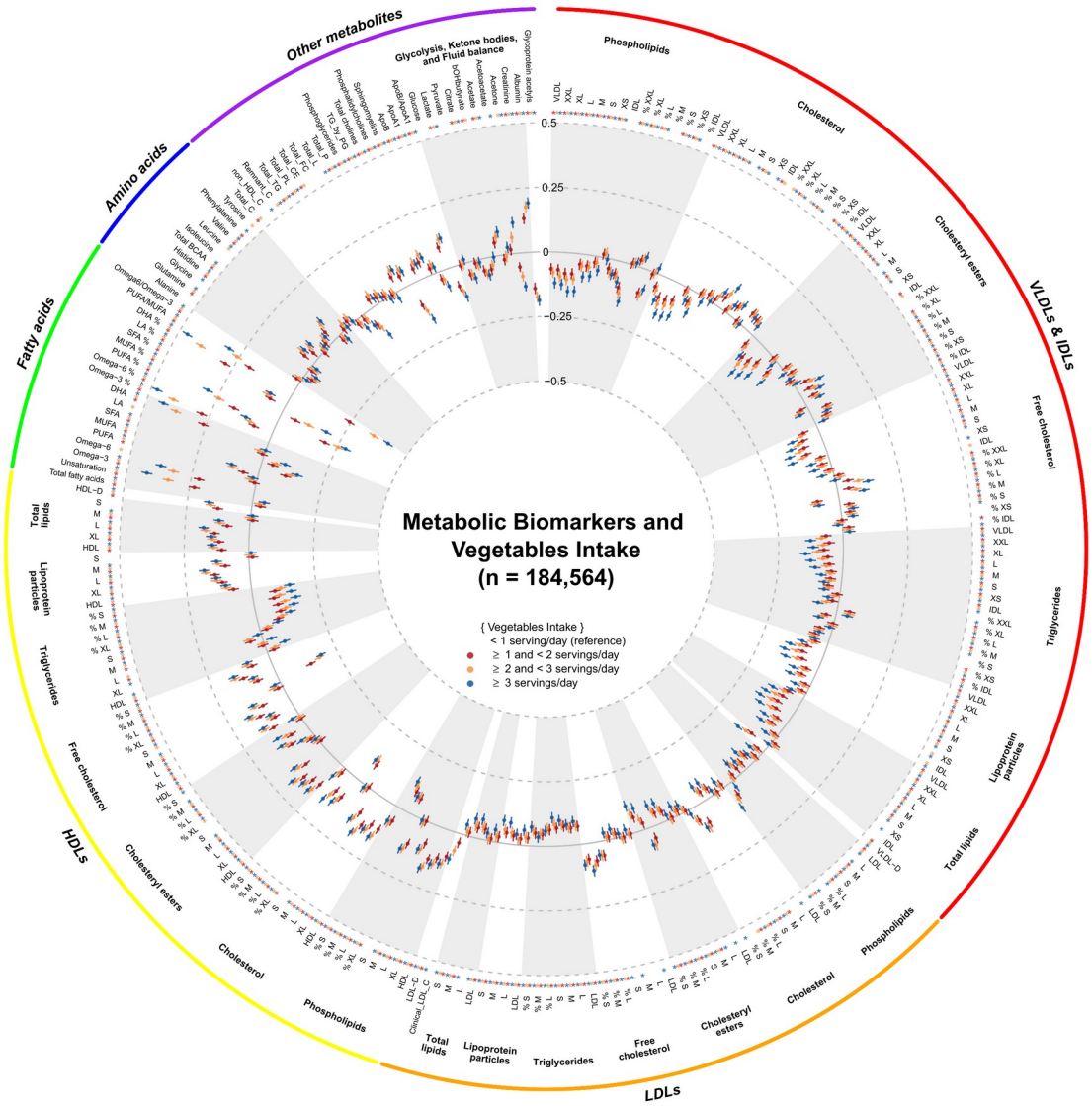
(M)


Metabolic Biomarkers and Decaffeinated Coffee Intake (n = 184,564)
< Decaffeinated Coffee Intake >
0 cup/day (reference)
> 0 and < 3 cups/day
≥ 3 cups/day
0.18
0.09
0
−0.09
−0.18
VLDLs & IDLs
LDLs
HDLs
Fatty acids
Amino acids
Other metabolites
Phospholipids
Cholesterol
Cholesteryl esters
Free cholesterol
Triglycerides
Lipoprotein particles
Total lipids
Glycolysis, Ketone bodies, and Fluid balance

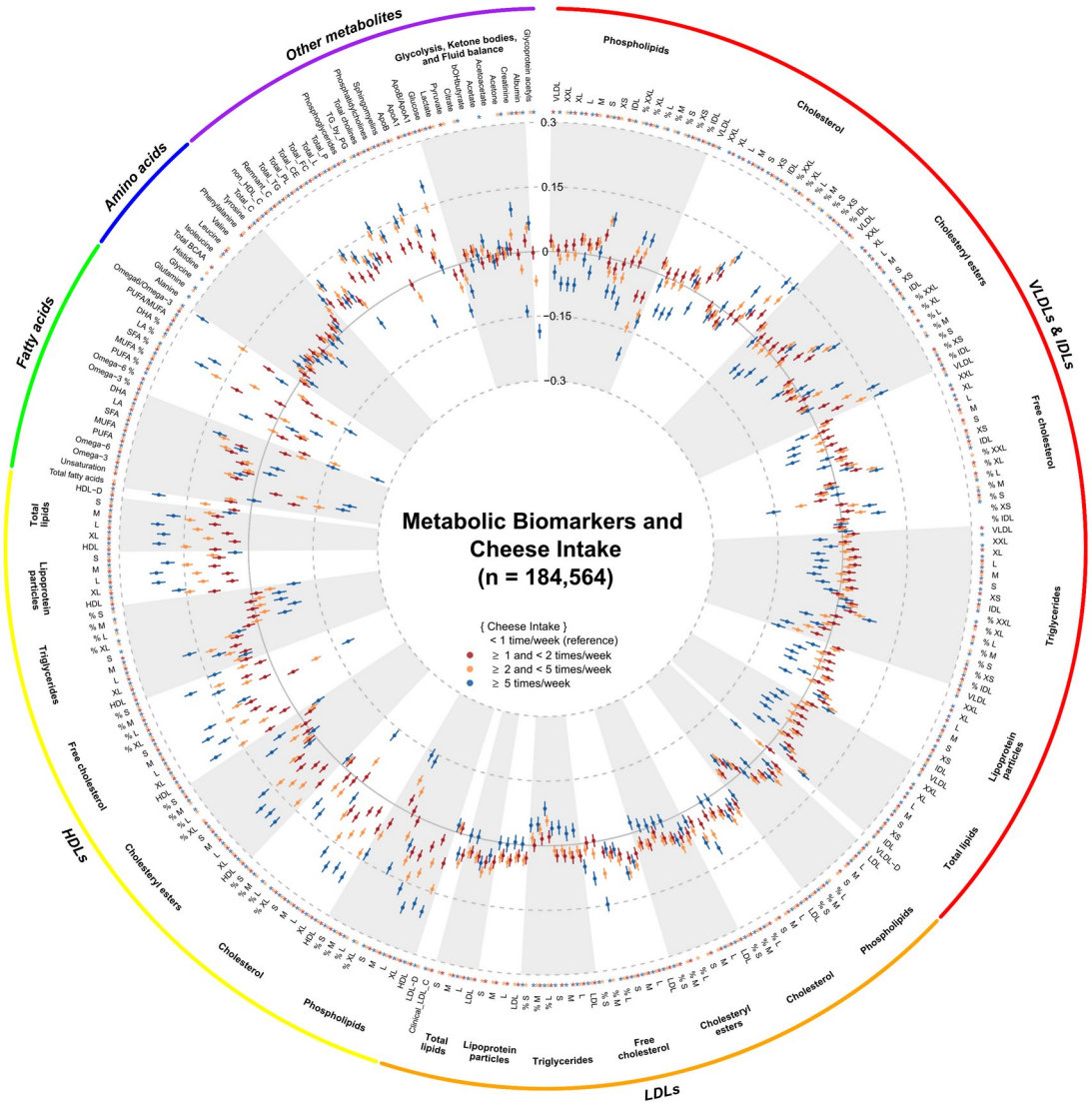
(N)


Metabolic Biomarkers and Tea Intake (n = 184,564)
{ Tea Intake }
0 cup/day (reference)
> 0 and < 3 cups/day
≥ 3 and < 5 cups/day
≥ 5 cups/day
Other metabolites
Amino acids
Fatty acids
VLDLs & IDLs
LDLs
HDLs
Phospholipids
Cholesterol
Cholesteryl esters
Free cholesterol
Triglycerides
Lipoprotein particles
Total lipids
Glycolysis, Ketone bodies, and Fluid balance

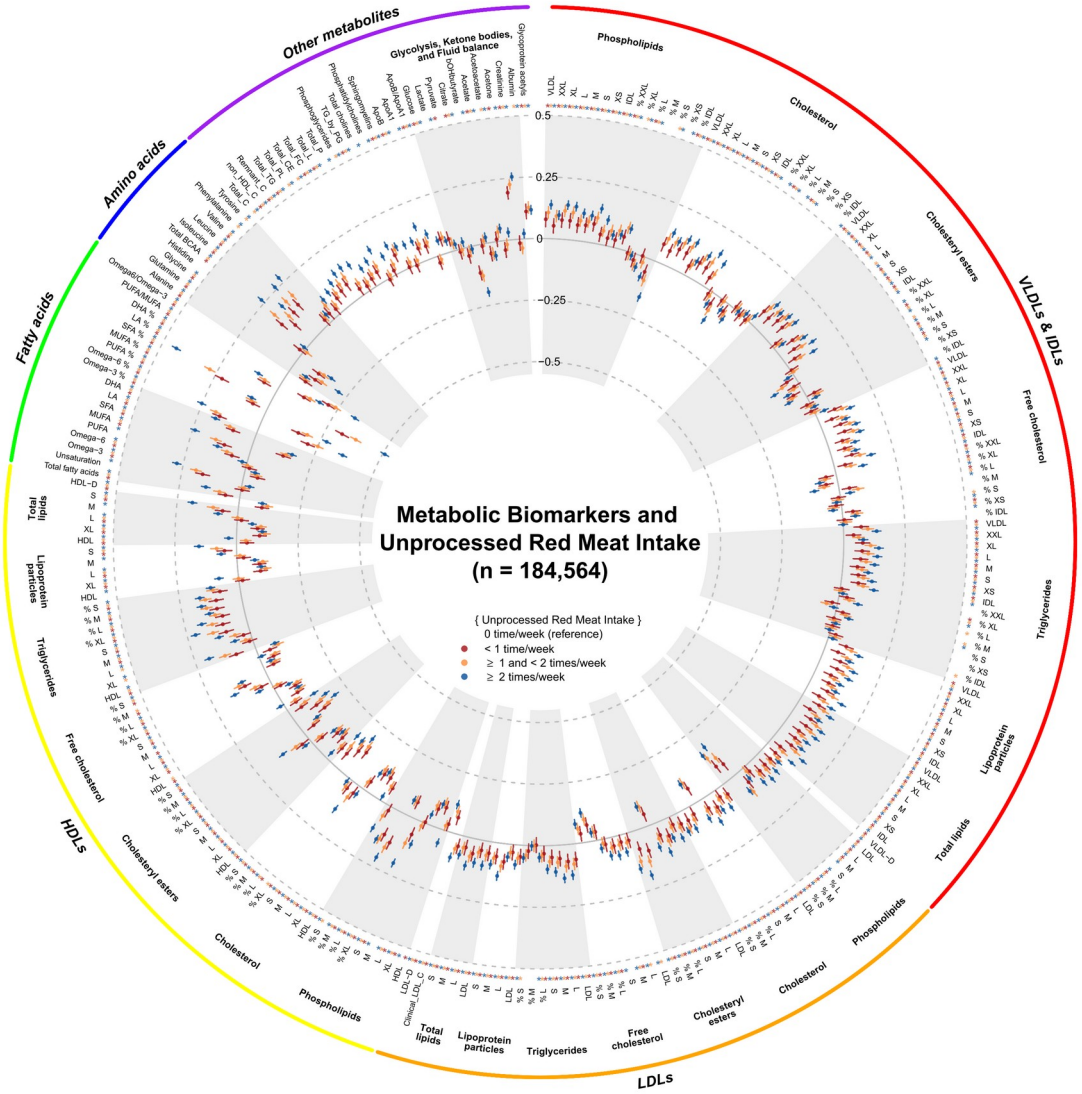
(O)


Metabolic Biomarkers and Vegetables Intake (n = 184,564)
{ Vegetables Intake }
< 1 serving/day (reference)
≥ 1 and < 2 servings/day
≥ 2 and < 3 servings/day
≥ 3 servings/day
VLDLs & IDLs
LDLs
HDLs
Fatty acids
Amino acids
Other metabolites
Phospholipids
Cholesterol
Cholesteryl esters
Free cholesterol
Triglycerides
Lipoprotein particles
Total lipids
Glycolysis, Ketone bodies, and Fluid balance

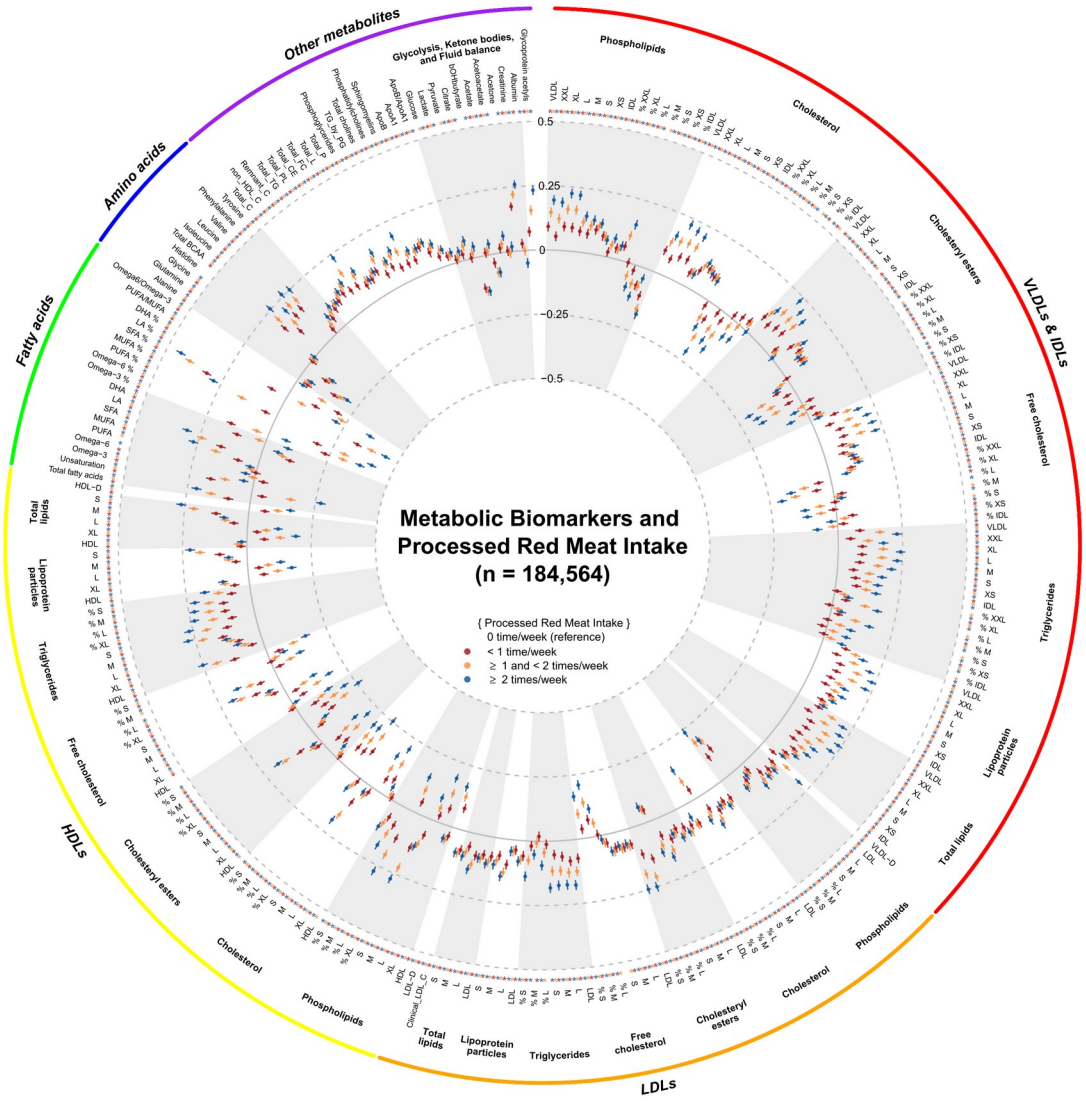
(P)


Metabolic Biomarkers and Cheese Intake (n = 184,564)
{ Cheese Intake }
< 1 time/week (reference)
≥ 1 and < 2 times/week
≥ 2 and < 5 times/week
≥ 5 times/week
VLDLs & IDLs
LDLs
HDLs
Fatty acids
Amino acids
Other metabolites
Phospholipids
Cholesterol
Cholesteryl esters
Free cholesterol
Triglycerides
Lipoprotein particles
Total lipids
Glycolysis, Ketone bodies, and Fluid balance

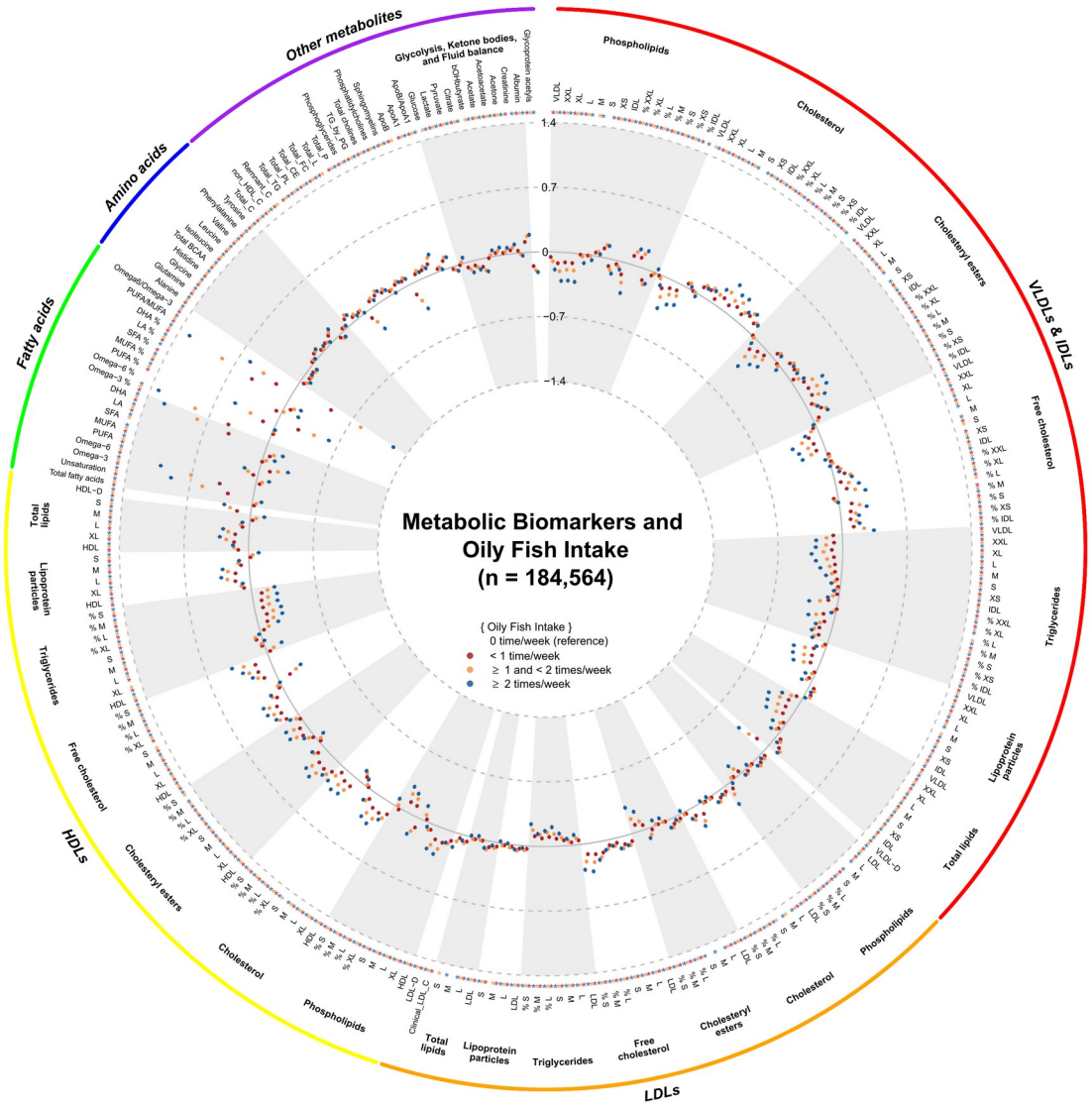
(Q)


Metabolic Biomarkers and
Unprocessed Red Meat Intake
(n = 184,564)
{ Unprocessed Red Meat Intake }
0 time/week (reference)
< 1 time/week
≥ 1 and < 2 times/week
≥ 2 times/week
VLDLs & IDLs
LDLs
HDLs
Fatty acids
Amino acids
Other metabolites
Glycolysis, Ketone bodies, and Fluid balance
Phospholipids
Cholesterol
Cholesteryl esters
Free cholesterol
Triglycerides
Lipoprotein particles
Total lipids
0.5
0.25
0
−0.25
−0.5

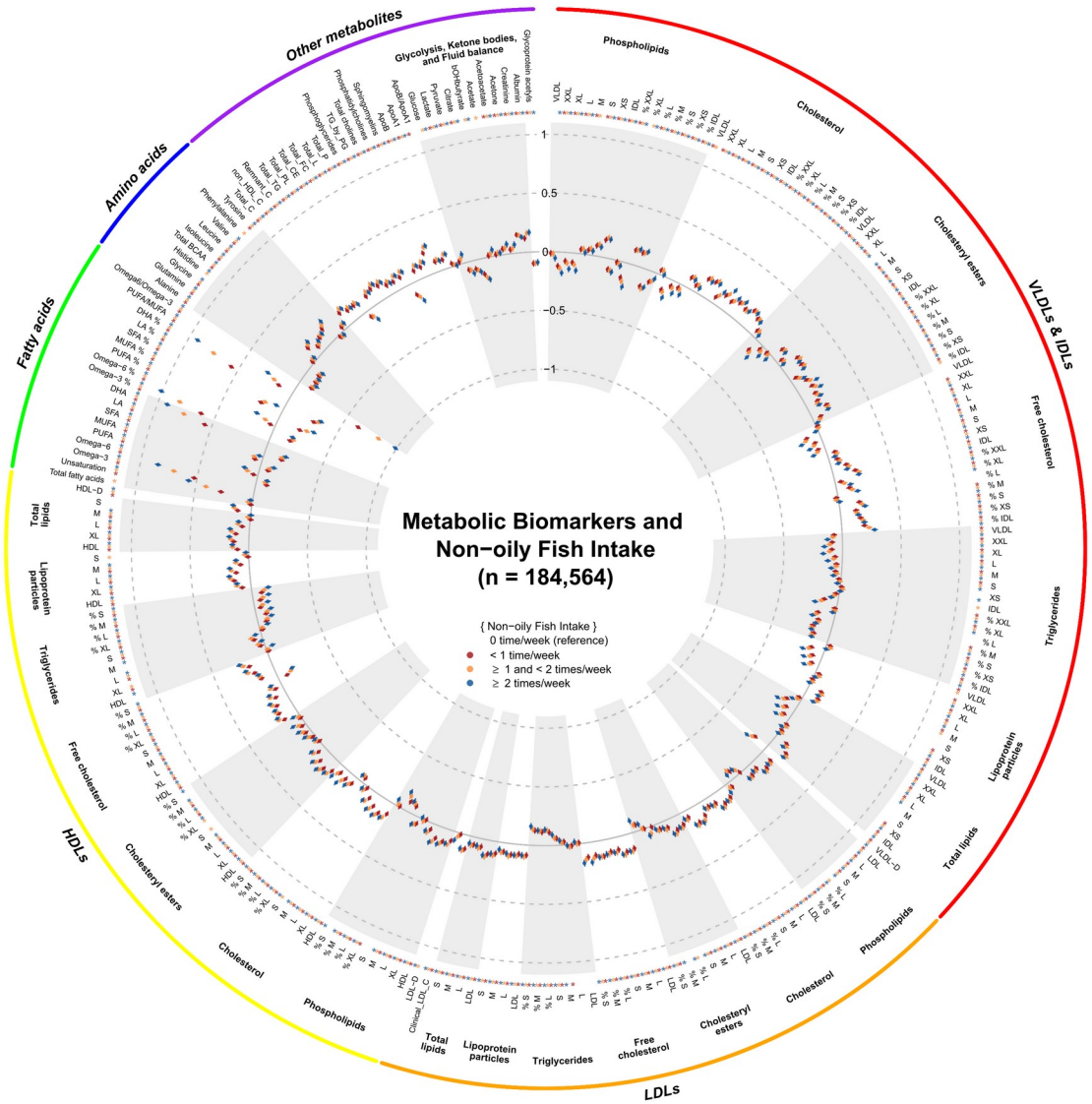
(R)


Metabolic Biomarkers and
Processed Red Meat Intake
(n = 184,564)
{ Processed Red Meat Intake }
0 time/week (reference)
< 1 time/week
≥ 1 and < 2 times/week
≥ 2 times/week
0.5
0.25
0
−0.25
−0.5
VLDLs & IDLs
LDLs
HDLs
Fatty acids
Amino acids
Other metabolites
Phospholipids
Cholesterol
Cholesteryl esters
Free cholesterol
Triglycerides
Lipoprotein particles
Total lipids
Glycolysis, Ketone bodies, and Fluid balance

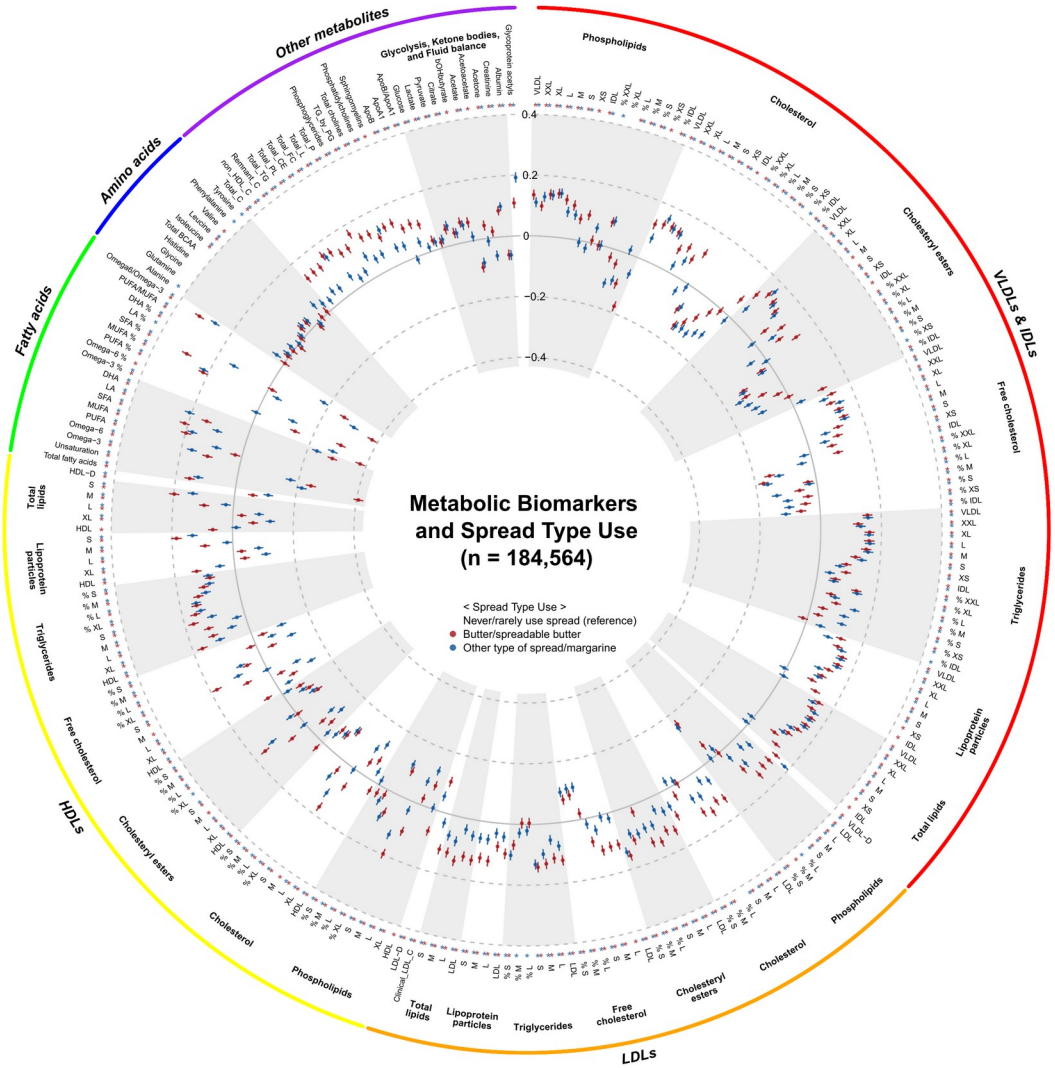
(S)


Metabolic Biomarkers and
Oily Fish Intake
(n = 184,564)
{ Oily Fish Intake }
0 time/week (reference)
< 1 time/week
≥ 1 and < 2 times/week
≥ 2 times/week
VLDLs & IDLs
LDLs
HDLs
Fatty acids
Amino acids
Other metabolites
Glycolysis, Ketone bodies, and Fluid balance
Phospholipids
Cholesterol
Cholesteryl esters
Free cholesterol
Triglycerides
Lipoprotein particles
Total lipids

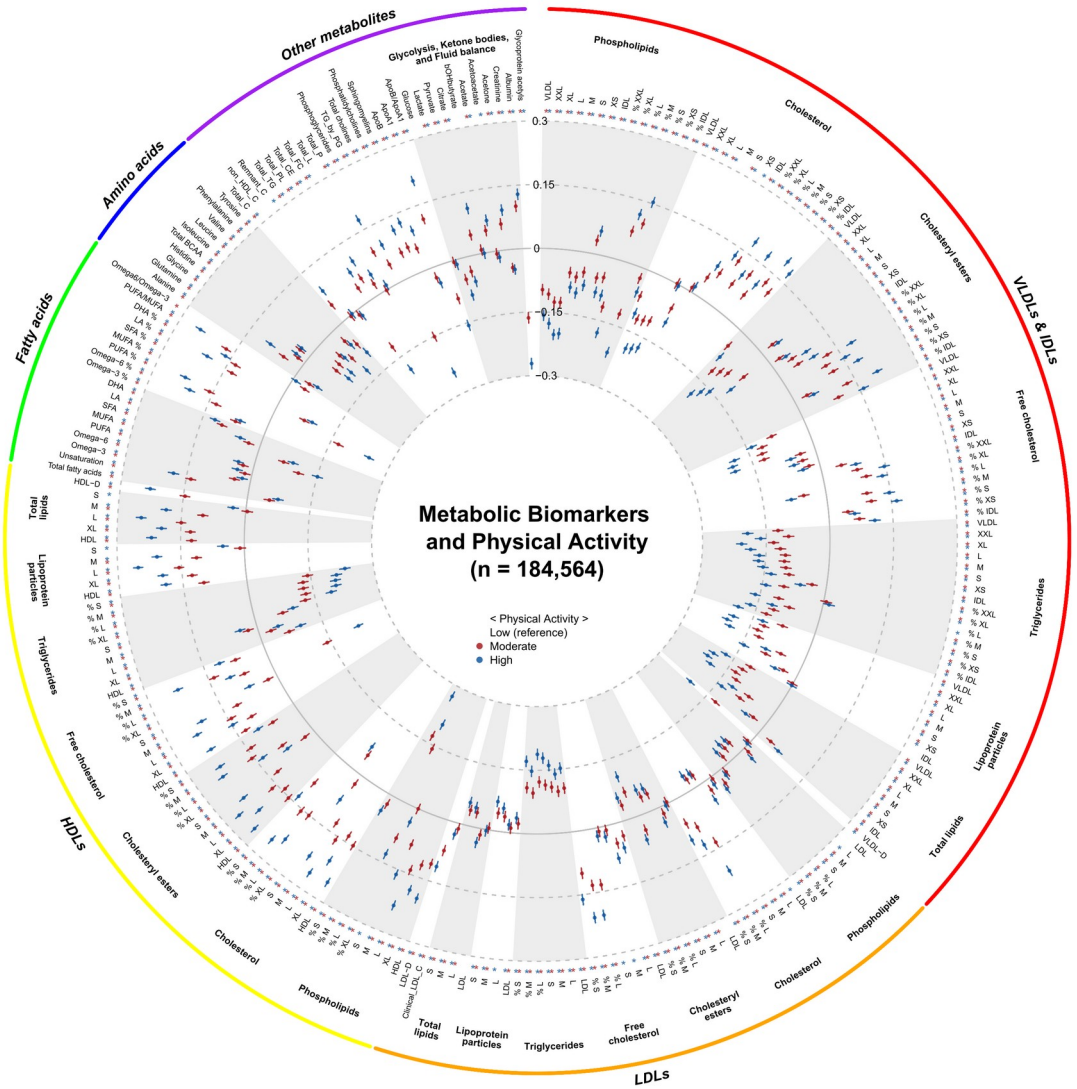
(T)


Metabolic Biomarkers and
Non−oily Fish Intake
(n = 184,564)
{ Non−oily Fish Intake }
0 time/week (reference)
< 1 time/week
≥ 1 and < 2 times/week
≥ 2 times/week
VLDLs & IDLs
LDLs
HDLs
Fatty acids
Amino acids
Other metabolites
Glycolysis, Ketone bodies, and Fluid balance
Phospholipids
Cholesterol
Cholesteryl esters
Free cholesterol
Triglycerides
Lipoprotein particles
Total lipids

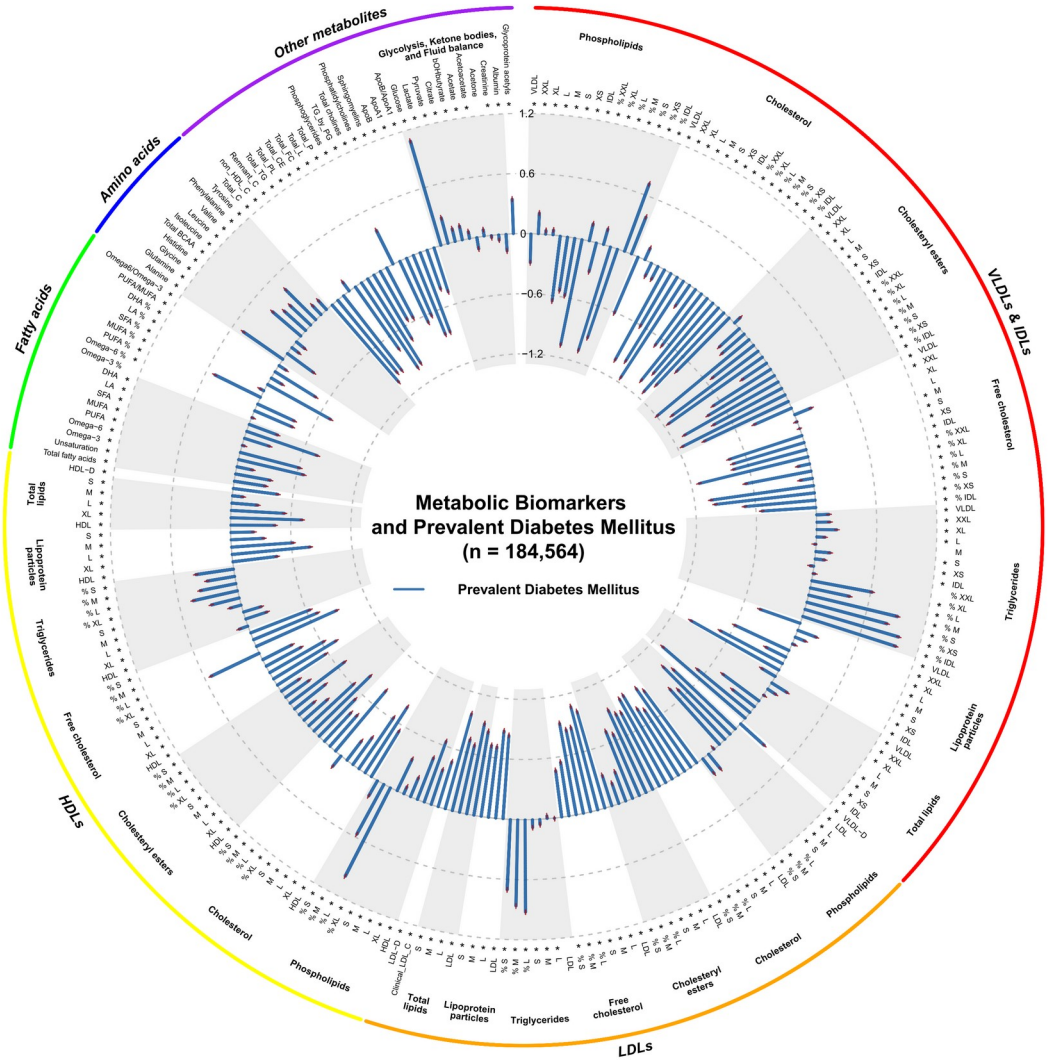
(U)


Metabolic Biomarkers and Spread Type Use (n = 184,564)
< Spread Type Use >
Never/rarely use spread (reference)
Butter/spreadable butter
Other type of spread/margarine
VLDLs & IDLs
LDLs
HDLs
Fatty acids
Amino acids
Other metabolites
Glycolysis, Ketone bodies, and Fluid balance
Phospholipids
Cholesterol
Cholesteryl esters
Free cholesterol
Triglycerides
Lipoprotein particles
Total lipids

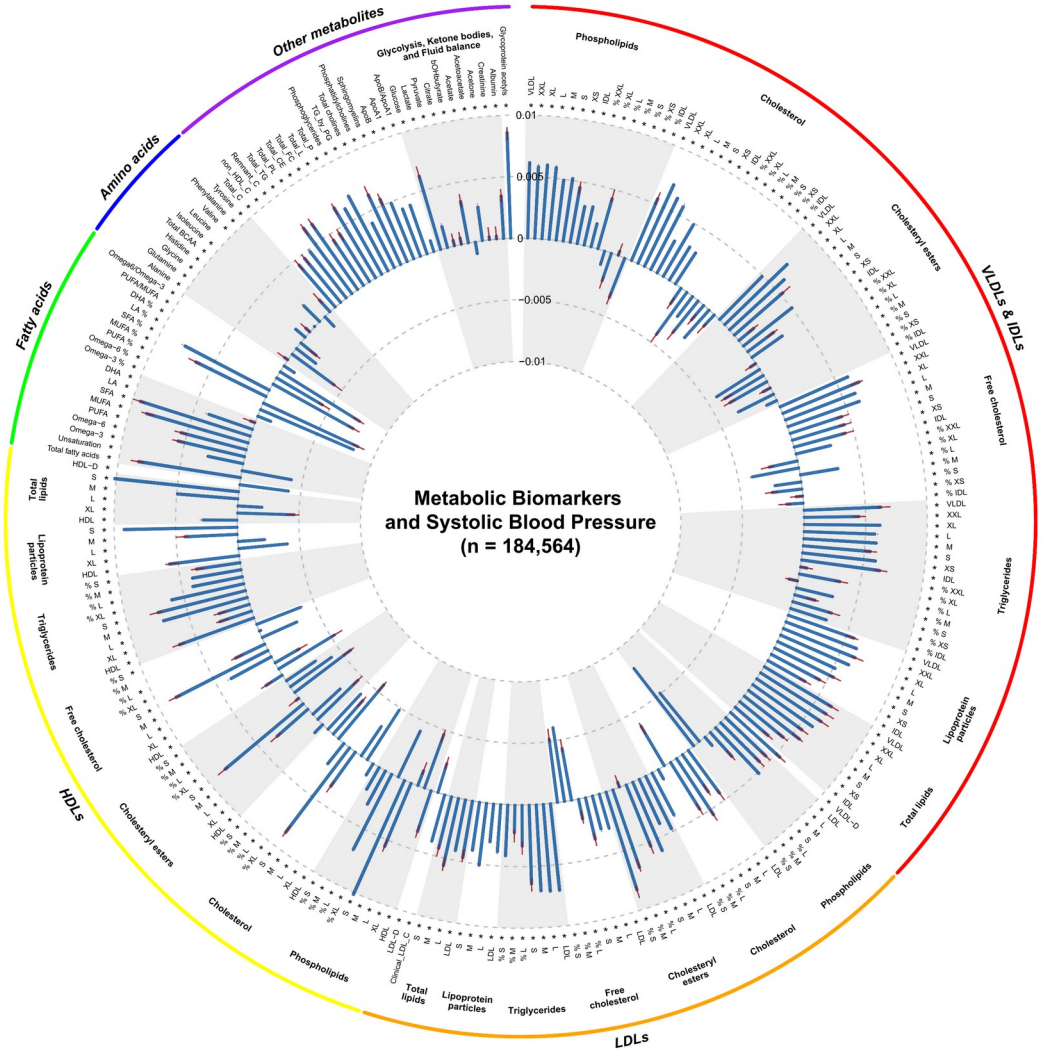
(V)


Metabolic Biomarkers
and Physical Activity
(n = 184,564)
< Physical Activity >
Low (reference)
Moderate
High
VLDLs & IDLs
LDLs
HDLs
Fatty acids
Amino acids
Other metabolites
Phospholipids
Cholesterol
Cholesteryl esters
Free cholesterol
Triglycerides
Lipoprotein particles
Total lipids
Glycolysis, Ketone bodies, and Fluid balance
0.3
0.15
0
-0.15
-0.3

(W)


Metabolic Biomarkers
and Prevalent Diabetes Mellitus
(n = 184,564)
Prevalent Diabetes Mellitus
VLDLs & IDLs
LDLs
HDLs
Fatty acids
Amino acids
Other metabolites
Phospholipids
Cholesterol
Cholesteryl esters
Free cholesterol
Triglycerides
Lipoprotein particles
Total lipids
Glycolysis, Ketone bodies, and Fluid balance
1.2
0.6
0
−0.6
−1.2

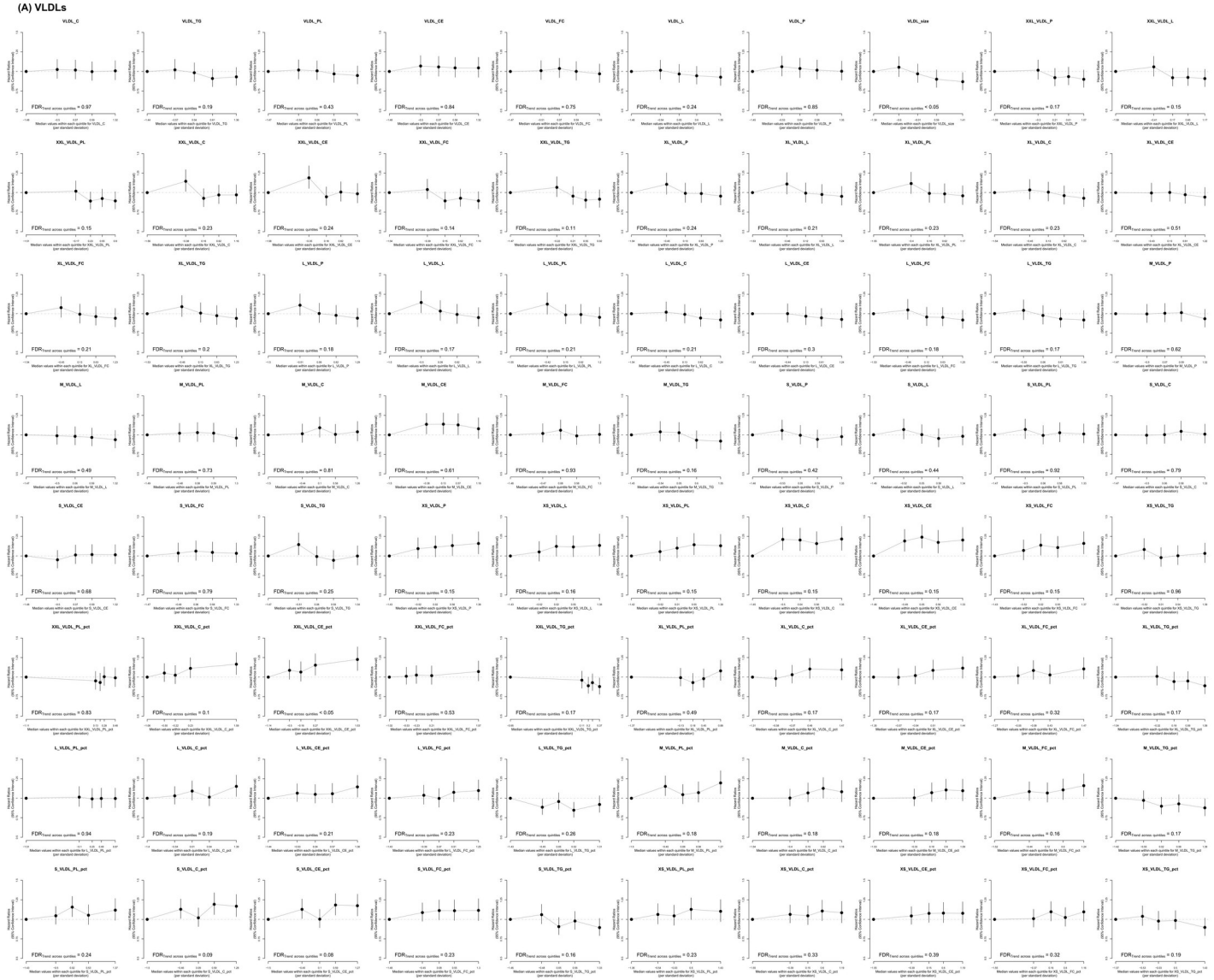
(X)



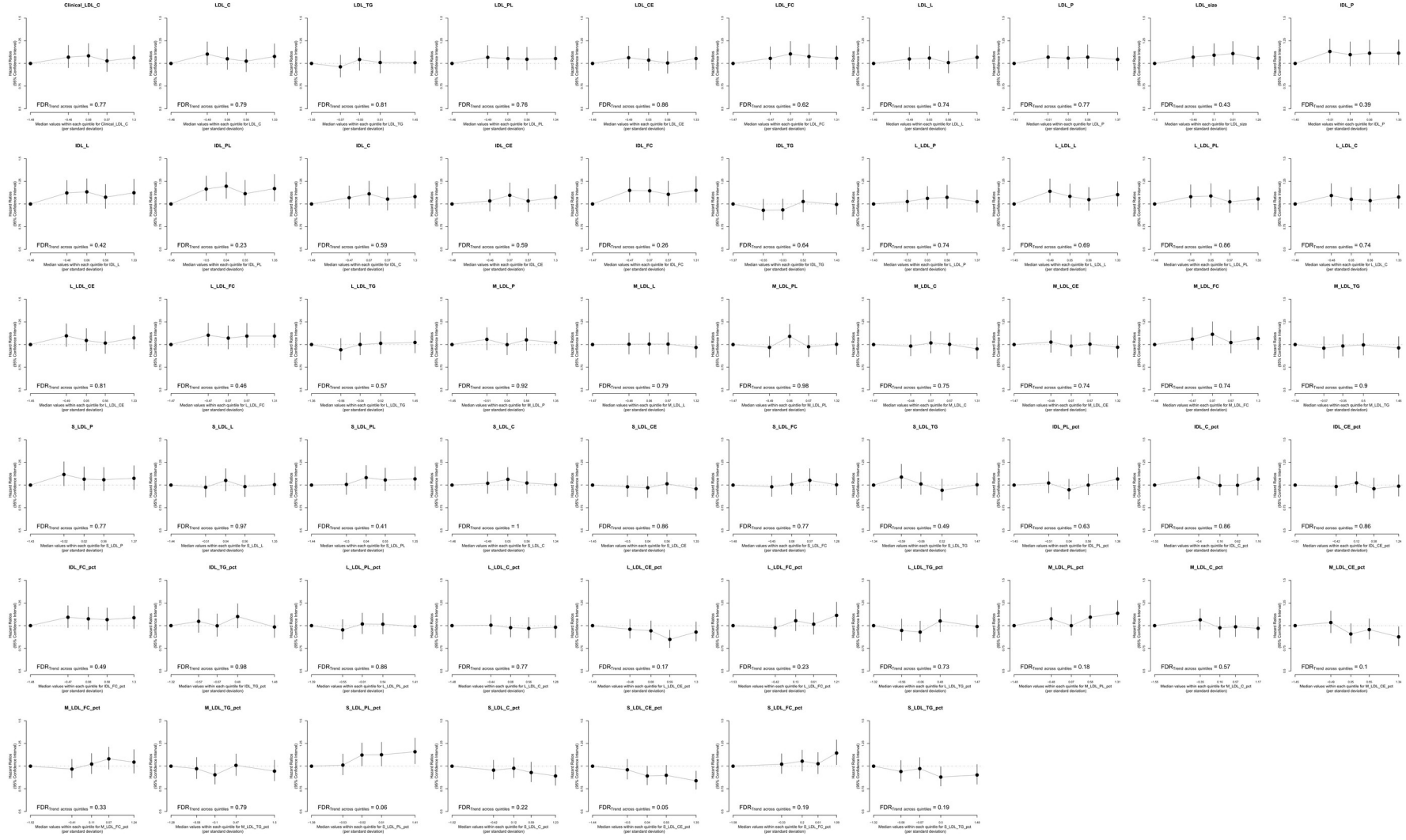
Supplementary Figure 3.2 Associations of each NMR metabolite with selected baseline characteristics

The linear regression models are adjusted for age and sex (if applicable) to evaluate associations of each NMR metabolite with selected baseline characteristics ((A) sex (male), (B) ethnicity, (C) lipid modifying medication, (D) fasting time, (E) body mass index, (F) smoking status, (G) alcohol intake, (H)-(U) dietary supplement use and 13 dietary variables, (V) physical activity, (W) prevalent diabetes mellitus, and (X) systolic blood pressure). The beta coefficients (with 95% confidence intervals) are presented per one standard deviation higher metabolic biomarker on the natural log scale. Stars

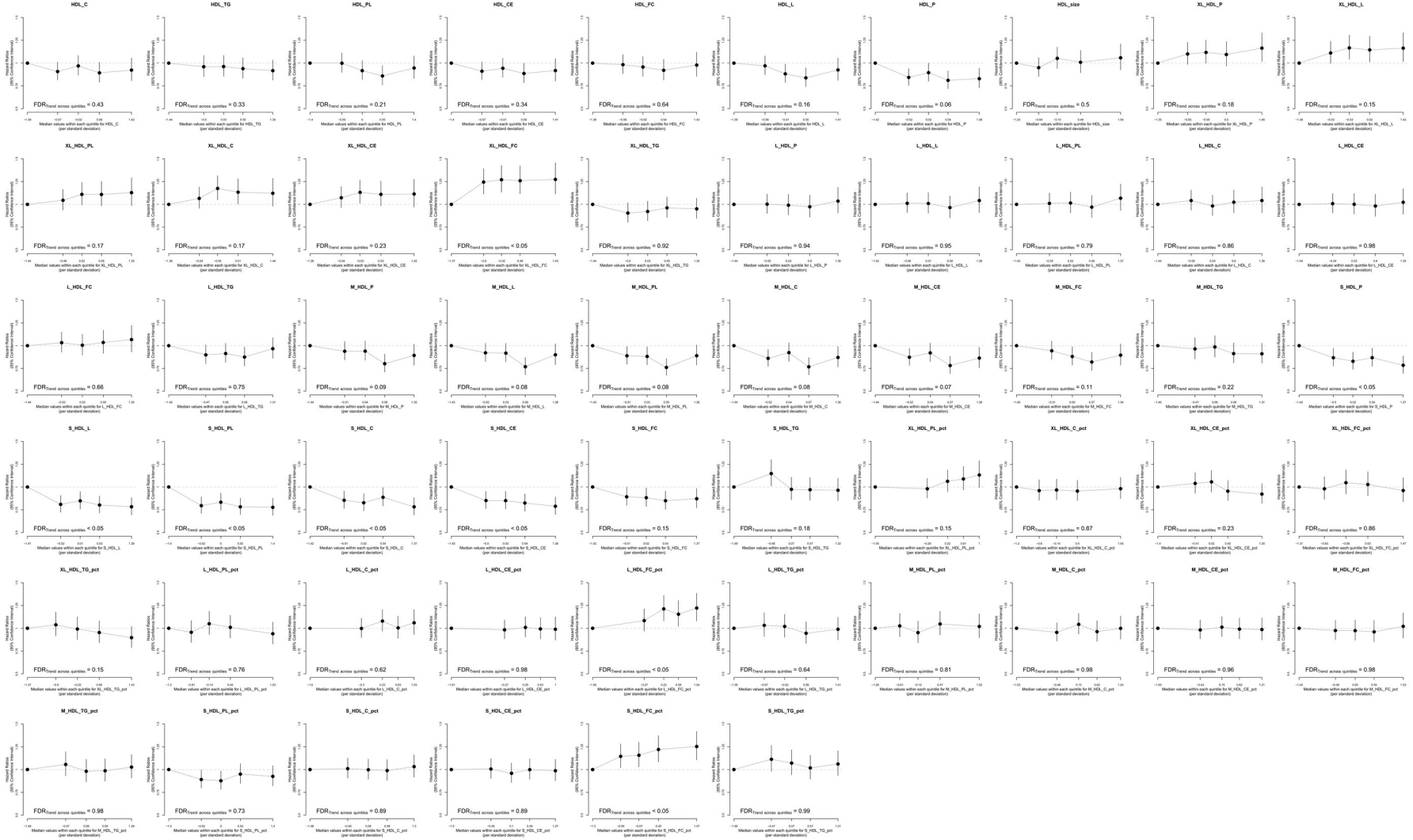
indicate false discovery rate < 0.05. Abbreviations are in line with **Supplementary Table 3.1**.


(A) VLDLs
VLDL_C
FDRTrend across quintiles = 0.97
VLDL_TG
FDRTrend across quintiles = 0.19
VLDL_PL
FDRTrend across quintiles = 0.43
VLDL_CE
FDRTrend across quintiles = 0.84
VLDL_FC
FDRTrend across quintiles = 0.75
VLDL_L
FDRTrend across quintiles = 0.24
VLDL_P
FDRTrend across quintiles = 0.85
VLDL_size
FDRTrend across quintiles < 0.05
XXL_VLDL_P
FDRTrend across quintiles = 0.17
XXL_VLDL_L
FDRTrend across quintiles = 0.15
XXL_VLDL_PL
FDRTrend across quintiles = 0.15
XXL_VLDL_C
FDRTrend across quintiles = 0.23
XXL_VLDL_CE
FDRTrend across quintiles = 0.24
XXL_VLDL_FC
FDRTrend across quintiles = 0.14
XXL_VLDL_TG
FDRTrend across quintiles = 0.11
XL_VLDL_P
FDRTrend across quintiles = 0.24
XL_VLDL_L
FDRTrend across quintiles = 0.21
XL_VLDL_PL
FDRTrend across quintiles = 0.23
XL_VLDL_C
FDRTrend across quintiles = 0.23
XL_VLDL_CE
FDRTrend across quintiles = 0.51
XL_VLDL_FC
FDRTrend across quintiles = 0.21
XL_VLDL_TG
FDRTrend across quintiles = 0.2
L_VLDL_P
FDRTrend across quintiles = 0.18
L_VLDL_L
FDRTrend across quintiles = 0.17
L_VLDL_PL
FDRTrend across quintiles = 0.21
L_VLDL_C
FDRTrend across quintiles = 0.21
L_VLDL_CE
FDRTrend across quintiles = 0.3
L_VLDL_FC
FDRTrend across quintiles = 0.18
L_VLDL_TG
FDRTrend across quintiles = 0.17
M_VLDL_P
FDRTrend across quintiles = 0.62
M_VLDL_L
FDRTrend across quintiles = 0.49
M_VLDL_PL
FDRTrend across quintiles = 0.73
M_VLDL_C
FDRTrend across quintiles = 0.81
M_VLDL_CE
FDRTrend across quintiles = 0.61
M_VLDL_FC
FDRTrend across quintiles = 0.93
M_VLDL_TG
FDRTrend across quintiles = 0.16
S_VLDL_P
FDRTrend across quintiles = 0.42
S_VLDL_L
FDRTrend across quintiles = 0.44
S_VLDL_PL
FDRTrend across quintiles = 0.92
S_VLDL_C
FDRTrend across quintiles = 0.79
S_VLDL_CE
FDRTrend across quintiles = 0.66
S_VLDL_FC
FDRTrend across quintiles = 0.79
S_VLDL_TG
FDRTrend across quintiles = 0.25
XS_VLDL_P
FDRTrend across quintiles = 0.15
XS_VLDL_L
FDRTrend across quintiles = 0.16
XS_VLDL_PL
FDRTrend across quintiles = 0.15
XS_VLDL_C
FDRTrend across quintiles = 0.15
XS_VLDL_CE
FDRTrend across quintiles = 0.15
XS_VLDL_FC
FDRTrend across quintiles = 0.15
XS_VLDL_TG
FDRTrend across quintiles = 0.96
XXL_VLDL_PL_pct
FDRTrend across quintiles = 0.83
XXL_VLDL_C_pct
FDRTrend across quintiles = 0.1
XXL_VLDL_CE_pct
FDRTrend across quintiles < 0.05
XXL_VLDL_FC_pct
FDRTrend across quintiles = 0.53
XXL_VLDL_TG_pct
FDRTrend across quintiles = 0.17
XL_VLDL_PL_pct
FDRTrend across quintiles = 0.49
XL_VLDL_C_pct
FDRTrend across quintiles = 0.17
XL_VLDL_CE_pct
FDRTrend across quintiles = 0.17
XL_VLDL_FC_pct
FDRTrend across quintiles = 0.32
XL_VLDL_TG_pct
FDRTrend across quintiles = 0.17
L_VLDL_PL_pct
FDRTrend across quintiles = 0.94
L_VLDL_C_pct
FDRTrend across quintiles = 0.19
L_VLDL_CE_pct
FDRTrend across quintiles = 0.21
L_VLDL_FC_pct
FDRTrend across quintiles = 0.23
L_VLDL_TG_pct
FDRTrend across quintiles = 0.26
M_VLDL_PL_pct
FDRTrend across quintiles = 0.18
M_VLDL_C_pct
FDRTrend across quintiles = 0.18
M_VLDL_CE_pct
FDRTrend across quintiles = 0.18
M_VLDL_FC_pct
FDRTrend across quintiles = 0.16
M_VLDL_TG_pct
FDRTrend across quintiles = 0.17
S_VLDL_PL_pct
FDRTrend across quintiles = 0.24
S_VLDL_C_pct
FDRTrend across quintiles = 0.09
S_VLDL_CE_pct
FDRTrend across quintiles = 0.06
S_VLDL_FC_pct
FDRTrend across quintiles = 0.23
S_VLDL_TG_pct
FDRTrend across quintiles = 0.16
XS_VLDL_PL_pct
FDRTrend across quintiles = 0.23
XS_VLDL_C_pct
FDRTrend across quintiles = 0.33
XS_VLDL_CE_pct
FDRTrend across quintiles = 0.39
XS_VLDL_FC_pct
FDRTrend across quintiles = 0.32
XS_VLDL_TG_pct
FDRTrend across quintiles = 0.19

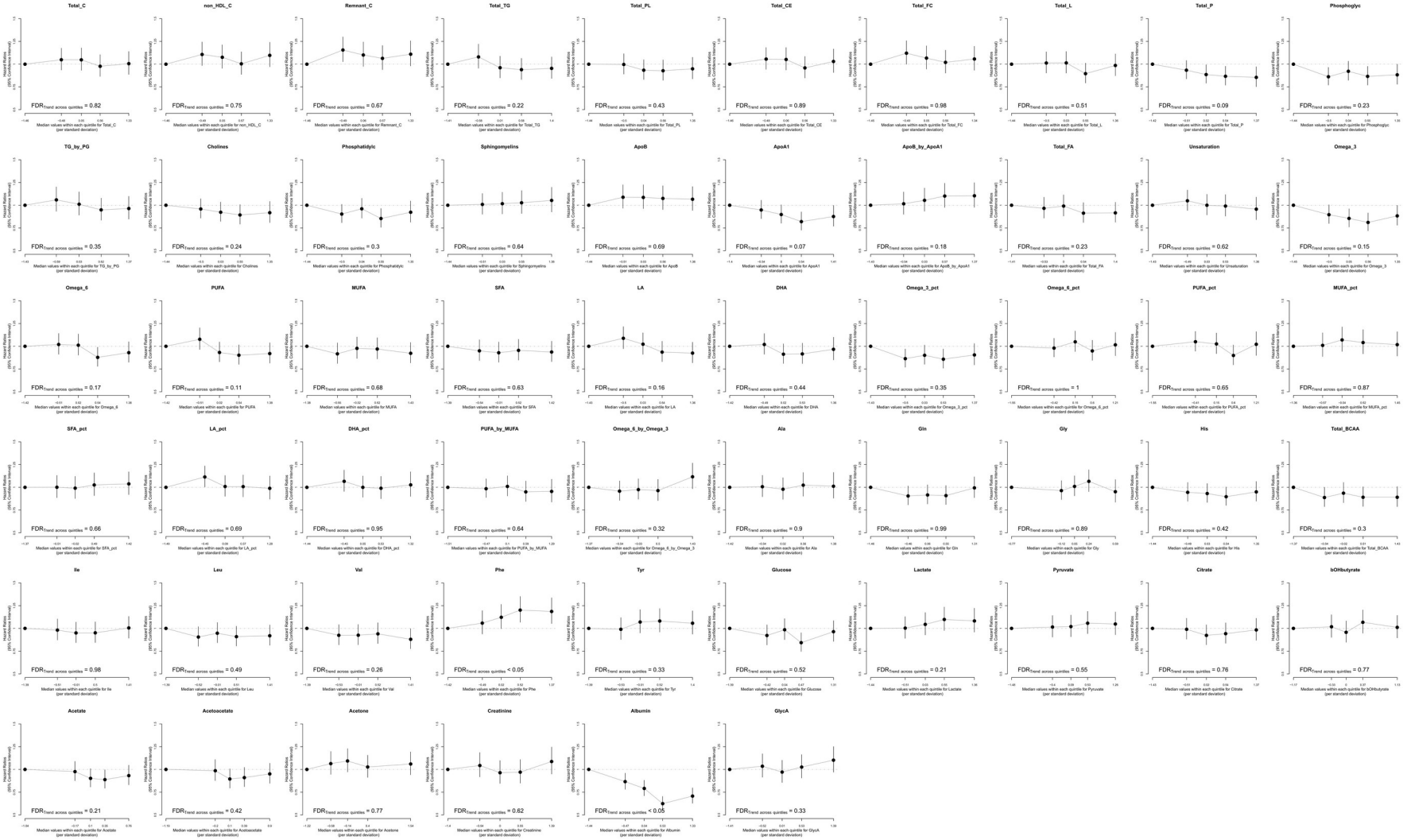
(B) IDLs and LDLs



(C) HDLs

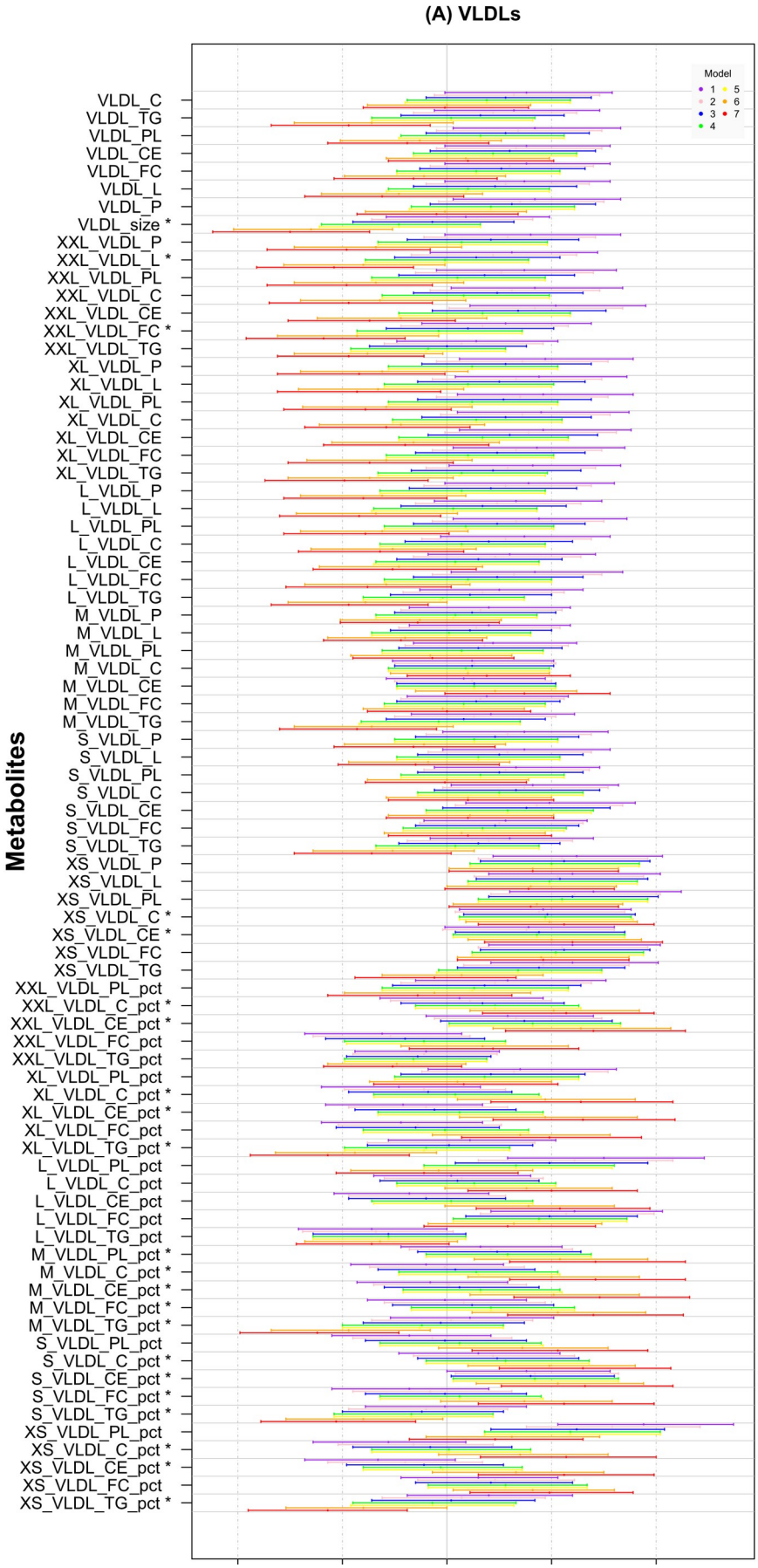

HDL_C
FDRTrend across quintiles = 0.43
HDL_TG
FDRTrend across quintiles = 0.33
HDL_PL
FDRTrend across quintiles = 0.21
HDL_CE
FDRTrend across quintiles = 0.34
HDL_FC
FDRTrend across quintiles = 0.64
HDL_L
FDRTrend across quintiles = 0.16
HDL_P
FDRTrend across quintiles = 0.06
HDL_size
FDRTrend across quintiles = 0.5
XL_HDL_P
FDRTrend across quintiles = 0.18
XL_HDL_L
FDRTrend across quintiles = 0.15
XL_HDL_PL
FDRTrend across quintiles = 0.17
XL_HDL_C
FDRTrend across quintiles = 0.17
XL_HDL_CE
FDRTrend across quintiles = 0.23
XL_HDL_FC
FDRTrend across quintiles < 0.05
XL_HDL_TG
FDRTrend across quintiles = 0.92
L_HDL_P
FDRTrend across quintiles = 0.94
L_HDL_L
FDRTrend across quintiles = 0.95
L_HDL_PL
FDRTrend across quintiles = 0.79
L_HDL_C
FDRTrend across quintiles = 0.86
L_HDL_CE
FDRTrend across quintiles = 0.98
L_HDL_FC
FDRTrend across quintiles = 0.66
L_HDL_TG
FDRTrend across quintiles = 0.75
M_HDL_P
FDRTrend across quintiles = 0.09
M_HDL_L
FDRTrend across quintiles = 0.08
M_HDL_PL
FDRTrend across quintiles = 0.08
M_HDL_C
FDRTrend across quintiles = 0.08
M_HDL_CE
FDRTrend across quintiles = 0.07
M_HDL_FC
FDRTrend across quintiles = 0.11
M_HDL_TG
FDRTrend across quintiles = 0.22
S_HDL_P
FDRTrend across quintiles < 0.05
S_HDL_L
FDRTrend across quintiles < 0.05
S_HDL_PL
FDRTrend across quintiles < 0.05
S_HDL_C
FDRTrend across quintiles < 0.05
S_HDL_CE
FDRTrend across quintiles < 0.05
S_HDL_FC
FDRTrend across quintiles = 0.15
S_HDL_TG
FDRTrend across quintiles = 0.18
XL_HDL_PL_pct
FDRTrend across quintiles = 0.15
XL_HDL_C_pct
FDRTrend across quintiles = 0.87
XL_HDL_CE_pct
FDRTrend across quintiles = 0.23
XL_HDL_FC_pct
FDRTrend across quintiles = 0.86
XL_HDL_TG_pct
FDRTrend across quintiles = 0.15
L_HDL_PL_pct
FDRTrend across quintiles = 0.76
L_HDL_C_pct
FDRTrend across quintiles = 0.62
L_HDL_CE_pct
FDRTrend across quintiles = 0.98
L_HDL_FC_pct
FDRTrend across quintiles < 0.05
L_HDL_TG_pct
FDRTrend across quintiles = 0.64
M_HDL_PL_pct
FDRTrend across quintiles = 0.81
M_HDL_C_pct
FDRTrend across quintiles = 0.98
M_HDL_CE_pct
FDRTrend across quintiles = 0.96
M_HDL_FC_pct
FDRTrend across quintiles = 0.98
M_HDL_TG_pct
FDRTrend across quintiles = 0.98
S_HDL_PL_pct
FDRTrend across quintiles = 0.73
S_HDL_C_pct
FDRTrend across quintiles = 0.89
S_HDL_CE_pct
FDRTrend across quintiles = 0.89
S_HDL_FC_pct
FDRTrend across quintiles < 0.05
S_HDL_TG_pct
FDRTrend across quintiles = 0.99

(D) Amino acids, fatty acids and other metabolites



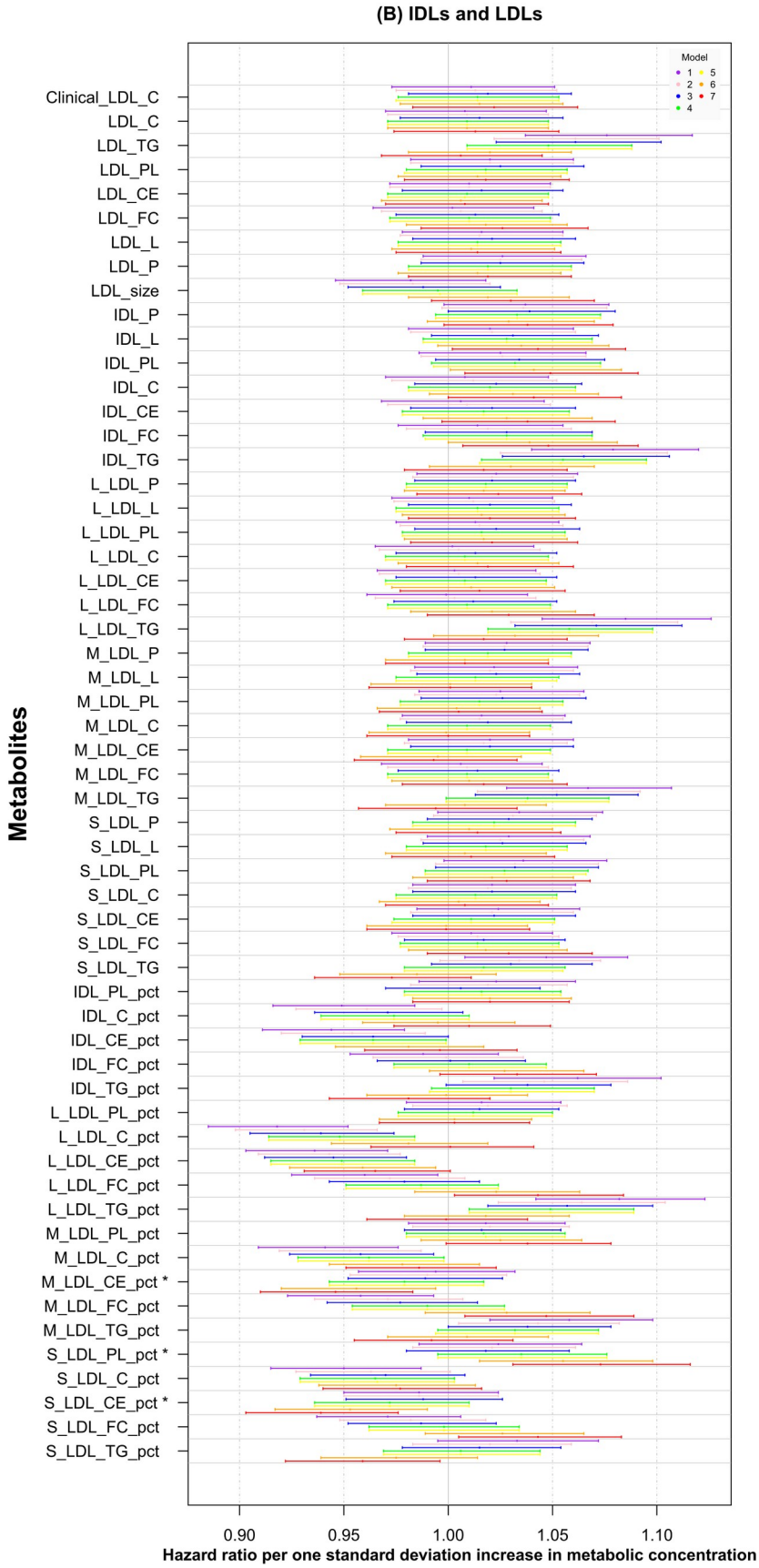
Supplementary Figure 4.1 Associations of incident ischaemic stroke with quintiles of levels of individual metabolites (n = 184,564)

Cox regression, controlling for age, sex, ethnicity, lipid modifying medication, fasting time, spectrometers, assessment centres, body mass index, smoking status, alcohol intake, dietary supplement use, 13 dietary variables, physical activity, prevalent diabetes, and systolic blood pressure, was used to evaluate associations of incident ischaemic stroke with quintiles of levels of individual metabolites. (A) VLDLs, (B) IDLs and LDLs, (C) HDLs, (D) Amino acids, fatty acids, and other metabolites. The trend tests were significant for 12 metabolites, including phenylalanine; albumin; free cholesterol to total lipids ratio in large HDL; cholesterol, cholesteryl esters, total lipids, phospholipids, free cholesterol to total lipids ratio in small HDL; concentration of small HDL particles; average diameter for VLDL particles; free cholesterol in very large HDL; cholesteryl esters to total lipids ratio in chylomicrons and extremely large VLDL. Abbreviations are in line with **Supplementary Table 3.1**.

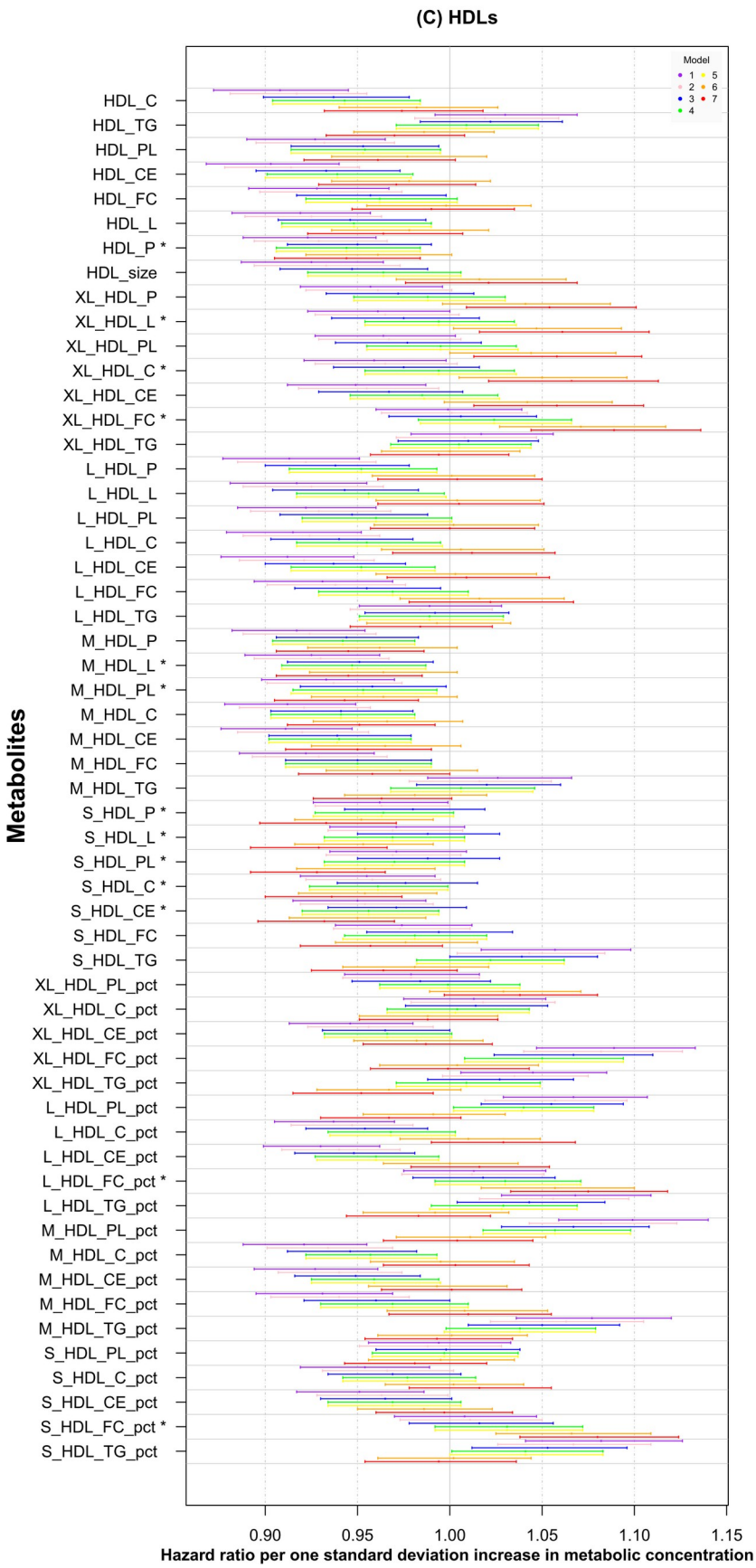

(A) VLDLs
Model
1
2
3
4
5
6
7
Metabolites
VLDL_C
VLDL_TG
VLDL_PL
VLDL_CE
VLDL_FC
VLDL_L
VLDL_P
VLDL_size *
XXL_VLDL_P
XXL_VLDL_L *
XXL_VLDL_PL
XXL_VLDL_C
XXL_VLDL_CE
XXL_VLDL_FC *
XXL_VLDL_TG
XL_VLDL_P
XL_VLDL_L
XL_VLDL_PL
XL_VLDL_C
XL_VLDL_CE
XL_VLDL_FC
XL_VLDL_TG
L_VLDL_P
L_VLDL_L
L_VLDL_PL
L_VLDL_C
L_VLDL_CE
L_VLDL_FC
L_VLDL_TG
M_VLDL_P
M_VLDL_L
M_VLDL_PL
M_VLDL_C
M_VLDL_CE
M_VLDL_FC
M_VLDL_TG
S_VLDL_P
S_VLDL_L
S_VLDL_PL
S_VLDL_C
S_VLDL_CE
S_VLDL_FC
S_VLDL_TG
XS_VLDL_P
XS_VLDL_L
XS_VLDL_PL
XS_VLDL_C *
XS_VLDL_CE *
XS_VLDL_FC
XS_VLDL_TG
XXL_VLDL_PL_pct
XXL_VLDL_C_pct *
XXL_VLDL_CE_pct *
XXL_VLDL_FC_pct
XXL_VLDL_TG_pct
XL_VLDL_PL_pct
XL_VLDL_C_pct *
XL_VLDL_CE_pct *
XL_VLDL_FC_pct
XL_VLDL_TG_pct *
L_VLDL_PL_pct
L_VLDL_C_pct
L_VLDL_CE_pct
L_VLDL_FC_pct
L_VLDL_TG_pct
M_VLDL_PL_pct *
M_VLDL_C_pct *
M_VLDL_CE_pct *
M_VLDL_FC_pct *
M_VLDL_TG_pct *
S_VLDL_PL_pct
S_VLDL_C_pct *
S_VLDL_CE_pct *
S_VLDL_FC_pct *
S_VLDL_TG_pct *
XS_VLDL_PL_pct
XS_VLDL_C_pct *
XS_VLDL_CE_pct *
XS_VLDL_FC_pct
XS_VLDL_TG_pct *

Supplementary Figure 4.2 Results of metabolome-wide association analysis

Circle dots and lines denote the hazard ratios with 95% confidence intervals per one standard deviation higher metabolic levels on the natural log scale between incident ischaemic stroke and individual metabolites from seven models of Cox regression. Model 1 was adjusted for age, sex, ethnicity, lipid modifying medication, fasting time, spectrometers, and assessment centres; model 2 was additionally adjusted for smoking status; model 3 for alcohol intake frequency; model 4 for dietary supplement use and 13 dietary variables (fresh fruit, whole and refined grain food, caffeinated and decaffeinated coffee, tea, vegetables, cheese, unprocessed red meat, processed meat, oily and non-oily fish, and type of spread); model 5 for physical activity; model 6 for body mass index; and model 7 for prevalent diabetes and systolic blood pressure. Stars indicate associations that surpassed false discovery rate significance threshold (0.05) in model 7. (A) VLDLs, (B) IDLs and LDLs, (C) HDLs, (D) Amino acids, fatty acids, and other metabolites. Abbreviations are in line with **Supplementary Table 3.1**.



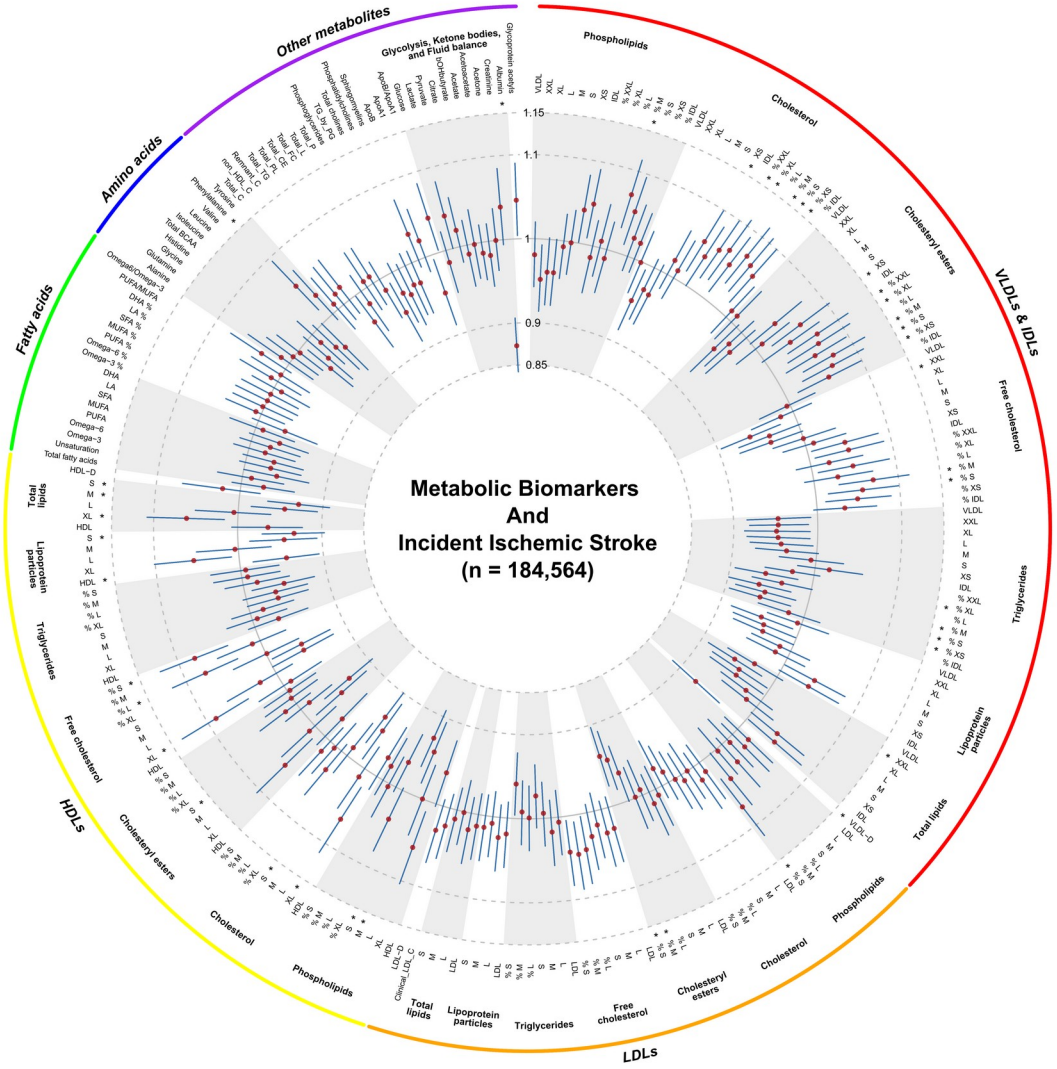
Supplementary Figure 4.2 (Continued) Results of metabolome-wide association analysis


(C) HDLs
Model
1
2
3
4
5
6
7
HDL_C
HDL_TG
HDL_PL
HDL_CE
HDL_FC
HDL_L
HDL_P *
HDL_size
XL_HDL_P
XL_HDL_L *
XL_HDL_PL
XL_HDL_C *
XL_HDL_CE
XL_HDL_FC *
XL_HDL_TG
L_HDL_P
L_HDL_L
L_HDL_PL
L_HDL_C
L_HDL_CE
L_HDL_FC
L_HDL_TG
M_HDL_P
M_HDL_L *
M_HDL_PL *
M_HDL_C
M_HDL_CE
M_HDL_FC
M_HDL_TG
S_HDL_P *
S_HDL_L *
S_HDL_PL *
S_HDL_C *
S_HDL_CE *
S_HDL_FC
S_HDL_TG
XL_HDL_PL_pct
XL_HDL_C_pct
XL_HDL_CE_pct
XL_HDL_FC_pct
XL_HDL_TG_pct
L_HDL_PL_pct
L_HDL_C_pct
L_HDL_CE_pct
L_HDL_FC_pct *
L_HDL_TG_pct
M_HDL_PL_pct
M_HDL_C_pct
M_HDL_CE_pct
M_HDL_FC_pct
M_HDL_TG_pct
S_HDL_PL_pct
S_HDL_C_pct
S_HDL_CE_pct
S_HDL_FC_pct *
S_HDL_TG_pct
Metabolites
0.90
0.95
1.00
1.05
1.10
1.15
Hazard ratio per one standard deviation increase in metabolic concentration

Supplementary Figure 4.2 (Continued) Results of metabolome-wide association analysis

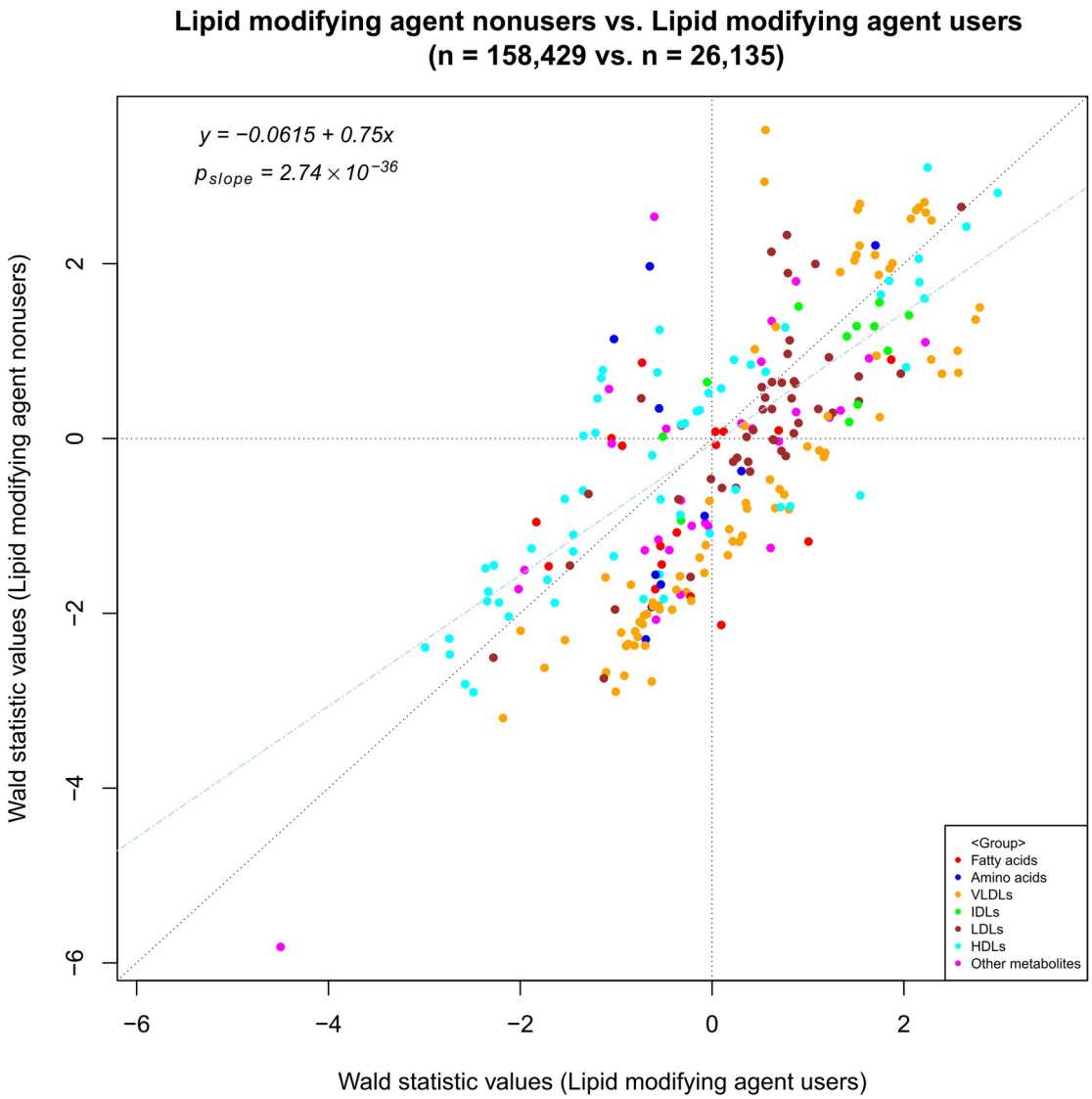


Supplementary Figure 4.2 (Continued) Results of metabolome-wide association analysis



Supplementary Figure 4.3 Associations of individual metabolites with incident ischaemic stroke in model 7 (n = 184,564)

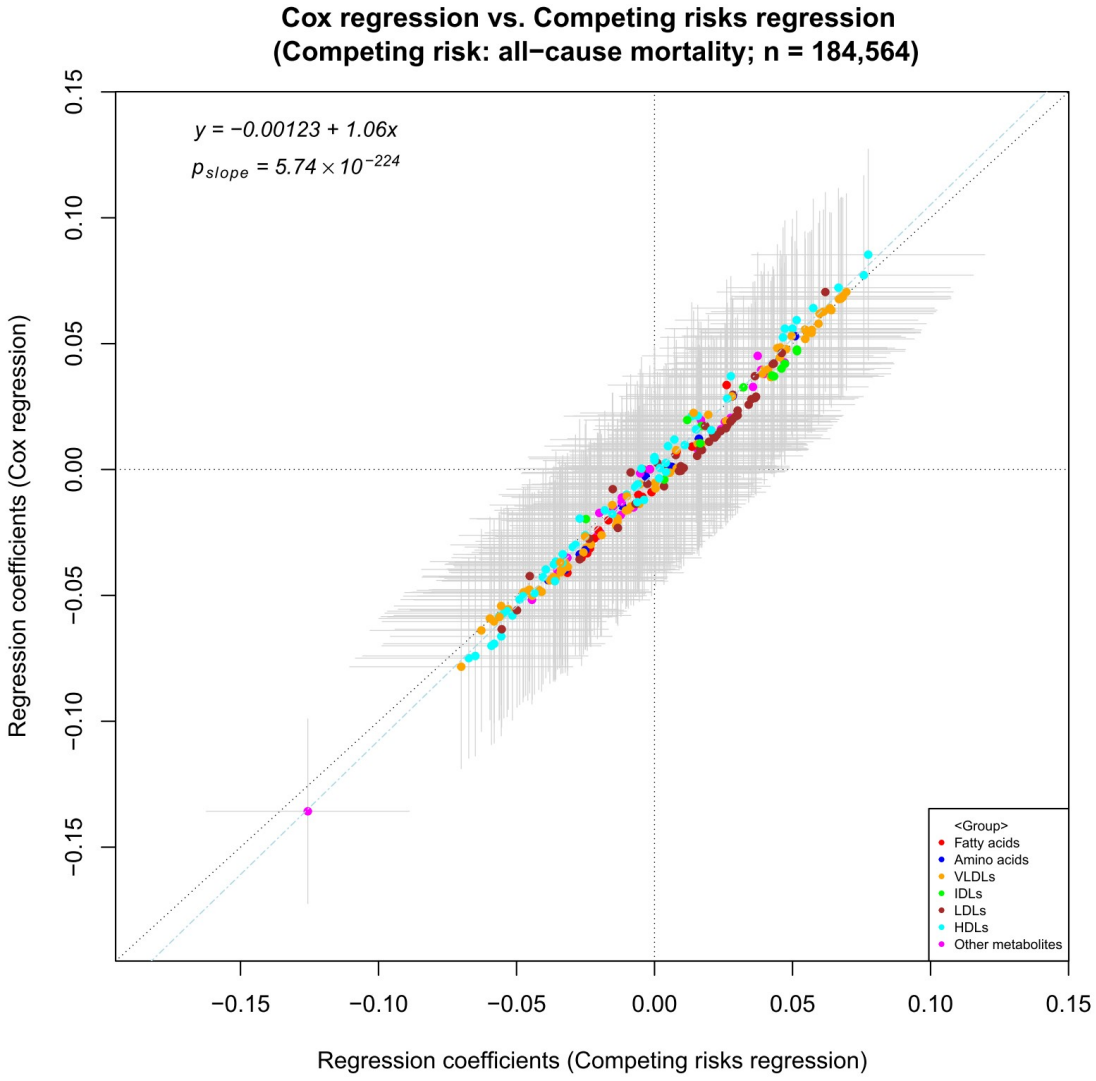
The model 7 was used to adjust for age, sex, ethnicity, lipid modifying medication, fasting time, spectrometers, assessment centres, body mass index, smoking status, alcohol intake, dietary supplement use, 13 dietary variables, physical activity, prevalent diabetes, and systolic blood pressure. The hazard ratios (with 95% confidence intervals) are presented per one standard deviation higher metabolic biomarker on the natural log scale. Stars indicate false discovery rate < 0.05. Abbreviations are in line with **Supplementary Table 3.1**.



Supplementary Figure 4.4 Comparison of associations of metabolic biomarkers with incident ischaemic stroke in the lipid modifying agent nonusers with users

The scatter plot was constructed by the Wald statistic values for the association of each metabolite with incident ischaemic stroke in Cox proportional hazards regression models (1) in the lipid modifying agent nonusers on the y-axis, and (2) in the lipid modifying agent users on the x-axis. The model 7 was used to adjust for age, sex, ethnicity, lipid modifying medication, fasting time, spectrometers, assessment centres, body mass index, smoking status, alcohol intake, dietary supplement use, 13 dietary variables, physical activity, prevalent diabetes, and systolic blood pressure. To compare

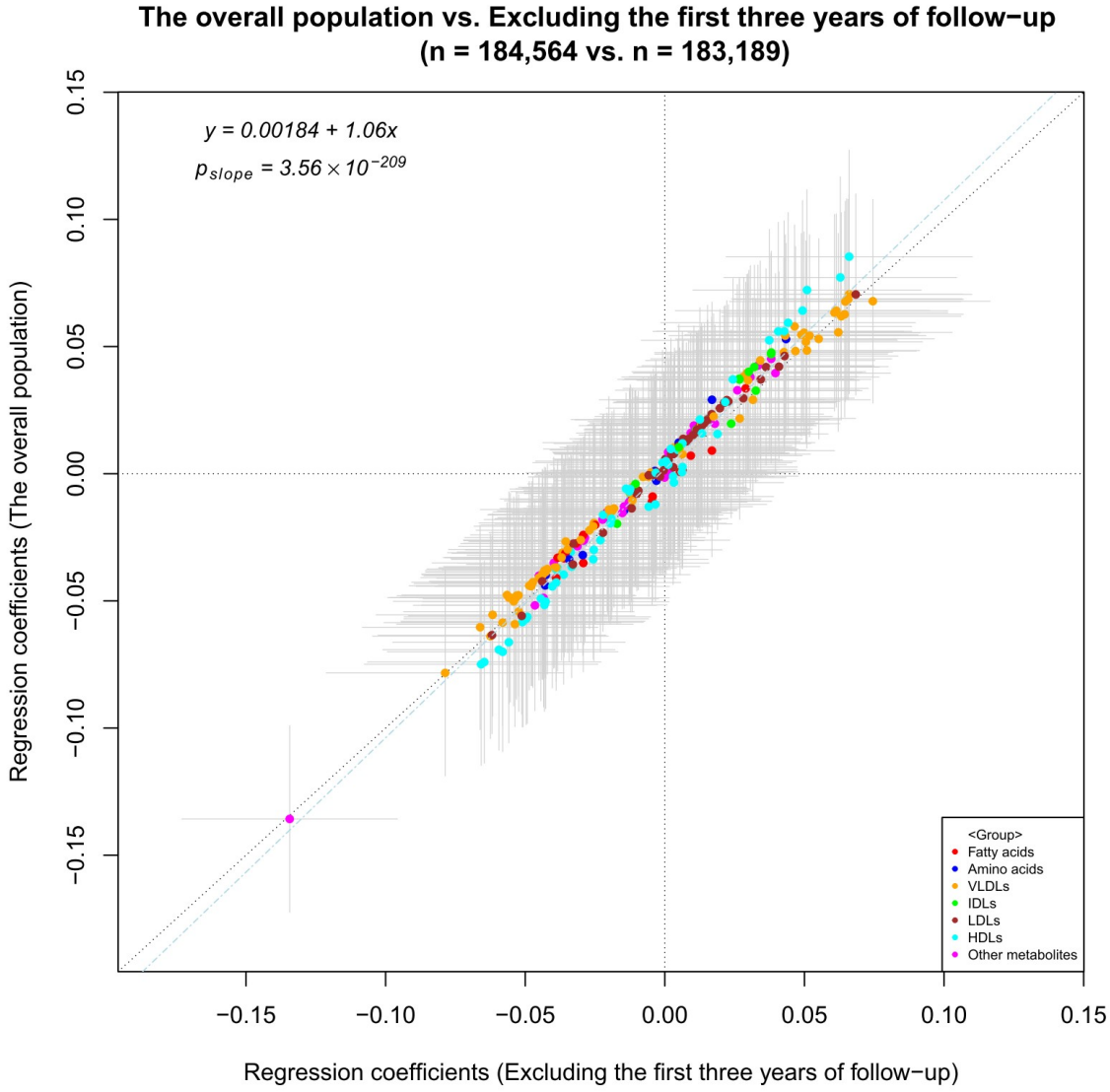
the metabolic signatures between two groups, a linear mixed effects model with 67 cluster numbers of metabolites as random intercepts was applied.



Supplementary Figure 4.5 Comparison of associations of metabolic biomarkers with risk of incident ischaemic stroke between Cox regression and competing risks regression

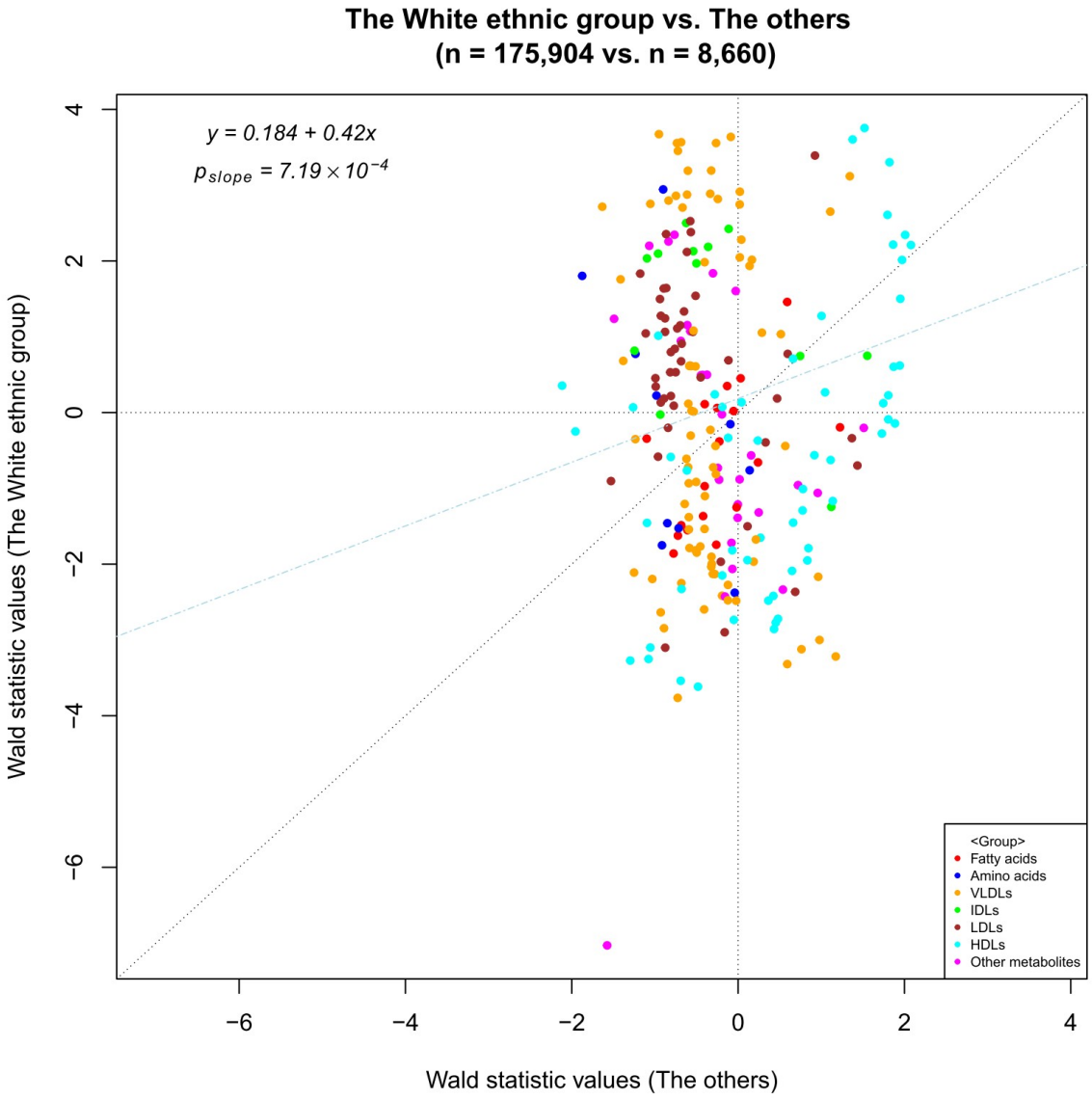
The scatter plot was constructed by the regression coefficients for the association of each metabolite with incident ischaemic stroke (1) in Cox proportional hazards regression models on the y-axis, and (2) in competing risks regression on the x-axis. The model 7 was used to adjust for age, sex, ethnicity, lipid modifying medication, fasting time, spectrometers, assessment centres, body mass index, smoking status, alcohol intake, dietary supplement use, 13 dietary variables, physical activity, prevalent

diabetes, and systolic blood pressure. To compare the metabolic signatures between Cox regression and competing risks regression, a linear mixed effects model with 67 cluster numbers of metabolites as random intercepts was applied.



Supplementary Figure 4.6 Comparison of associations of metabolic biomarkers with incident ischaemic stroke in the overall population with those excluding the first three years of follow-up

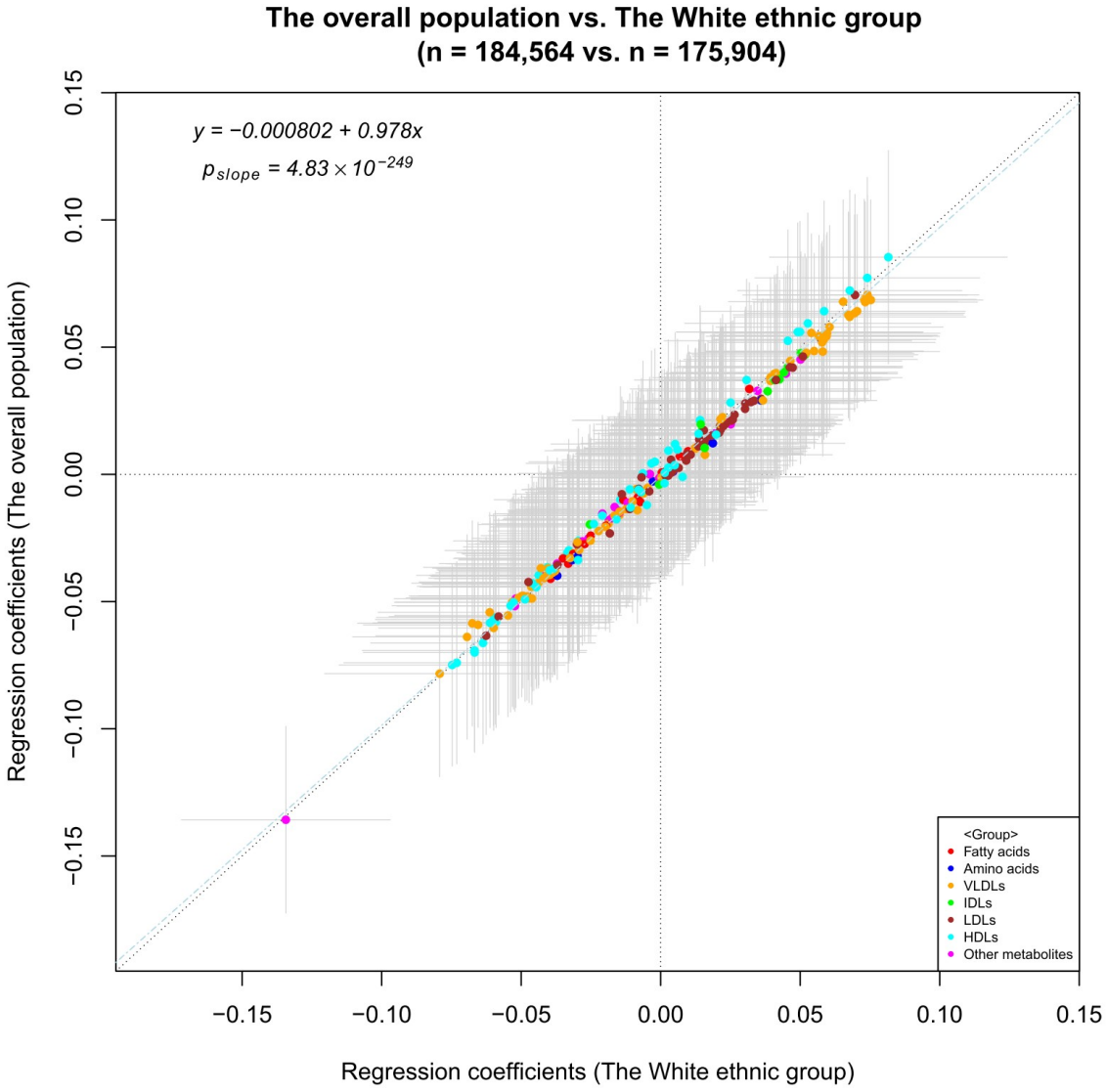
The scatter plot was constructed by the regression coefficients for the association of each metabolite with incident ischaemic stroke in Cox proportional hazards regression models (1) in the whole population on the y-axis, and (2) in those excluding the first three years of follow-up on the x-axis. The model 7 was used to adjust for age, sex, ethnicity, lipid modifying medication, fasting time, spectrometers, assessment centres, body mass index, smoking status, alcohol intake, dietary supplement use, 13 dietary variables, physical activity, prevalent diabetes, and systolic blood pressure. To compare the metabolic signatures between two groups, a linear mixed effects model with 67 cluster numbers of metabolites as random intercepts was applied.



Supplementary Figure 4.7 Comparison of associations of metabolic biomarkers with incident ischaemic stroke in the White ethnic group with the others

The scatter plot was constructed by the regression coefficients for the association of each metabolite with incident ischaemic stroke in Cox proportional hazards regression models (1) in the White ethnic group on the y-axis, and (2) in the others on the x-axis. The model 7 was used to adjust for age, sex, ethnicity, lipid modifying medication, fasting time, spectrometers, assessment centres, body mass index, smoking status, alcohol intake, dietary supplement use, 13 dietary variables, physical activity, prevalent diabetes, and systolic blood pressure. To compare the metabolic signatures between two

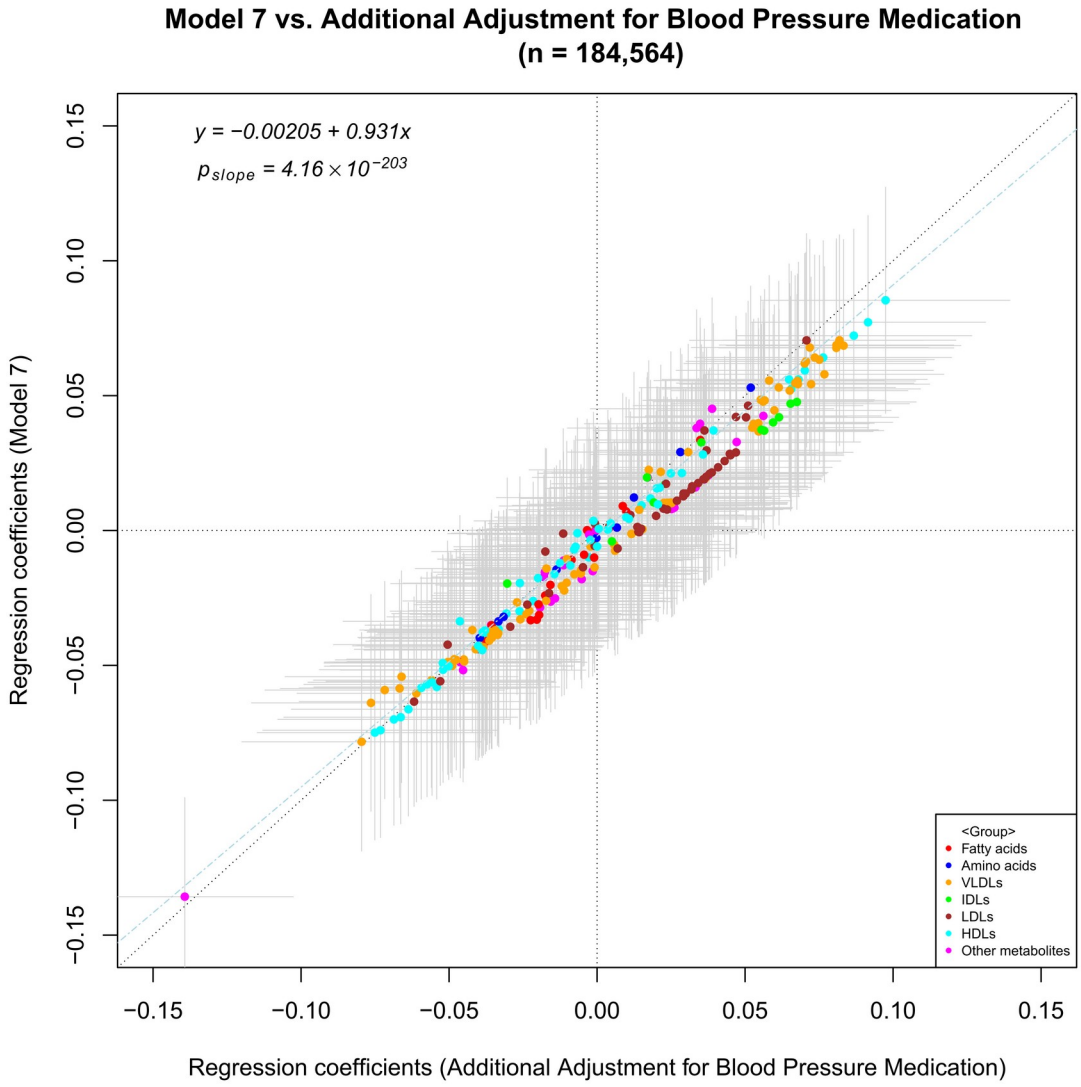
groups, a linear mixed effects model with 67 cluster numbers of metabolites as random intercepts was applied.



Supplementary Figure 4.8 Comparison of associations of metabolic biomarkers with incident ischaemic stroke in the overall population with the White ethnic group

The scatter plot was constructed by the regression coefficients for the association of each metabolite with incident ischaemic stroke in Cox proportional hazards regression models (1) in the overall population on the y-axis, and (2) in the White ethnic group on the x-axis. The model 7 was used to adjust for age, sex, ethnicity, lipid modifying medication, fasting time, spectrometers, assessment centres, body mass index, smoking status, alcohol intake, dietary supplement use, 13 dietary variables, physical activity, prevalent diabetes, and systolic blood pressure. To compare the metabolic signatures

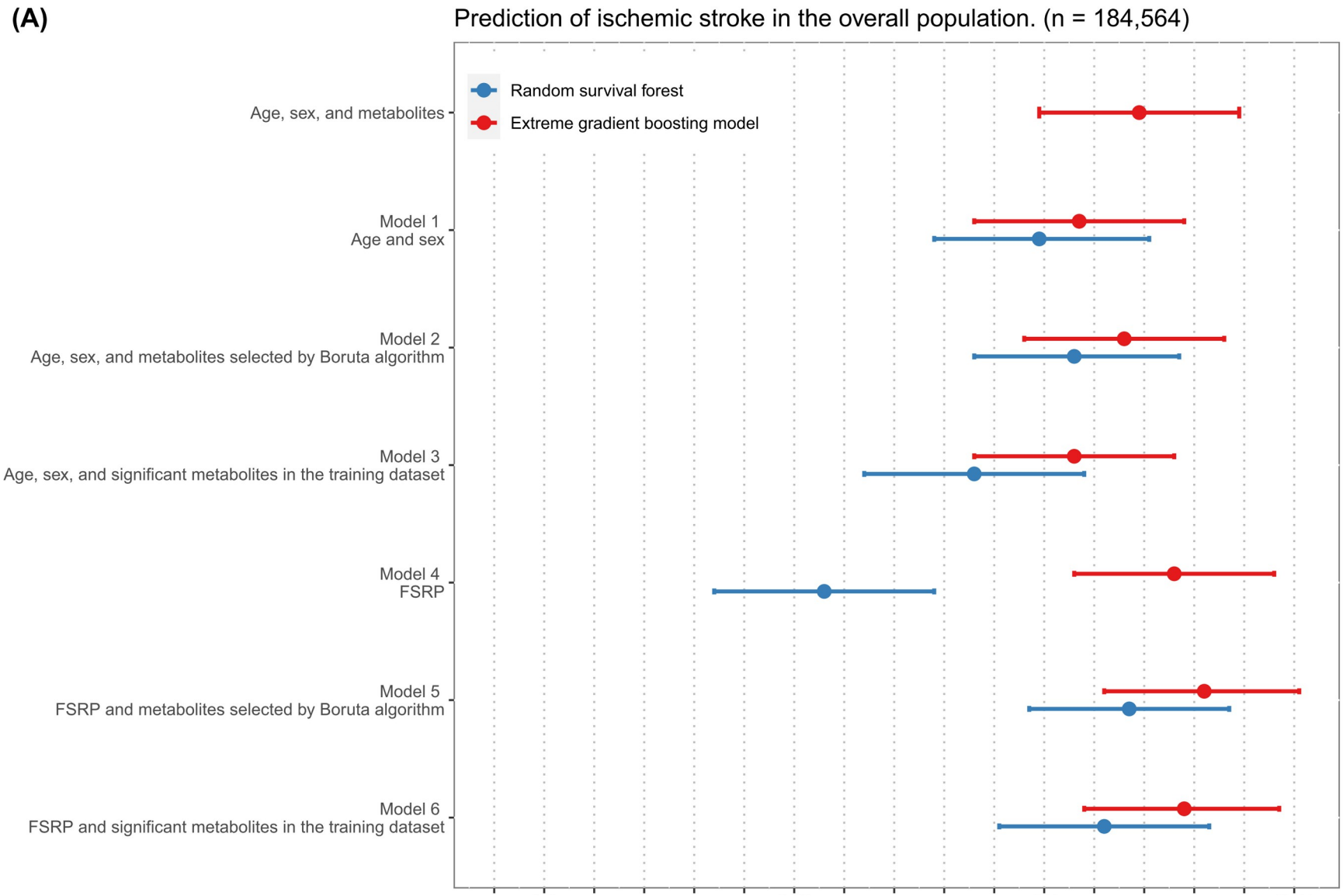
between two groups, a linear mixed effects model with 67 cluster numbers of metabolites as random intercepts was applied.



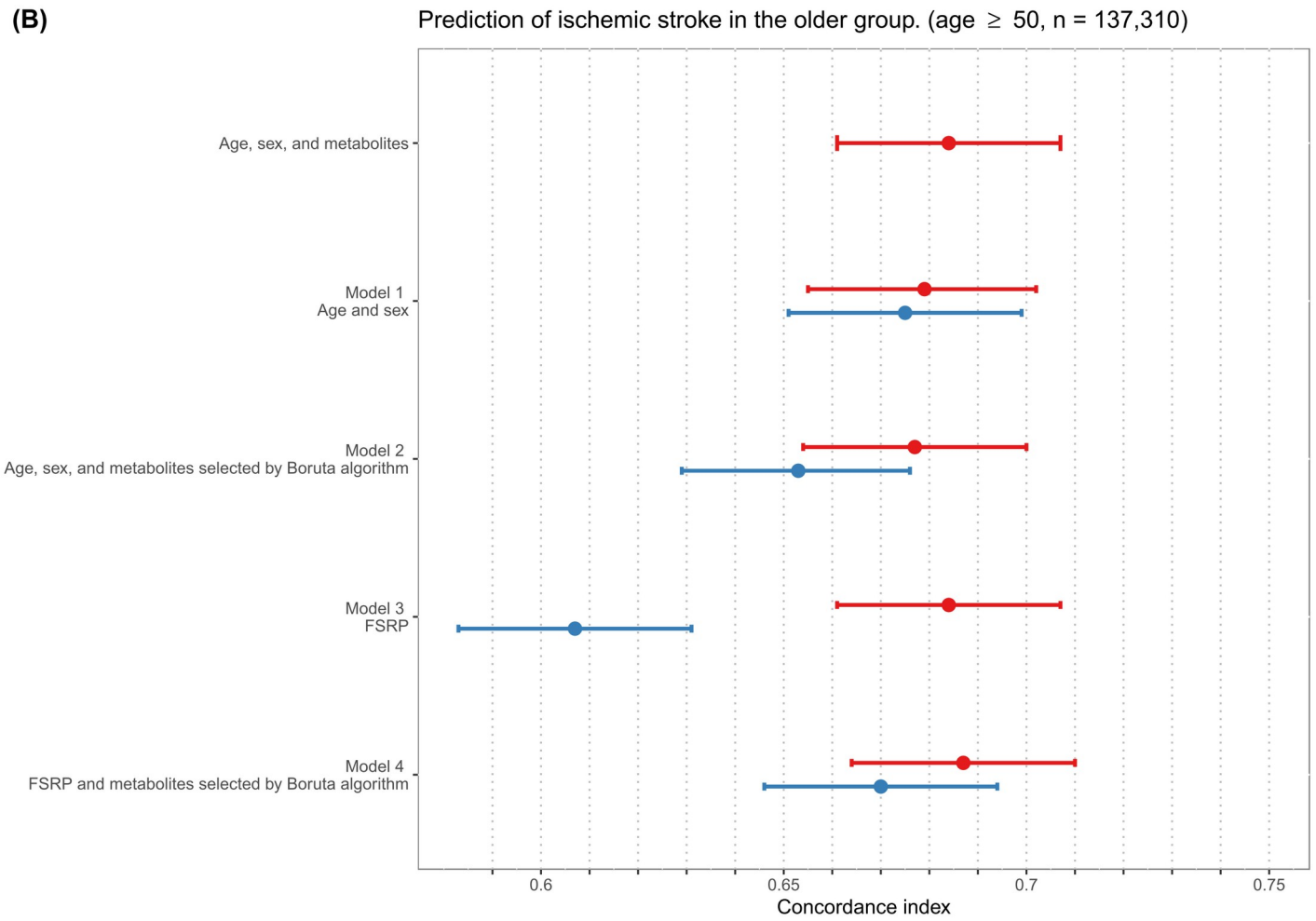
Supplementary Figure 4.9 Comparison of associations of metabolic biomarkers with incident ischaemic stroke in the model 7 with additional adjustment for blood pressure medication.

The scatter plot was constructed by the regression coefficients for the association of each metabolite with incident ischaemic stroke in Cox proportional hazards regression models in the overall population (1) in model 7 on the y-axis, and (2) additional adjustment for blood pressure medication on the x-axis. The model 7 was used to adjust for age, sex, ethnicity, lipid modifying medication, fasting time, spectrometers, assessment centres, body mass index, smoking status, alcohol intake, dietary supplement use, 13 dietary variables, physical activity, prevalent diabetes, and systolic

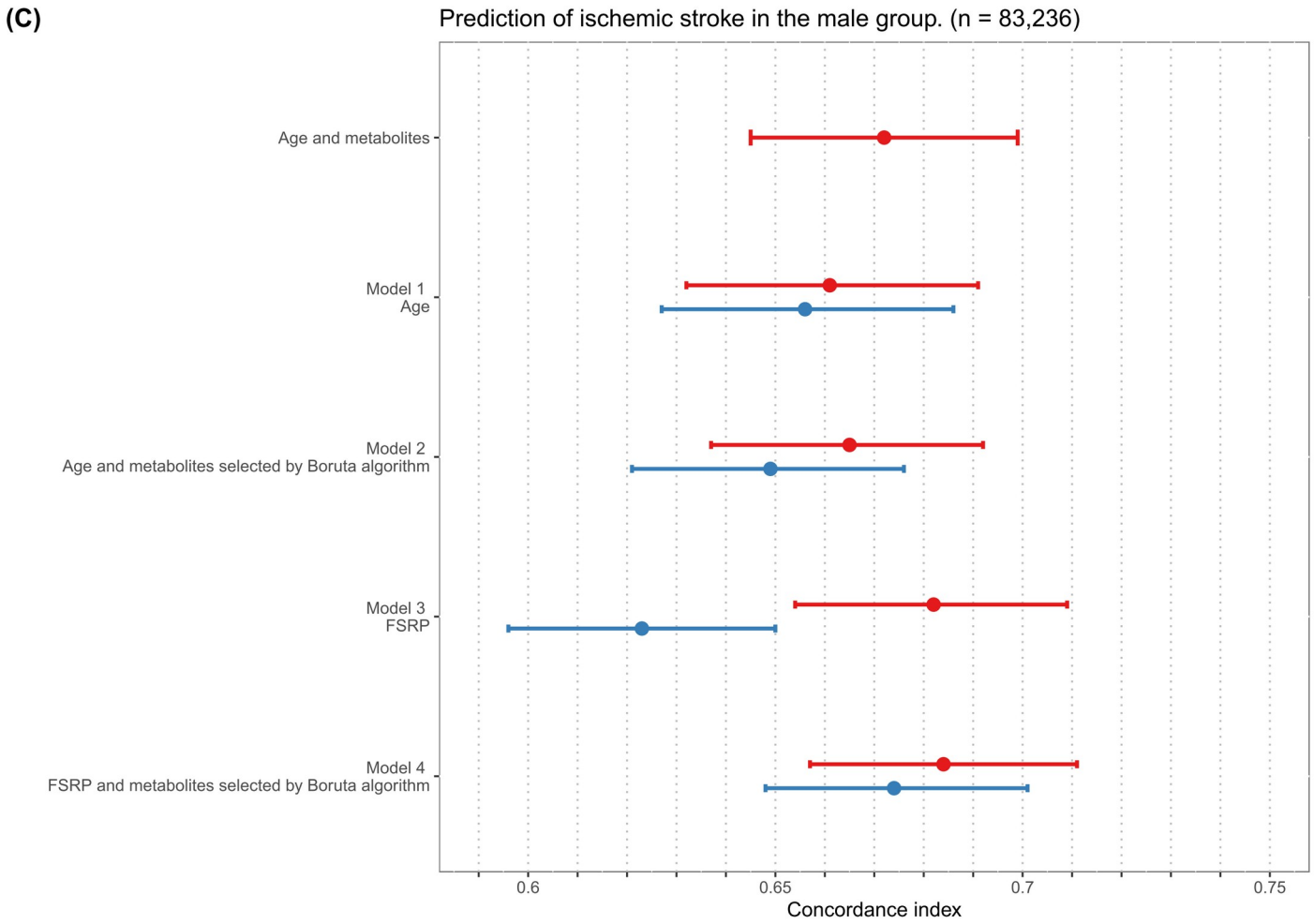
blood pressure. To compare the metabolic signatures between two groups, a linear mixed effects model with 67 cluster numbers of metabolites as random intercepts was applied.


(A)
Prediction of ischemic stroke in the overall population. (n = 184,564)
Random survival forest
Extreme gradient boosting model
Age, sex, and metabolites
Model 1
Age and sex
Model 2
Age, sex, and metabolites selected by Boruta algorithm
Model 3
Age, sex, and significant metabolites in the training dataset
Model 4
FSRP
Model 5
FSRP and metabolites selected by Boruta algorithm
Model 6
FSRP and significant metabolites in the training dataset

Supplementary Figure 5.1 Predictive values on ischaemic stroke by extreme gradient boosting model and random survival forests. Comparison of model performance of extreme gradient boosting model and random survival forests in (A) the overall population, (B) the older group, (C) the male group, and (D) the female group. The predictors of models are shown at the Y axis. Framingham Stroke Risk Profile (FSRP) comprises age, sex, systolic blood pressure, smoking status, prevalent diabetes mellitus, and blood pressure medication. Results are presented as the absolute concordance index (center of error bar) and the 95% confidence interval (line of error bar).

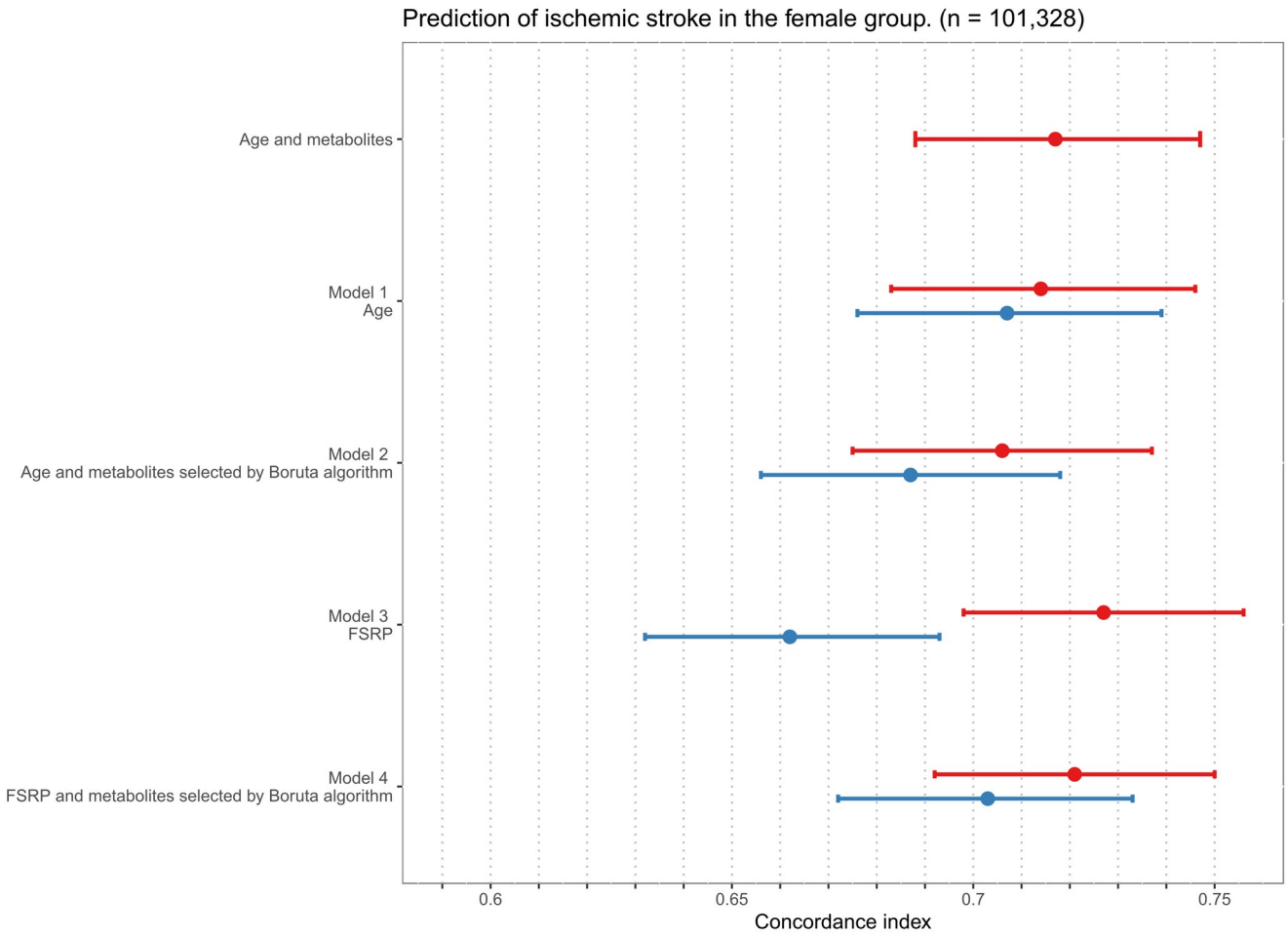


Supplementary Figure 5.1 (Continued) Predictive values on ischaemic stroke by extreme gradient boosting model and random survival forests.

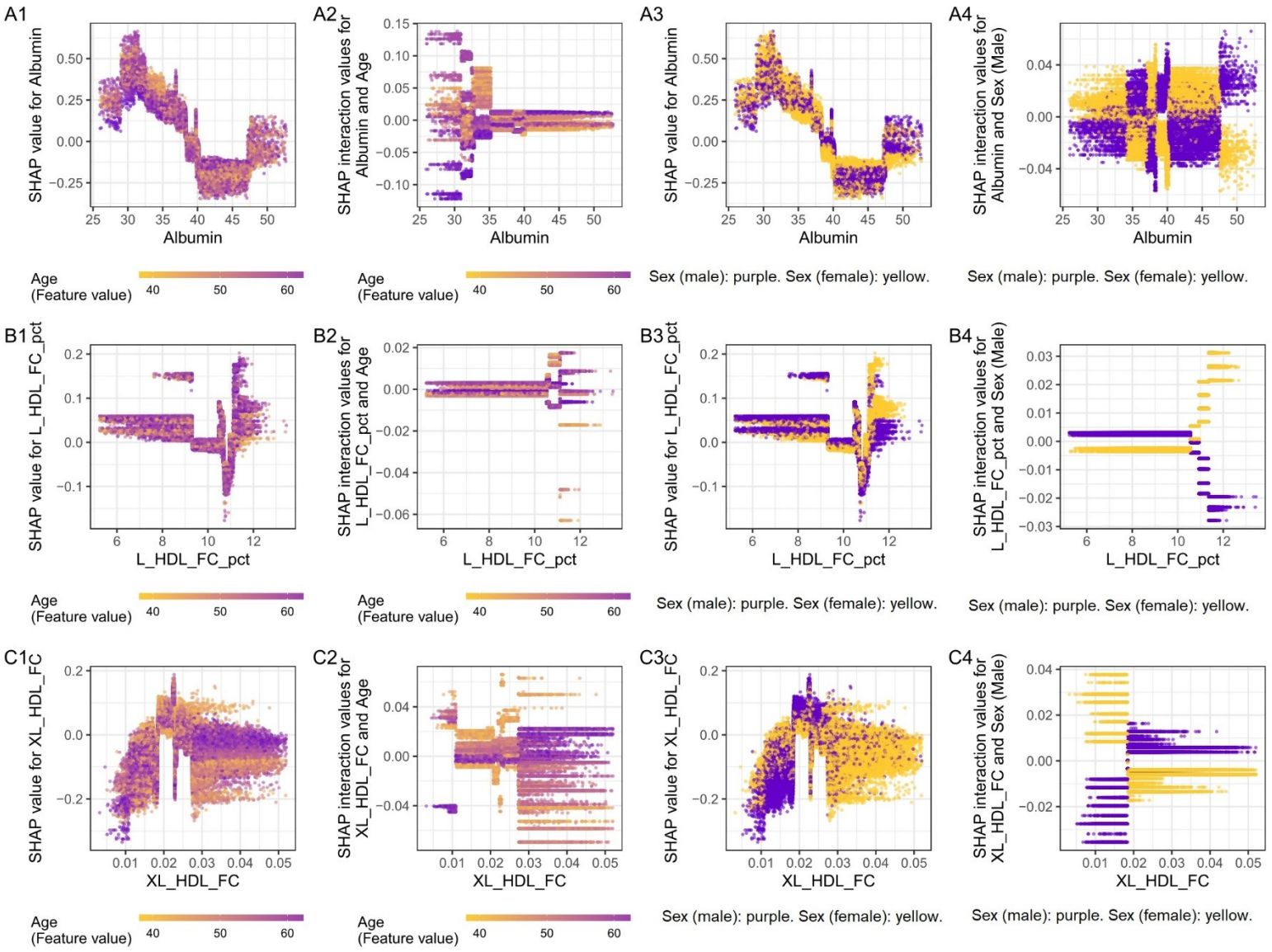


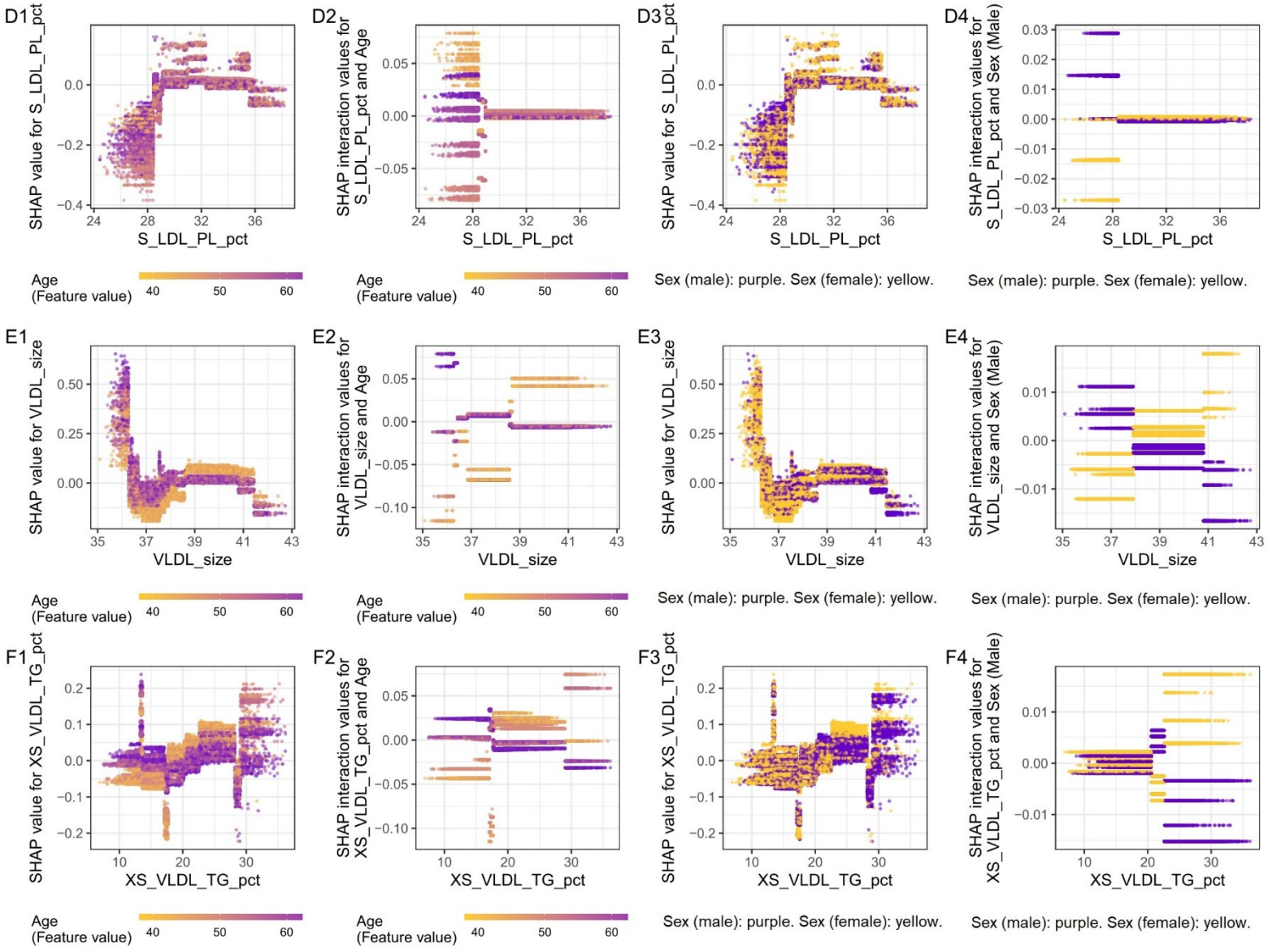
Supplementary Figure 5.1 (Continued) Predictive values on ischaemic stroke by extreme gradient boosting model and random survival forests.

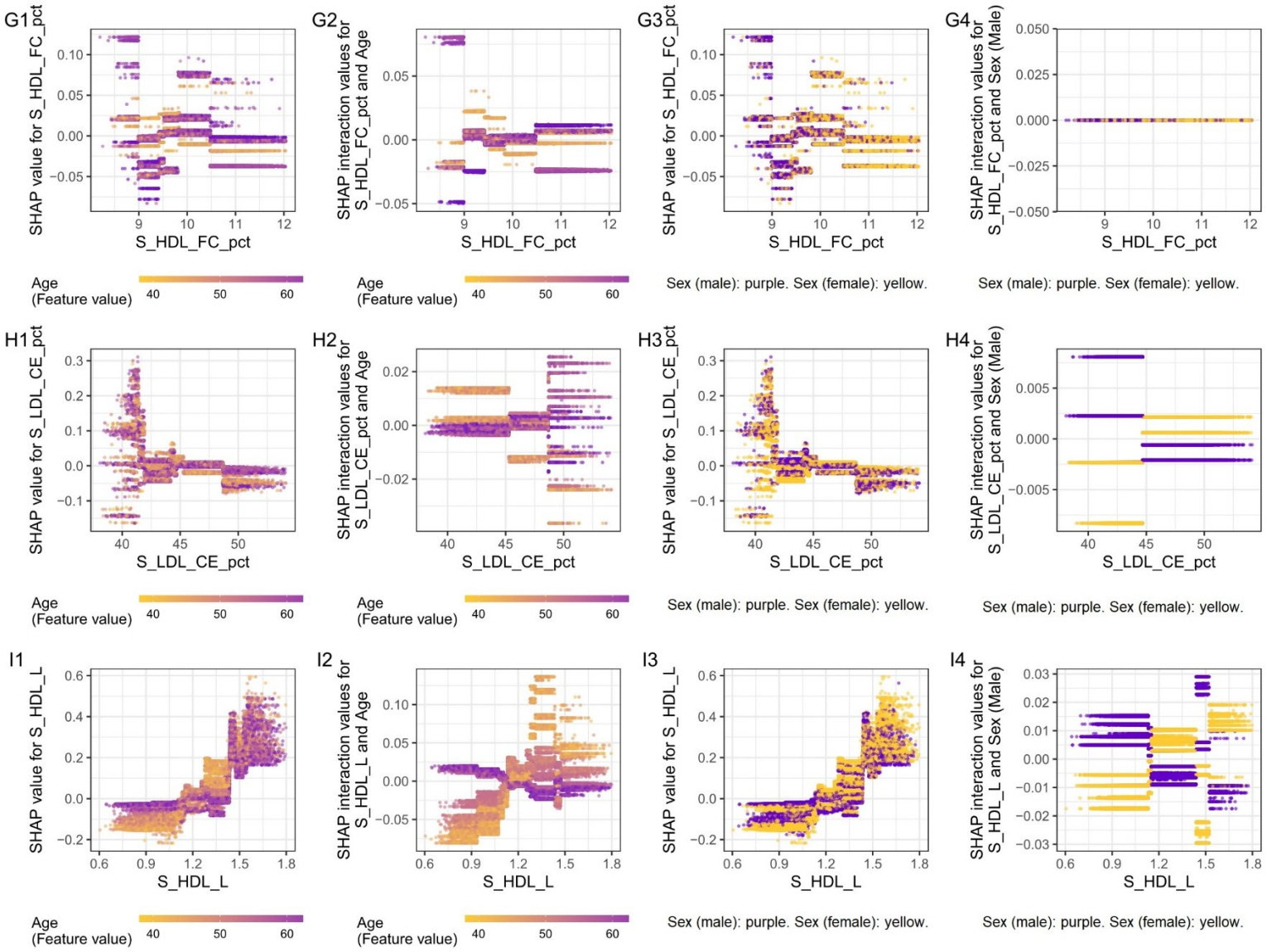
(D)

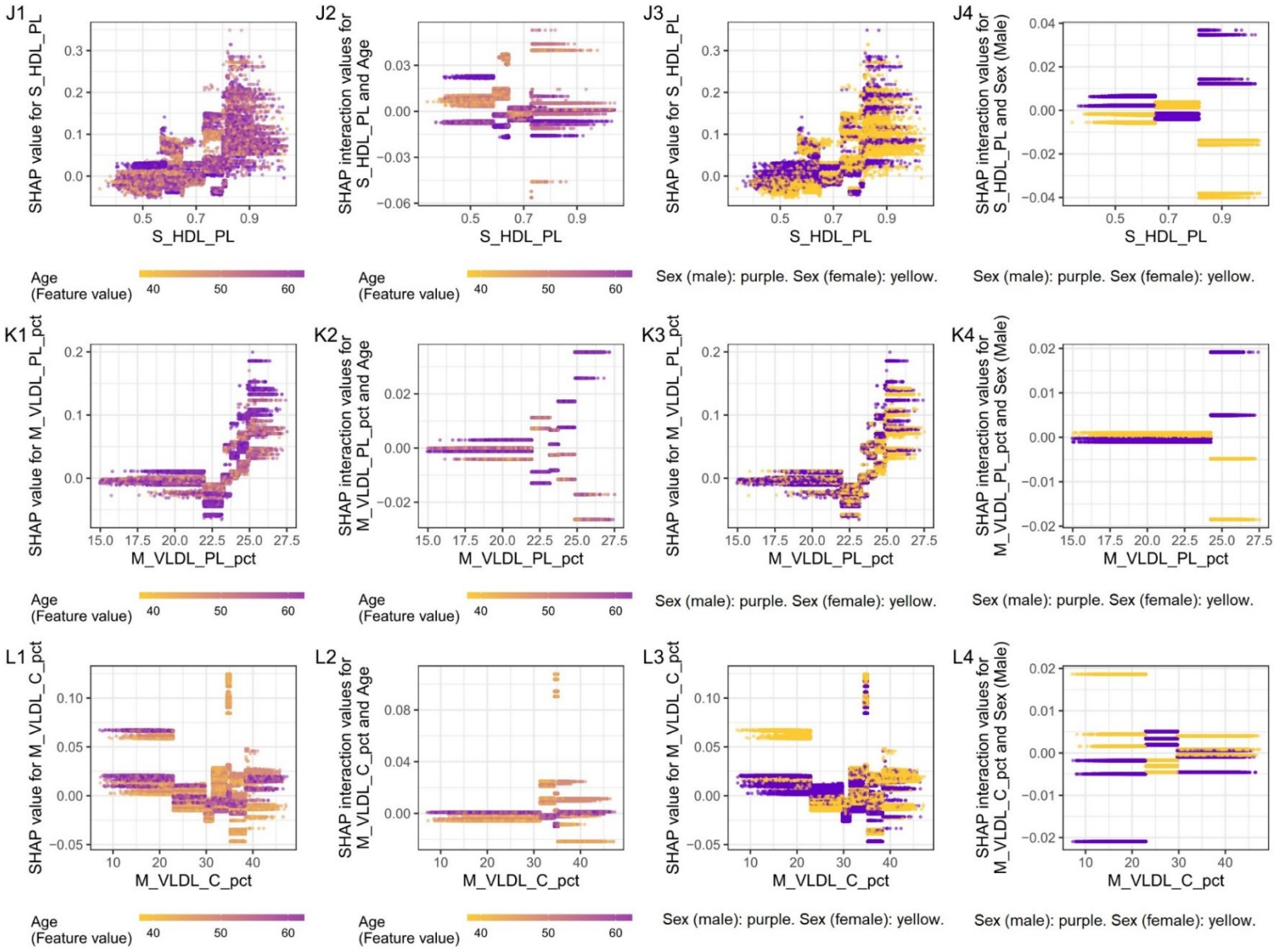

Prediction of ischemic stroke in the female group. (n = 101,328)
Age and metabolites
Model 1
Age
Model 2
Age and metabolites selected by Boruta algorithm
Model 3
FSRP
Model 4
FSRP and metabolites selected by Boruta algorithm
0.6
0.65
0.7
0.75
Concordance index

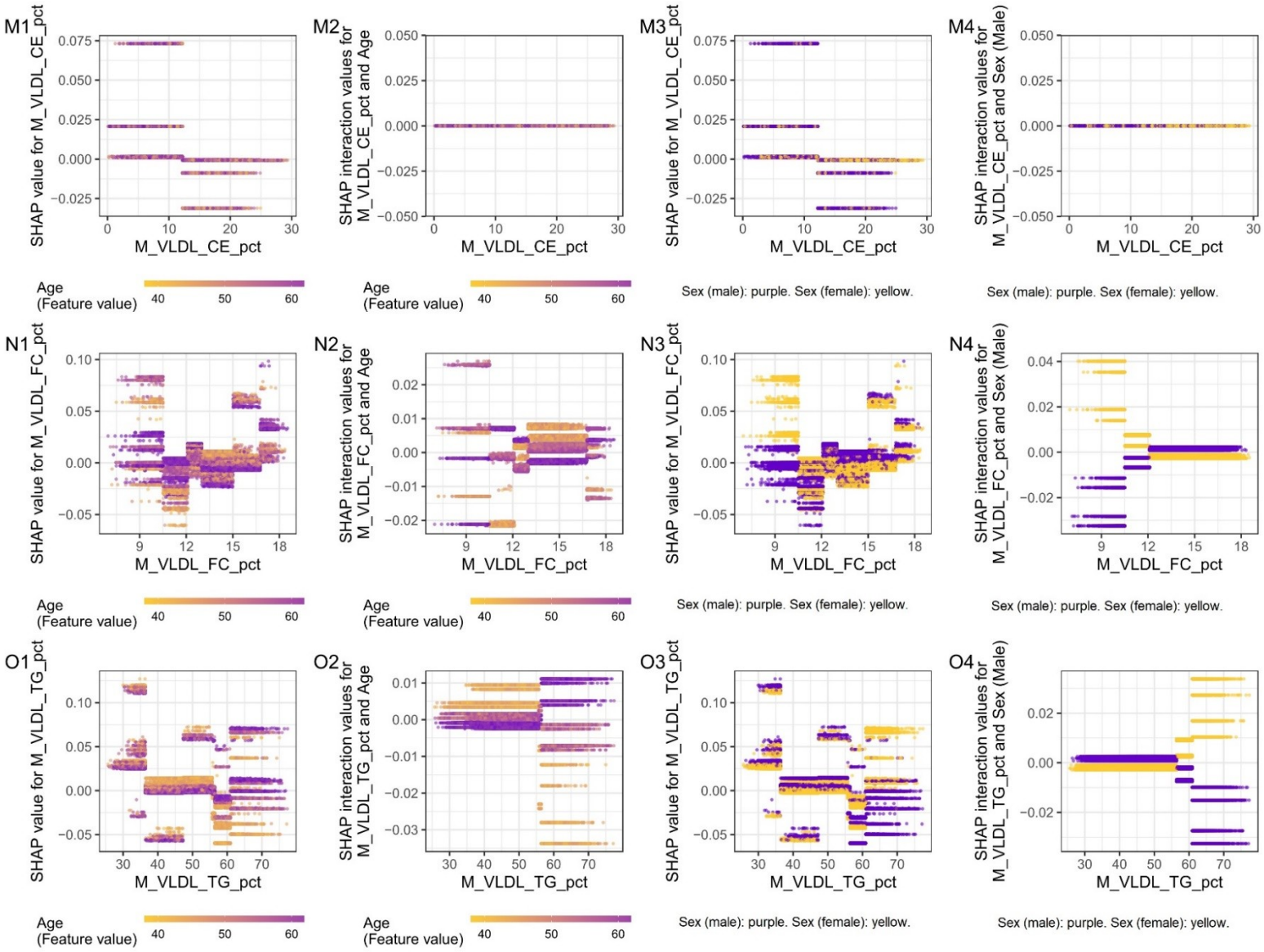
Supplementary Figure 5.1 (Continued) Predictive values on ischaemic stroke by extreme gradient boosting model and random survival forests.

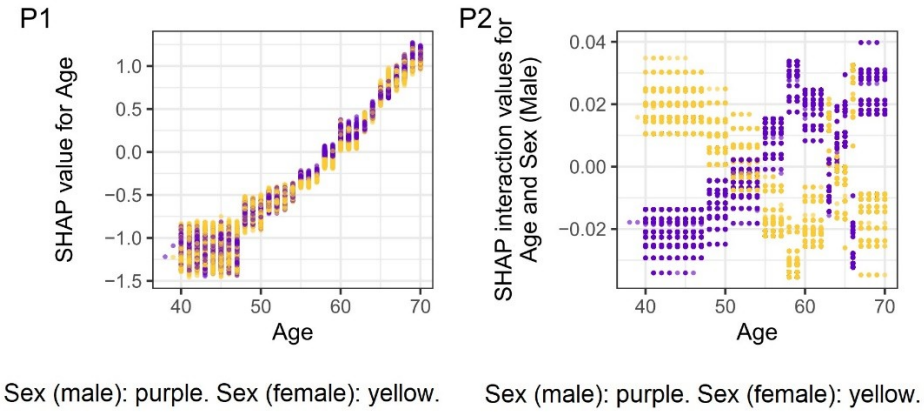

A1
SHAP value for Albumin
Albumin
Age
(Feature value)
40
50
60
A2
SHAP interaction values for Albumin and Age
Albumin
Age
(Feature value)
A3
SHAP value for Albumin
Albumin
Sex (male): purple. Sex (female): yellow.
A4
SHAP interaction values for Albumin and Sex (Male)
Albumin
Sex (male): purple. Sex (female): yellow.
B1
SHAP value for L_HDL_FC_pct
L_HDL_FC_pct
B2
SHAP interaction values for L_HDL_FC_pct and Age
L_HDL_FC_pct
B3
SHAP value for L_HDL_FC_pct
L_HDL_FC_pct
Sex (male): purple. Sex (female): yellow.
B4
SHAP interaction values for L_HDL_FC_pct and Sex (Male)
L_HDL_FC_pct
Sex (male): purple. Sex (female): yellow.
C1
SHAP value for XL_HDL_FC
XL_HDL_FC
C2
SHAP interaction values for XL_HDL_FC and Age
XL_HDL_FC
C3
SHAP value for XL_HDL_FC
XL_HDL_FC
Sex (male): purple. Sex (female): yellow.
C4
SHAP interaction values for XL_HDL_FC and Sex (Male)
XL_HDL_FC
Sex (male): purple. Sex (female): yellow.


D1
SHAP value for S_LDL_PL_pct
S_LDL_PL_pct
Age
(Feature value)
40
50
60
D2
SHAP interaction values for S_LDL_PL_pct and Age
S_LDL_PL_pct
Age
(Feature value)
40
50
60
D3
SHAP value for S_LDL_PL_pct
S_LDL_PL_pct
Sex (male): purple. Sex (female): yellow.
D4
SHAP interaction values for S_LDL_PL_pct and Sex (Male)
S_LDL_PL_pct
Sex (male): purple. Sex (female): yellow.
E1
SHAP value for VLDL_size
VLDL_size
Age
(Feature value)
40
50
60
E2
SHAP interaction values for VLDL_size and Age
VLDL_size
Age
(Feature value)
40
50
60
E3
SHAP value for VLDL_size
VLDL_size
Sex (male): purple. Sex (female): yellow.
E4
SHAP interaction values for VLDL_size and Sex (Male)
VLDL_size
Sex (male): purple. Sex (female): yellow.
F1
SHAP value for XS_VLDL_TG_pct
XS_VLDL_TG_pct
Age
(Feature value)
40
50
60
F2
SHAP interaction values for XS_VLDL_TG_pct and Age
XS_VLDL_TG_pct
Age
(Feature value)
40
50
60
F3
SHAP value for XS_VLDL_TG_pct
XS_VLDL_TG_pct
Sex (male): purple. Sex (female): yellow.
F4
SHAP interaction values for XS_VLDL_TG_pct and Sex (Male)
XS_VLDL_TG_pct
Sex (male): purple. Sex (female): yellow.


G1
SHAP value for S_HDL_FC_pct
S_HDL_FC_pct
G2
SHAP interaction values for S_HDL_FC_pct and Age
S_HDL_FC_pct
G3
SHAP value for S_HDL_FC_pct
S_HDL_FC_pct
G4
SHAP interaction values for S_HDL_FC_pct and Sex (Male)
S_HDL_FC_pct
Age (Feature value)
Sex (male): purple. Sex (female): yellow.
H1
SHAP value for S_LDL_CE_pct
S_LDL_CE_pct
H2
SHAP interaction values for S_LDL_CE_pct and Age
S_LDL_CE_pct
H3
SHAP value for S_LDL_CE_pct
S_LDL_CE_pct
H4
SHAP interaction values for S_LDL_CE_pct and Sex (Male)
S_LDL_CE_pct
I1
SHAP value for S_HDL_L
S_HDL_L
I2
SHAP interaction values for S_HDL_L and Age
S_HDL_L
I3
SHAP value for S_HDL_L
S_HDL_L
I4
SHAP interaction values for S_HDL_L and Sex (Male)
S_HDL_L


J1
SHAP value for S_HDL_PL
S_HDL_PL
J2
SHAP interaction values for S_HDL_PL and Age
S_HDL_PL
J3
SHAP value for S_HDL_PL
S_HDL_PL
J4
SHAP interaction values for S_HDL_PL and Sex (Male)
S_HDL_PL
Age (Feature value) 40 50 60
Age (Feature value) 40 50 60
Sex (male): purple. Sex (female): yellow.
Sex (male): purple. Sex (female): yellow.
K1
SHAP value for M_VLDL_PL_pct
M_VLDL_PL_pct
K2
SHAP interaction values for M_VLDL_PL_pct and Age
M_VLDL_PL_pct
K3
SHAP value for M_VLDL_PL_pct
M_VLDL_PL_pct
K4
SHAP interaction values for M_VLDL_PL_pct and Sex (Male)
M_VLDL_PL_pct
Age (Feature value) 40 50 60
Age (Feature value) 40 50 60
Sex (male): purple. Sex (female): yellow.
Sex (male): purple. Sex (female): yellow.
L1
SHAP value for M_VLDL_C_pct
M_VLDL_C_pct
L2
SHAP interaction values for M_VLDL_C_pct and Age
M_VLDL_C_pct
L3
SHAP value for M_VLDL_C_pct
M_VLDL_C_pct
L4
SHAP interaction values for M_VLDL_C_pct and Sex (Male)
M_VLDL_C_pct
Age (Feature value) 40 50 60
Age (Feature value) 40 50 60
Sex (male): purple. Sex (female): yellow.
Sex (male): purple. Sex (female): yellow.


M1
SHAP value for M_VLDL_CE_pct
M_VLDL_CE_pct
M2
SHAP interaction values for M_VLDL_CE_pct and Age
M_VLDL_CE_pct
M3
SHAP value for M_VLDL_CE_pct
M_VLDL_CE_pct
M4
SHAP interaction values for M_VLDL_CE_pct and Sex (Male)
M_VLDL_CE_pct
Age (Feature value)
Sex (male): purple. Sex (female): yellow.
N1
SHAP value for M_VLDL_FC_pct
M_VLDL_FC_pct
N2
SHAP interaction values for M_VLDL_FC_pct and Age
M_VLDL_FC_pct
N3
SHAP value for M_VLDL_FC_pct
M_VLDL_FC_pct
N4
SHAP interaction values for M_VLDL_FC_pct and Sex (Male)
M_VLDL_FC_pct
O1
SHAP value for M_VLDL_TG_pct
M_VLDL_TG_pct
O2
SHAP interaction values for M_VLDL_TG_pct and Age
M_VLDL_TG_pct
O3
SHAP value for M_VLDL_TG_pct
M_VLDL_TG_pct
O4
SHAP interaction values for M_VLDL_TG_pct and Sex (Male)
M_VLDL_TG_pct

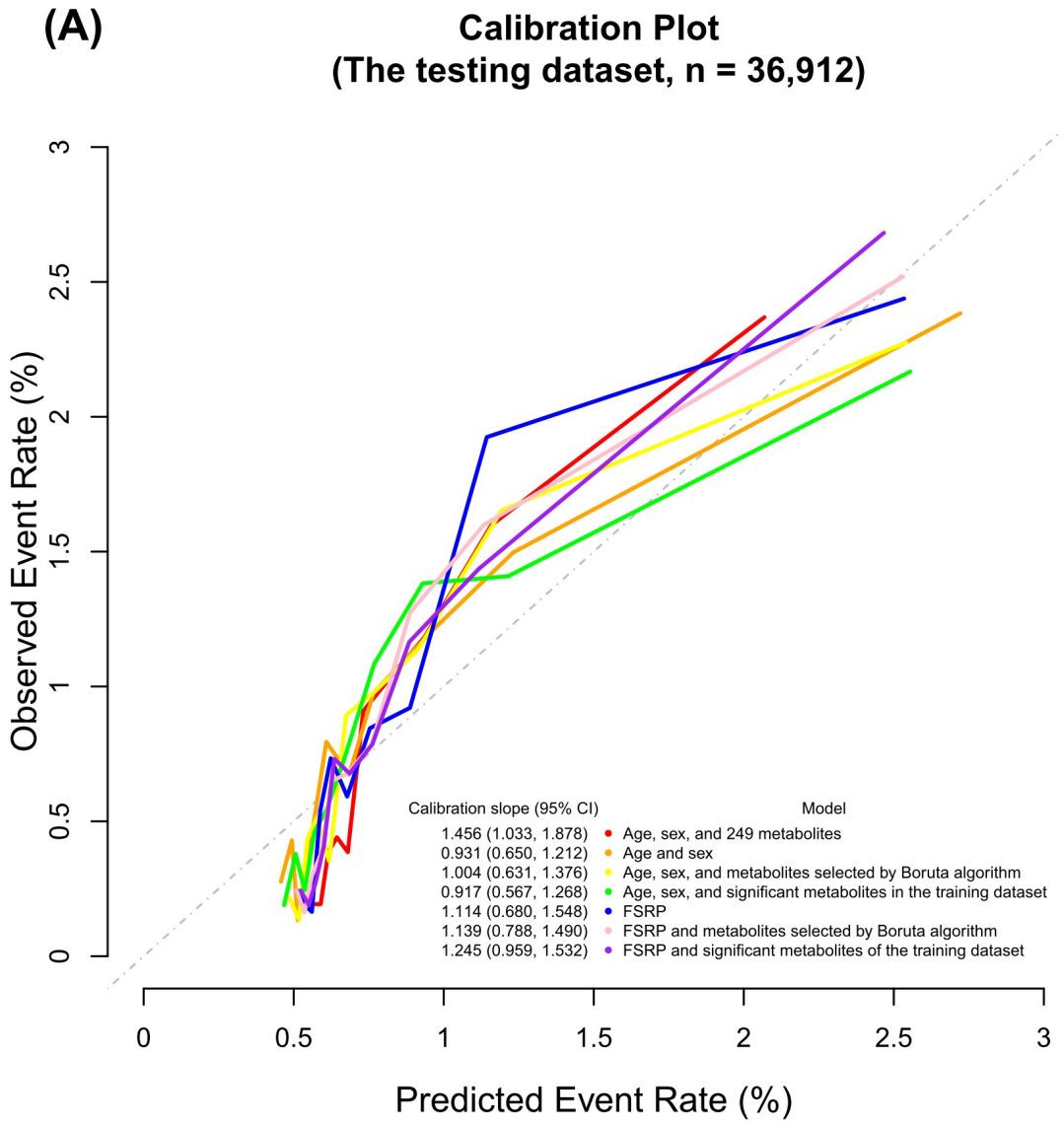


Sex (male): purple. Sex (female): yellow.

Sex (male): purple. Sex (female): yellow.

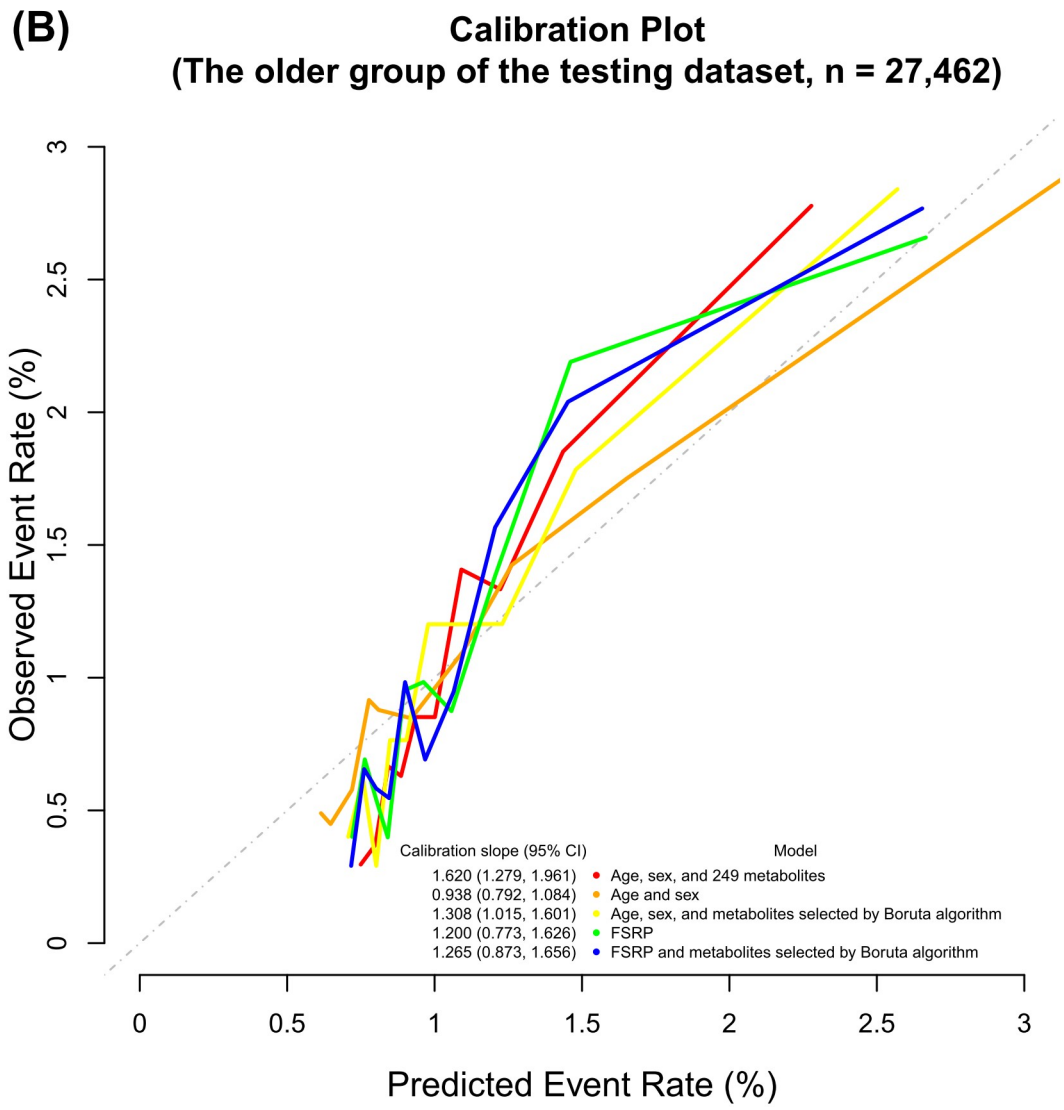
Supplementary Figure 5.2 SHAP dependence plots of the extreme gradient boosting model with age, sex, and significant metabolites

The SHAP value of a participant was displayed as each point on the y-axis corresponding to the feature on the x-axis. SHAP dependence plots (the 1st and 3rd ones) and interaction plots (the 2nd and 4th ones) were shown for the XGBoost-model 3 with age, sex and 15 significant metabolites in the training dataset (n = 147,652). In SHAP dependence plots, the SHAP value of a participant was displayed as each point on the y-axis corresponding to the metabolic value on the x-axis. In SHAP interaction plots, the SHAP interaction value of a participant was displayed as each point on the y-axis corresponding to the metabolic value on the x-axis. Abbreviations are in line with **Supplementary Table 3.1**.

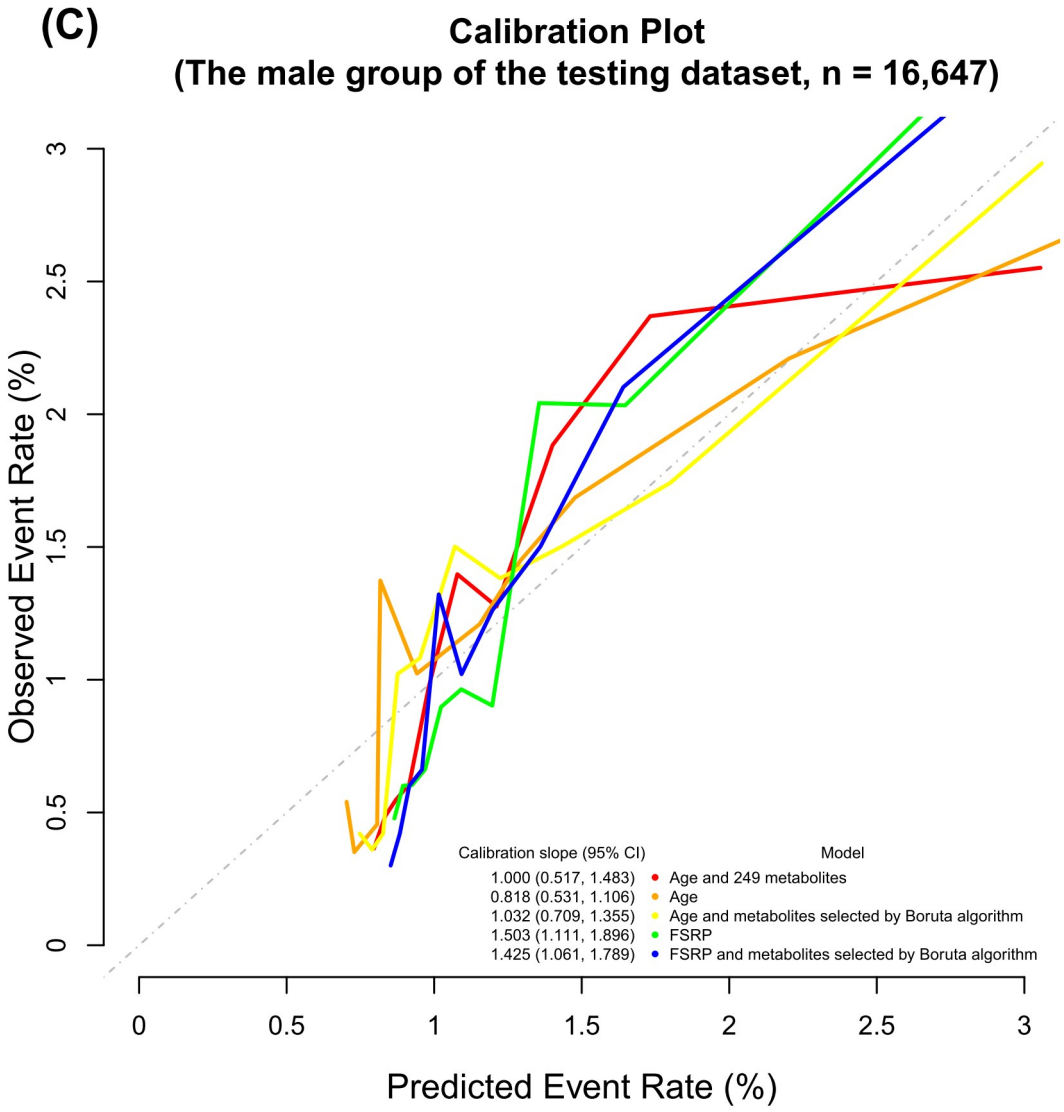


Supplementary Figure 5.3 Extreme gradient boosting model calibration

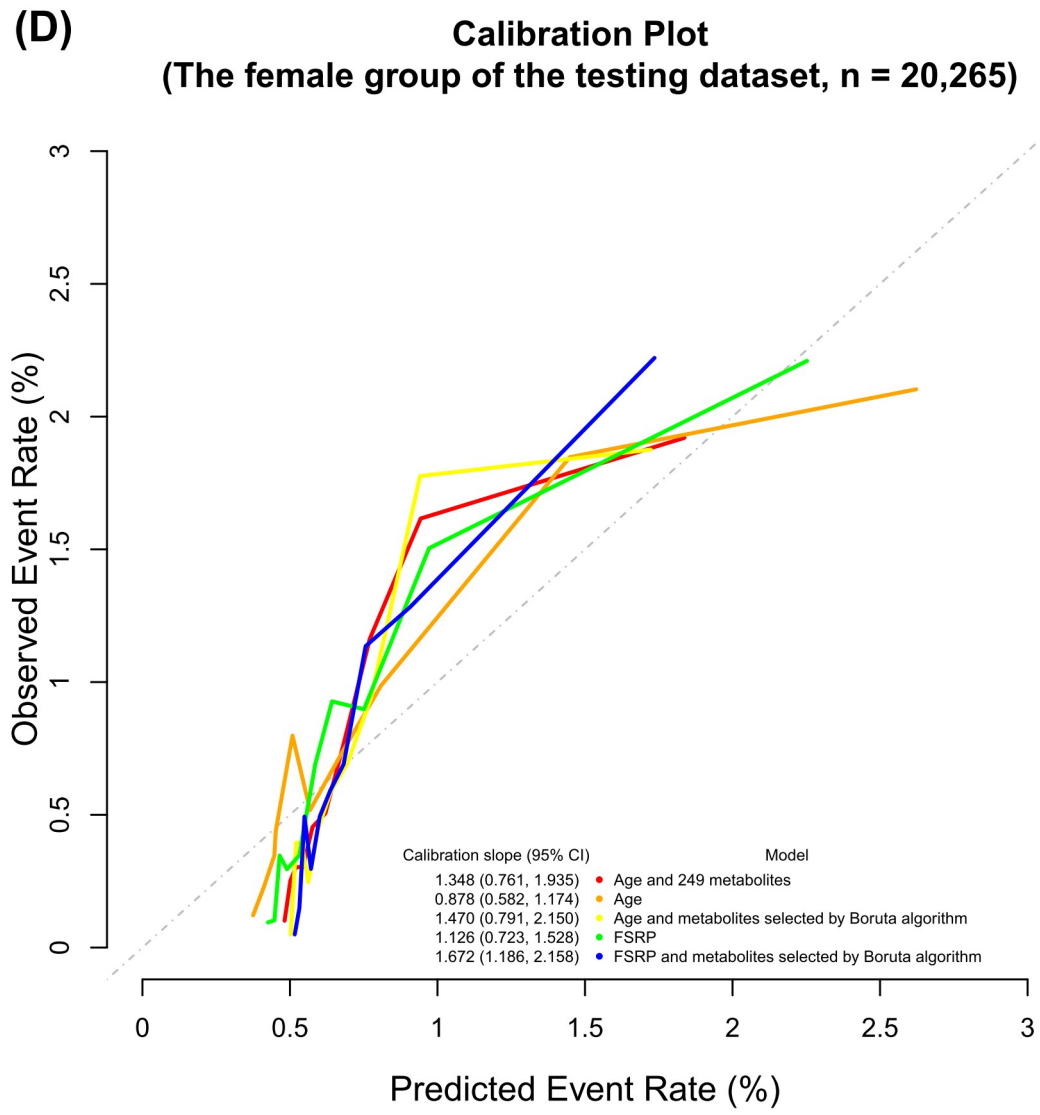
Calibration curves of risk prediction models for incident ischaemic stroke from cross-validation in (A) the overall population, (B) the older group, (C) the male group, and (D) the female group. For each model, the calibration of 10-year observed and predicted event rates are shown by 10 equally sized deciles of absolute predicted risk. Calibration slopes and intercepts are derived from a linear regression of the observed risk on the predicted risk. CI, confidence interval.



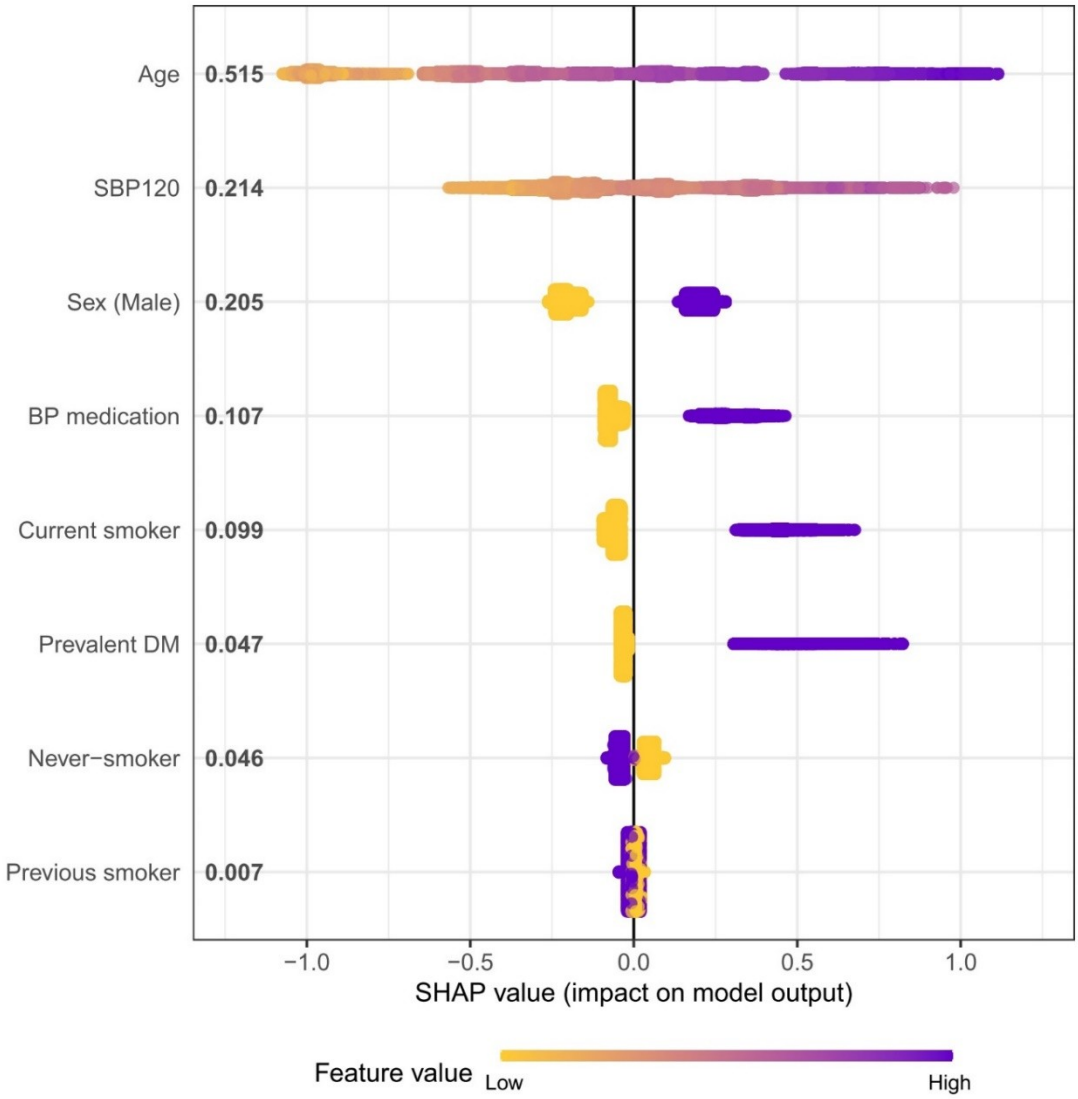
Supplementary Figure 5.3 (Continued) Extreme gradient boosting model calibration



Supplementary Figure 5.3 (Continued) Extreme gradient boosting model calibration



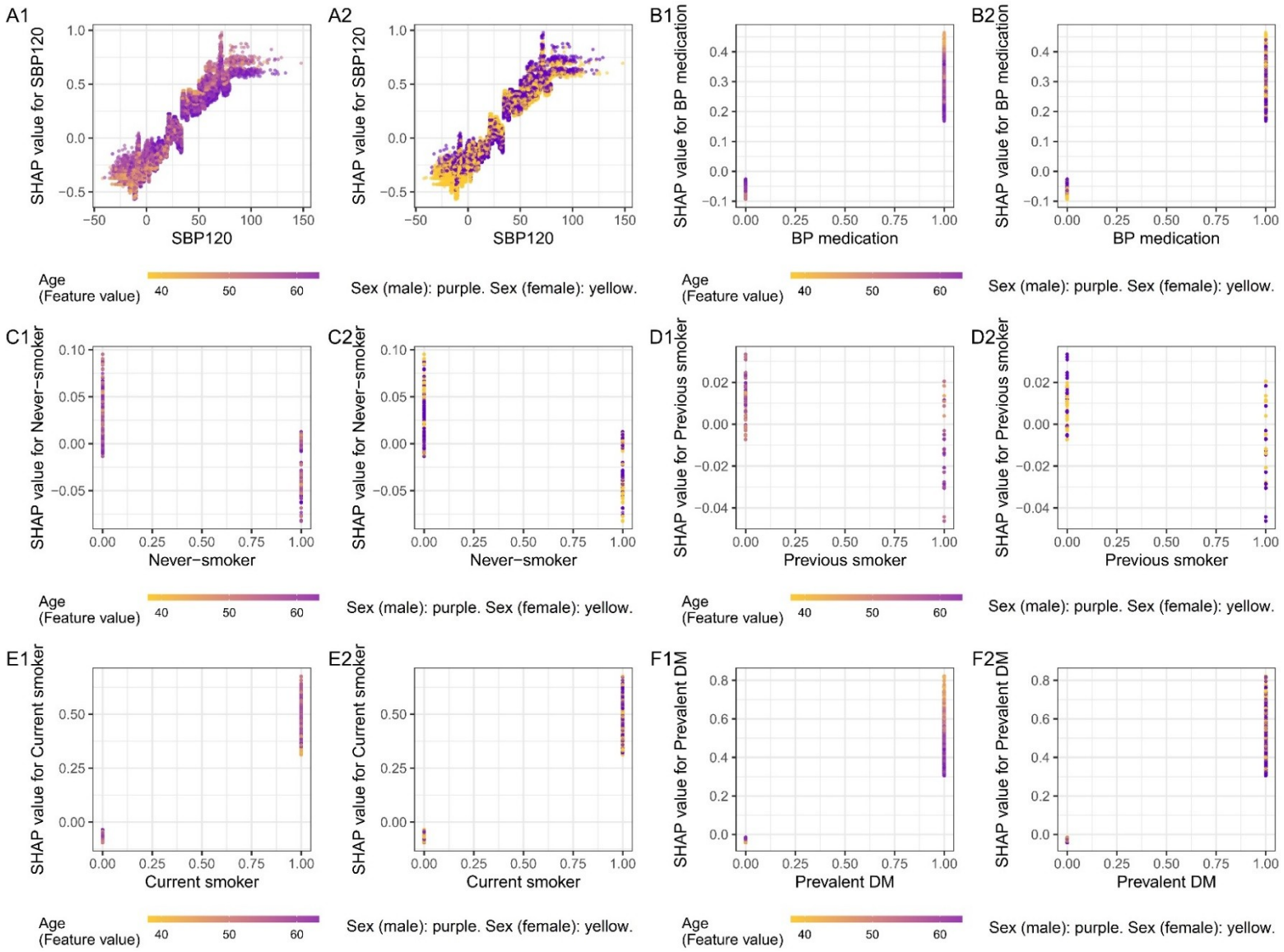
Supplementary Figure 5.3 (Continued) Extreme gradient boosting model calibration

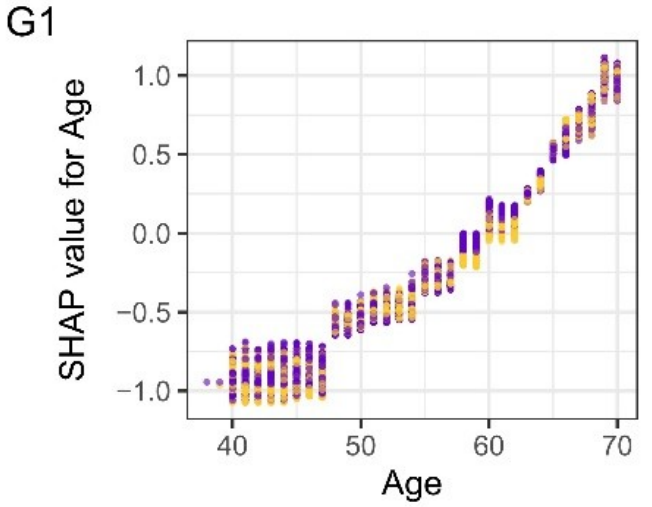


Supplementary Figure 5.4 The SHAP summary plot of extreme gradient boosting model with risk factors in Framingham Stroke Risk Profile

The plot illustrates the global interpretability with how much each feature contributes, either positively or negatively, to the prediction of incident ischaemic stroke. A positive association with ischaemic stroke was depicted as SHAP values above 0, and a negative association with ischaemic stroke was depicted as SHAP values below 0 (i.e., the greater the SHAP value the greater the hazard ratio for ischaemic stroke. It also illustrates the local interpretability with each dot corresponding to a participant in the training dataset (n = 147,652). The dot's position on the x axis indicates the impact that feature has on the model's incident ischaemic stroke prediction for that participant.

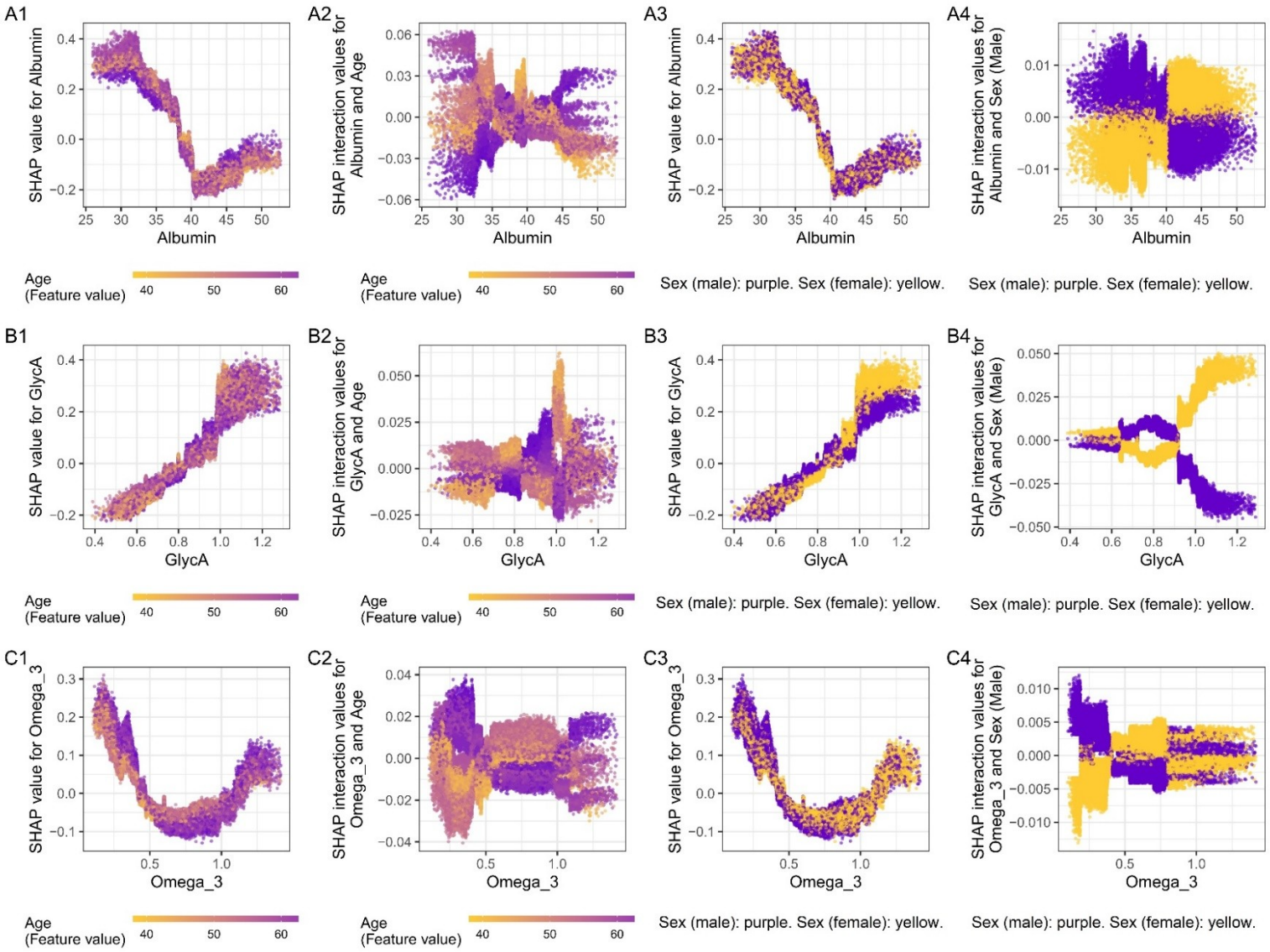
When numerous dots are located at the same x position, they are aggregated to reveal density. The bold figures on the left side indicate the mean of absolute SHAP values of the features. SBP120, systolic blood pressure – 120 mmHg. BP medication, blood pressure medication. Prevalent DM, prevalent diabetes mellitus.

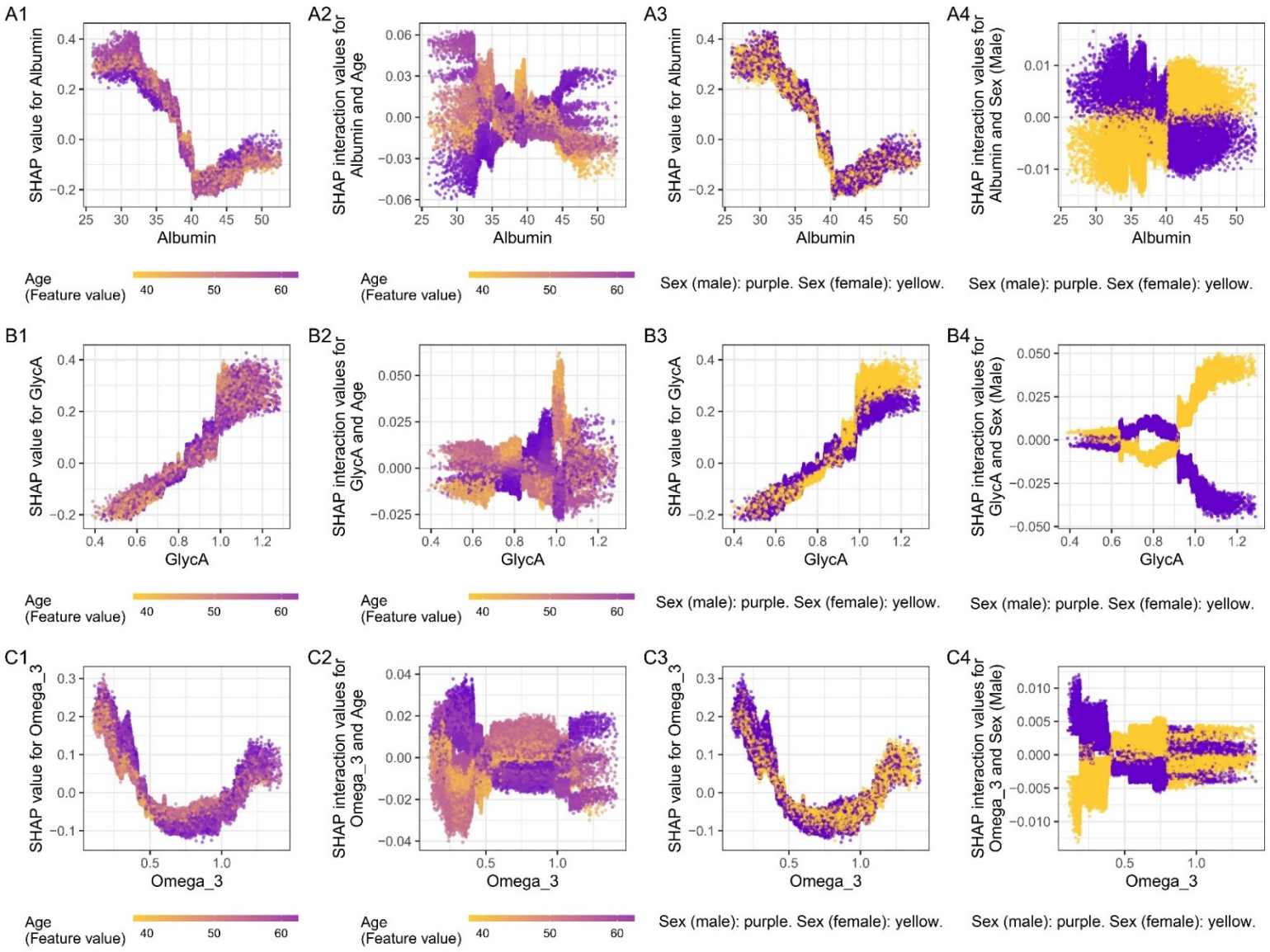

A1
SHAP value for SBP120
SBP120
A2
SHAP value for SBP120
SBP120
B1
SHAP value for BP medication
BP medication
B2
SHAP value for BP medication
BP medication
Age
(Feature value)
40
50
60
Sex (male): purple. Sex (female): yellow.
C1
SHAP value for Never−smoker
Never−smoker
C2
SHAP value for Never−smoker
Never−smoker
D1
SHAP value for Previous smoker
Previous smoker
D2
SHAP value for Previous smoker
Previous smoker
E1
SHAP value for Current smoker
Current smoker
E2
SHAP value for Current smoker
Current smoker
F1
SHAP value for Prevalent DM
Prevalent DM
F2
SHAP value for Prevalent DM
Prevalent DM

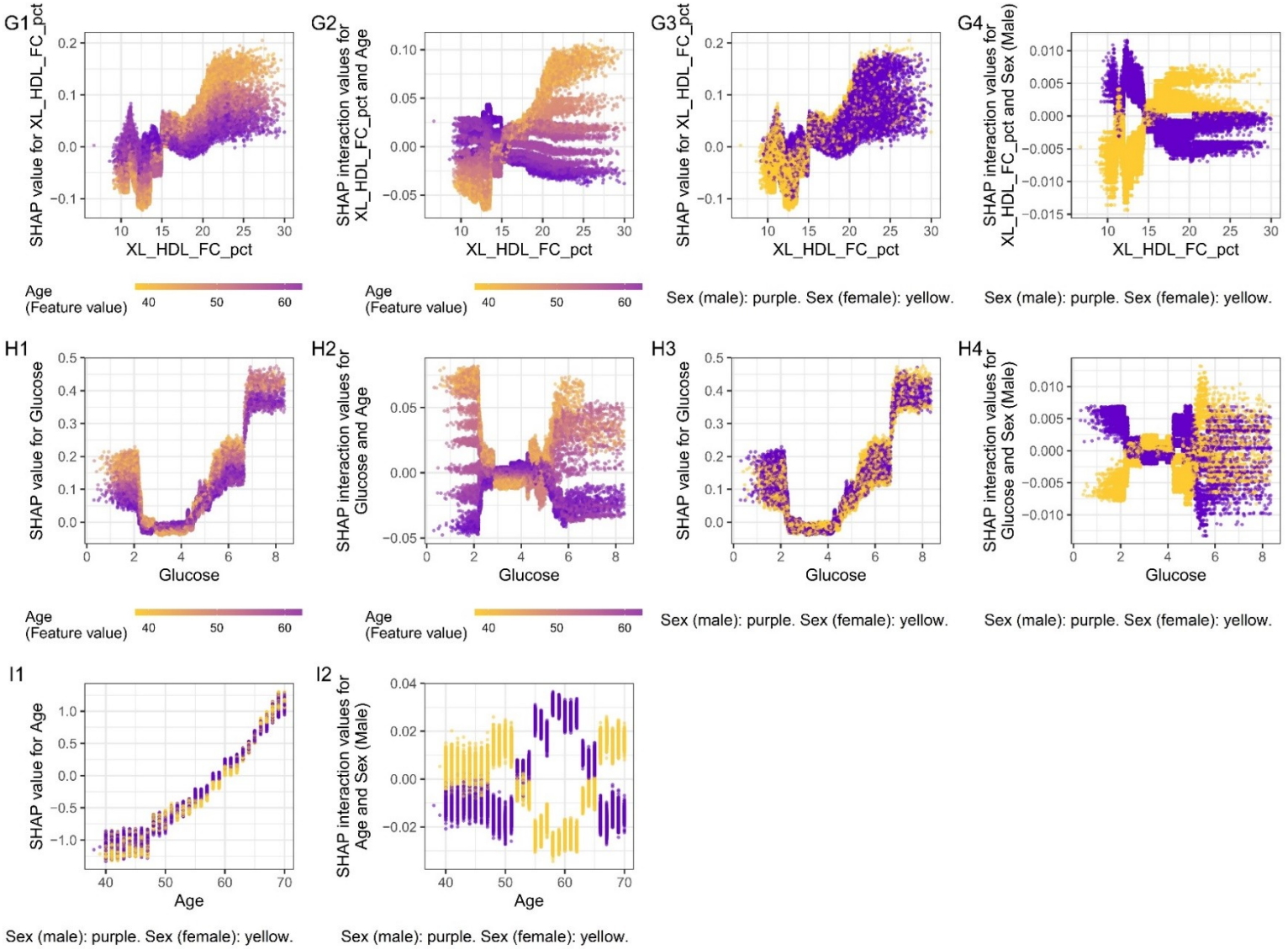


Supplementary Figure 5.5 SHAP dependence plots of the extreme gradient boosting model with risk factors in Framingham Stroke Risk Profile

The SHAP value of a participant was displayed as each point on the y-axis corresponding to the feature on the x-axis. SBP120, systolic blood pressure – 120 mmHg. BP medication, blood pressure medication. Prevalent DM, prevalent diabetes mellitus.

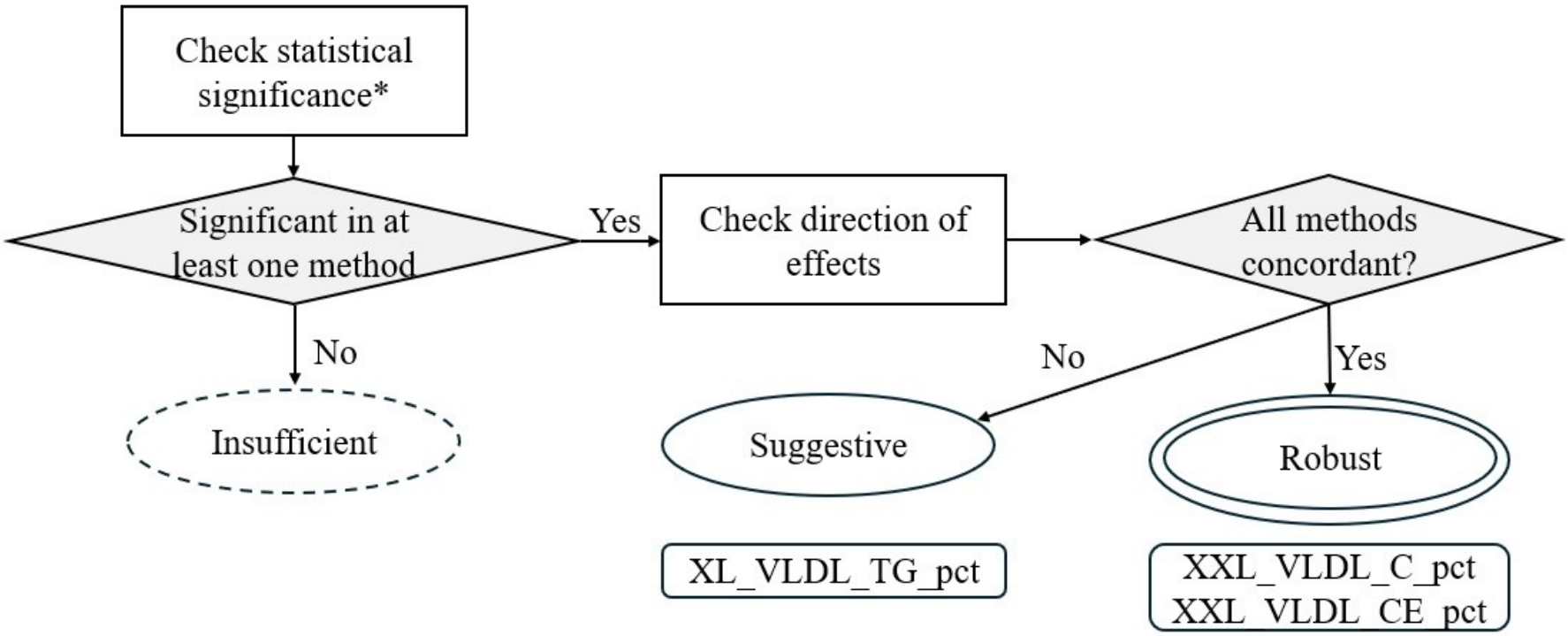

A1
SHAP value for Albumin
Albumin
Age
(Feature value)
40
50
60
A2
SHAP interaction values for Albumin and Age
Albumin
Age
(Feature value)
A3
SHAP value for Albumin
Albumin
Sex (male): purple. Sex (female): yellow.
A4
SHAP interaction values for Albumin and Sex (Male)
Albumin
Sex (male): purple. Sex (female): yellow.
B1
SHAP value for GlycA
GlycA
B2
SHAP interaction values for GlycA and Age
GlycA
B3
SHAP value for GlycA
GlycA
Sex (male): purple. Sex (female): yellow.
B4
SHAP interaction values for GlycA and Sex (Male)
GlycA
Sex (male): purple. Sex (female): yellow.
C1
SHAP value for Omega_3
Omega_3
C2
SHAP interaction values for Omega_3 and Age
Omega_3
C3
SHAP value for Omega_3
Omega_3
Sex (male): purple. Sex (female): yellow.
C4
SHAP interaction values for Omega_3 and Sex (Male)
Omega_3
Sex (male): purple. Sex (female): yellow.


A1
SHAP value for Albumin
Albumin
Age
(Feature value)
40
50
60
A2
SHAP interaction values for Albumin and Age
Albumin
Age
(Feature value)
A3
SHAP value for Albumin
Albumin
Sex (male): purple. Sex (female): yellow.
A4
SHAP interaction values for Albumin and Sex (Male)
Albumin
Sex (male): purple. Sex (female): yellow.
B1
SHAP value for GlycA
GlycA
B2
SHAP interaction values for GlycA and Age
GlycA
B3
SHAP value for GlycA
GlycA
B4
SHAP interaction values for GlycA and Sex (Male)
GlycA
C1
SHAP value for Omega_3
Omega_3
C2
SHAP interaction values for Omega_3 and Age
Omega_3
C3
SHAP value for Omega_3
Omega_3
C4
SHAP interaction values for Omega_3 and Sex (Male)
Omega_3

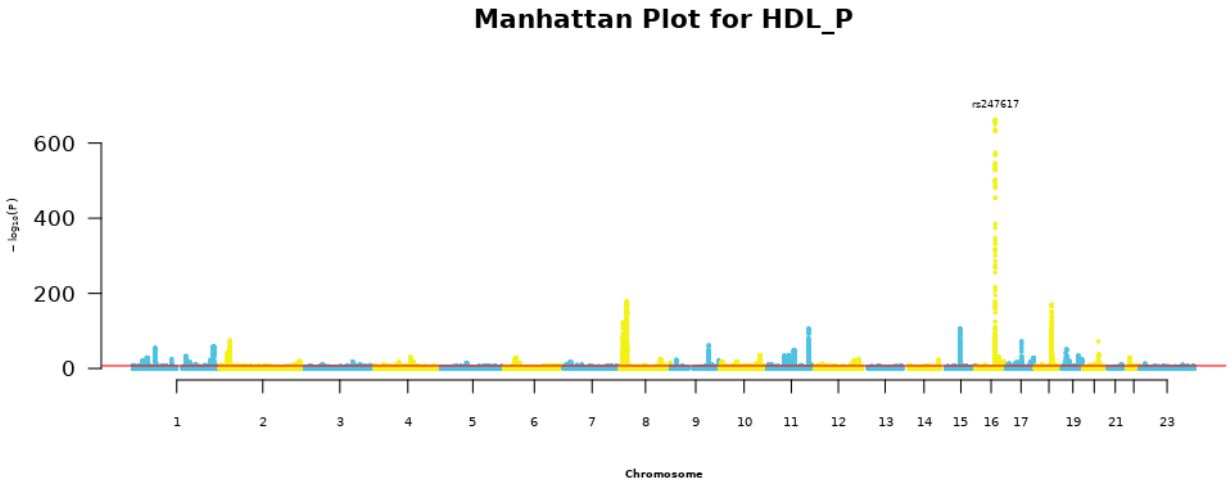
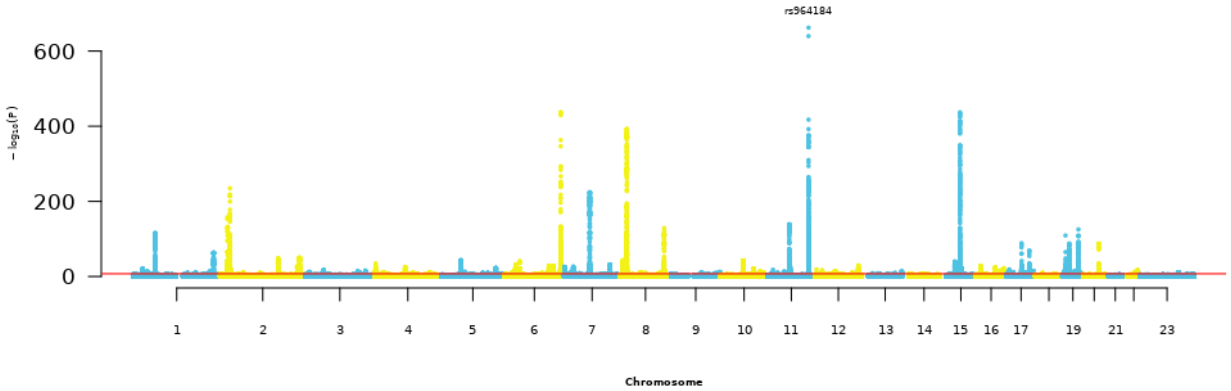

G1
SHAP value for XL_HDL_FC_pct
XL_HDL_FC_pct
Age
(Feature value)
40
50
60
G2
SHAP interaction values for XL_HDL_FC_pct and Age
XL_HDL_FC_pct
Age
(Feature value)
40
50
60
G3
SHAP value for XL_HDL_FC_pct
XL_HDL_FC_pct
Sex (male): purple. Sex (female): yellow.
G4
SHAP interaction values for XL_HDL_FC_pct and Sex (Male)
XL_HDL_FC_pct
Sex (male): purple. Sex (female): yellow.
H1
SHAP value for Glucose
Glucose
Age
(Feature value)
40
50
60
H2
SHAP interaction values for Glucose and Age
Glucose
Age
(Feature value)
40
50
60
H3
SHAP value for Glucose
Glucose
Sex (male): purple. Sex (female): yellow.
H4
SHAP interaction values for Glucose and Sex (Male)
Glucose
Sex (male): purple. Sex (female): yellow.
I1
SHAP value for Age
Age
Sex (male): purple. Sex (female): yellow.
I2
SHAP interaction values for Age and Sex (Male)
Age
Sex (male): purple. Sex (female): yellow.

Supplementary Figure 5.6 SHAP dependence plots and interaction plots of the extreme gradient boosting model with age, sex and 8 metabolites.

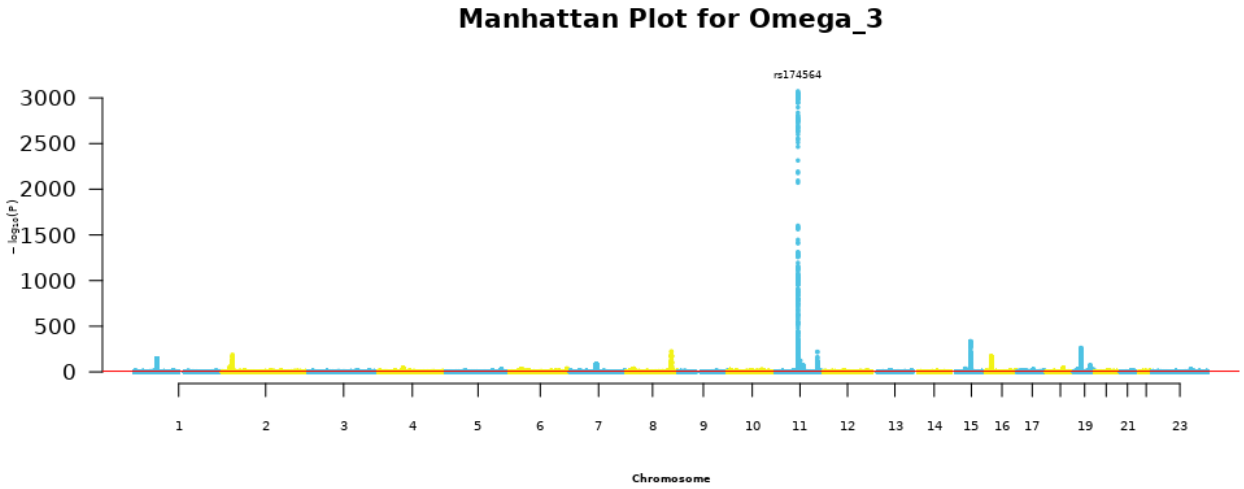
SHAP dependence plots (the 1st and 3rd ones) and interaction plots (the 2nd and 4th ones) were shown for the XGBoost-model 2 with age, sex and 8 metabolites in the training dataset (n = 147,652). In SHAP dependence plots, the SHAP value of a participant was displayed as each point on the y-axis corresponding to the metabolic value on the x-axis. In SHAP interaction plots, the SHAP interaction value of a participant was displayed as each point on the y-axis corresponding to the metabolic value on the x-axis. GlycA, glycoprotein acetyls. Omega_3, omega-3 fatty acids. LA_pct, ratio of linoleic acid to total fatty acids. PUFA_pct, ratio of monounsaturated fatty acids to total fatty acids. XL_HDL_FC_pct, free cholesterol to total lipids ratio in very large HDL.



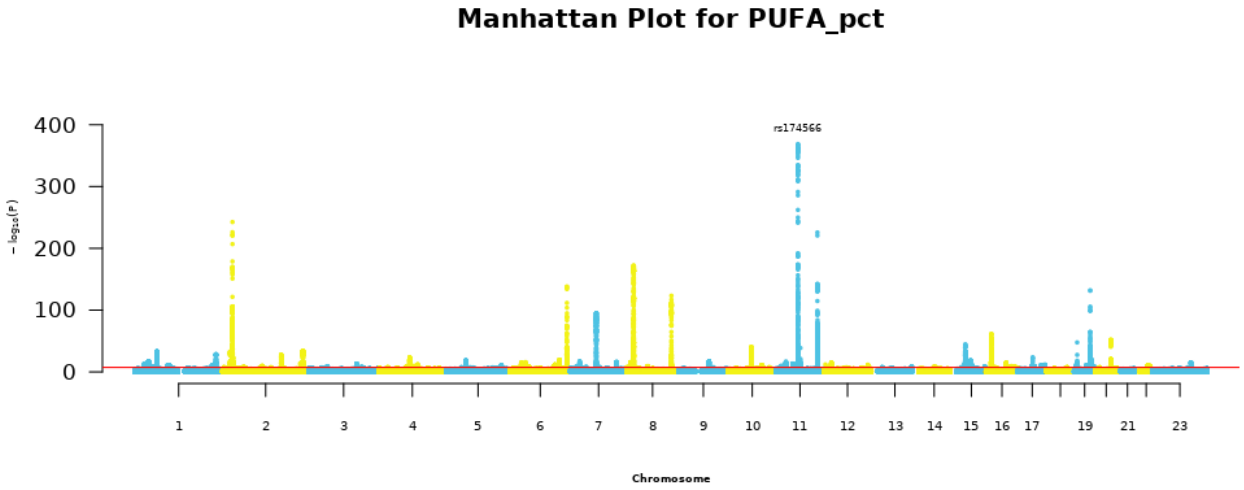
Supplementary Figure 6.1 A framework for assessing causal relationships between metabolites and ischaemic stroke.
MR, Mendelian randomisation. MR-PRESSO, Mendelian randomisation pleiotropy residual sum and outlier. Abbreviations are in line with **Supplementary Table 3.1**. Modified from: Markozannes et al. (2022) *BMC Medicine*.
*At least one of the performed methods (i.e. inverse variance-weighted method, MR-Egger, weighted median estimator, the simple mode estimator, weighted mode estimator, and MR-PRESSO) for the specific association presented a significant p value after adjusting for multiple testing.

ID	Metabolite	Manhattan plot
1	HDL_P	
2	VLDL_size	

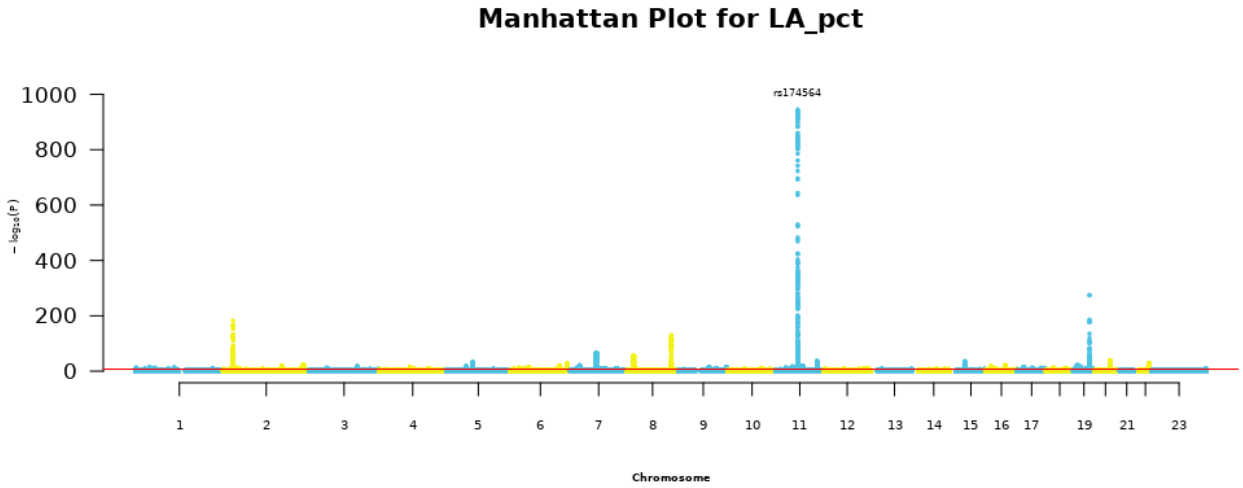
3 Omega_3



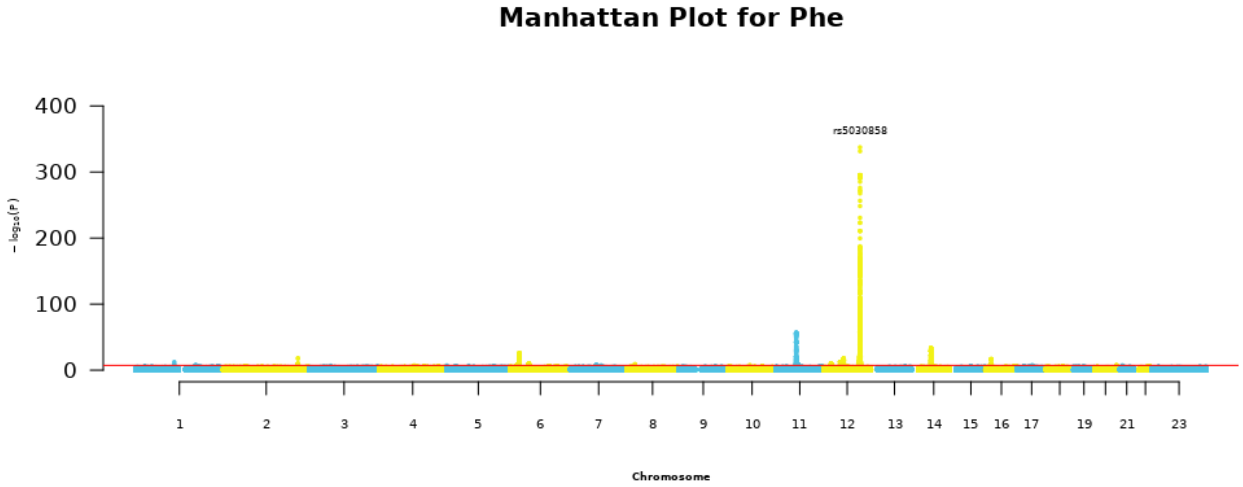
4 PUFA_pct



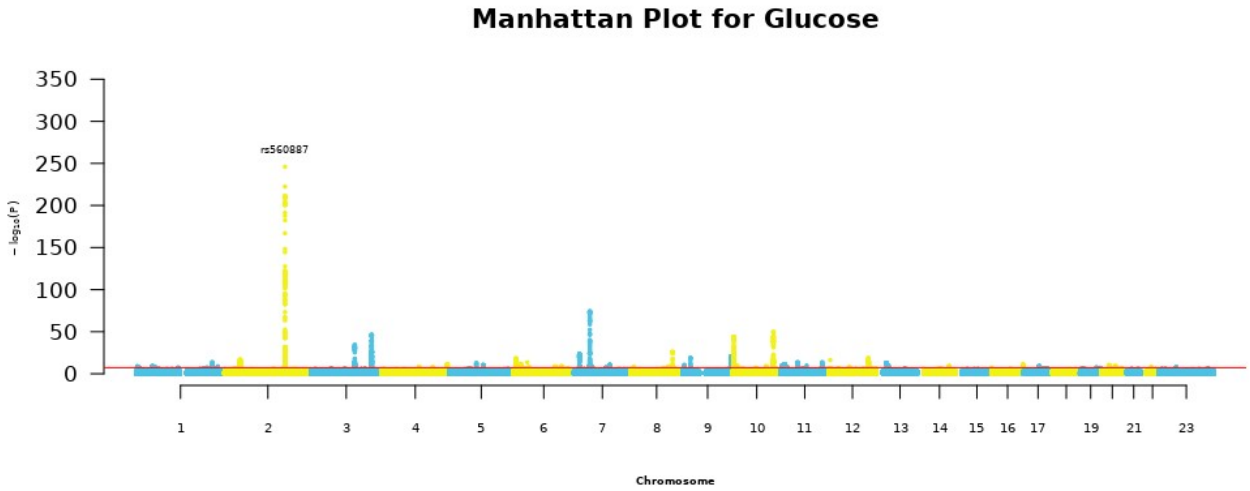
5 LA_pct


Manhattan Plot for LA_pct
rs174564
1000
800
600
400
200
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
19
21
23
Chromosome

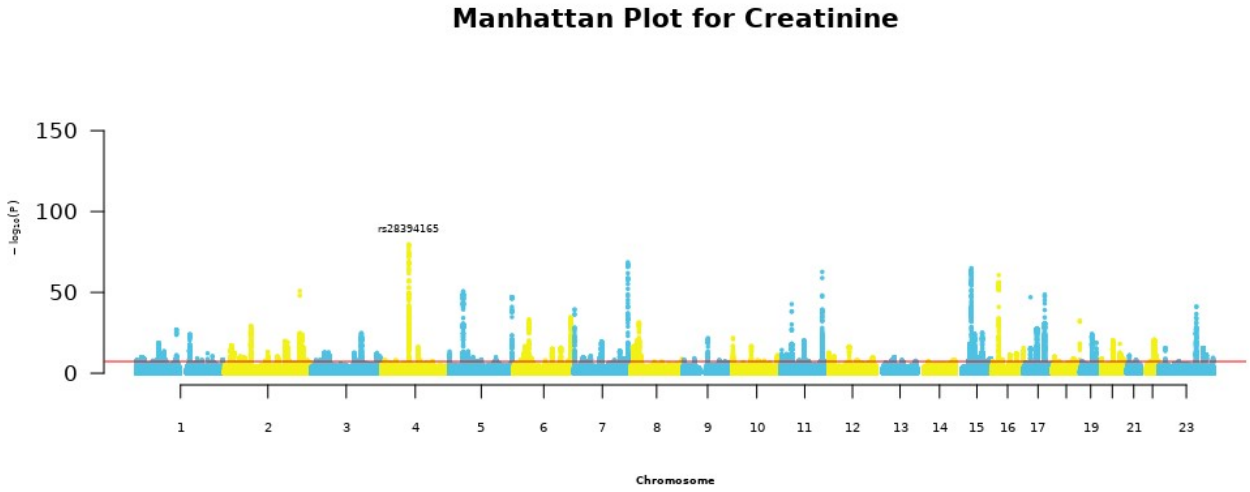
6 Phe


Manhattan Plot for Phe
rs5030858
400
300
200
100
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
19
21
23
Chromosome

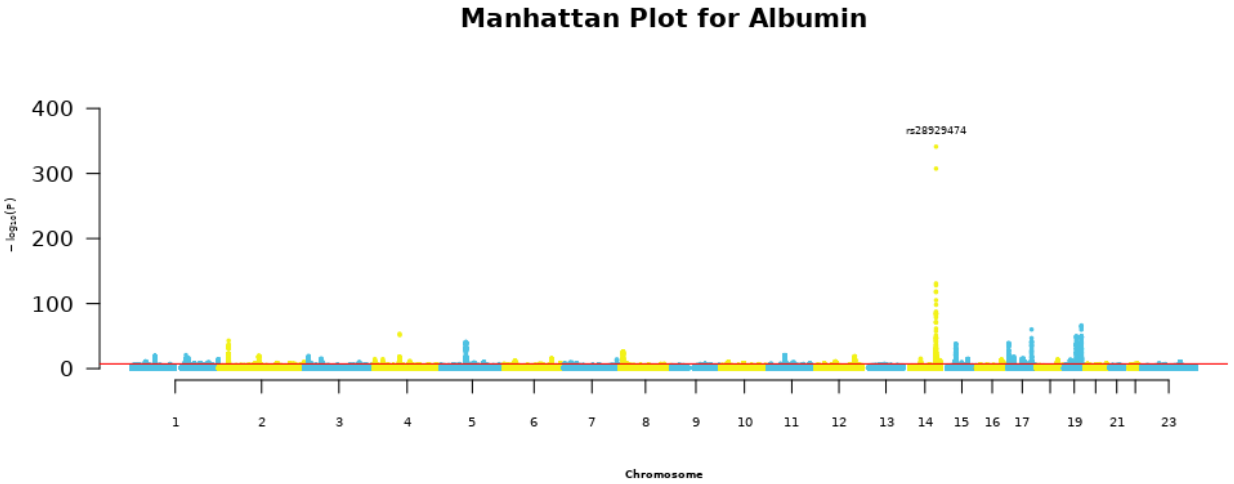
7 Glucose



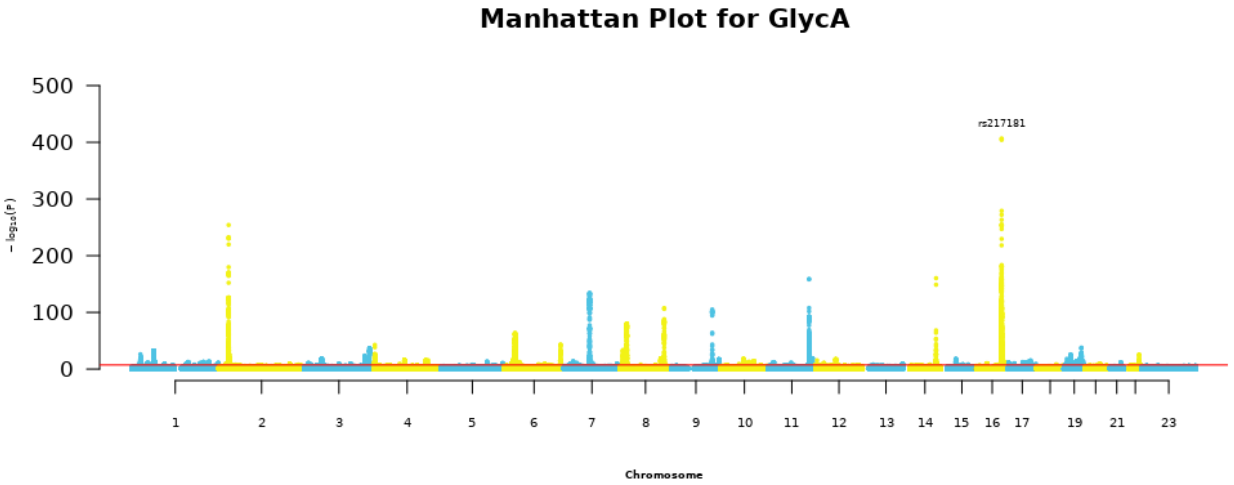
8 Creatinine



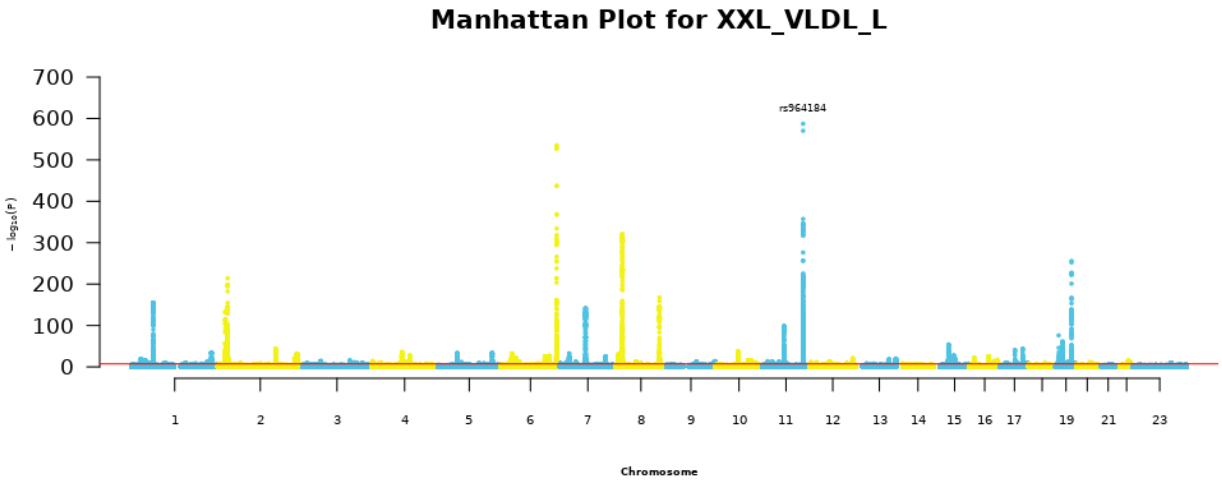
9 Albumin



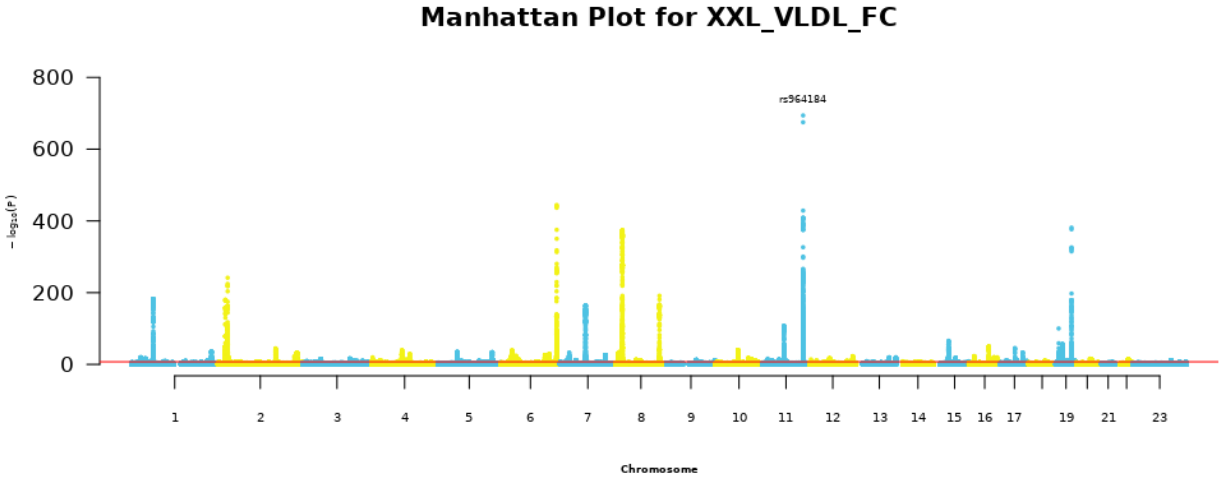
10 GlycA



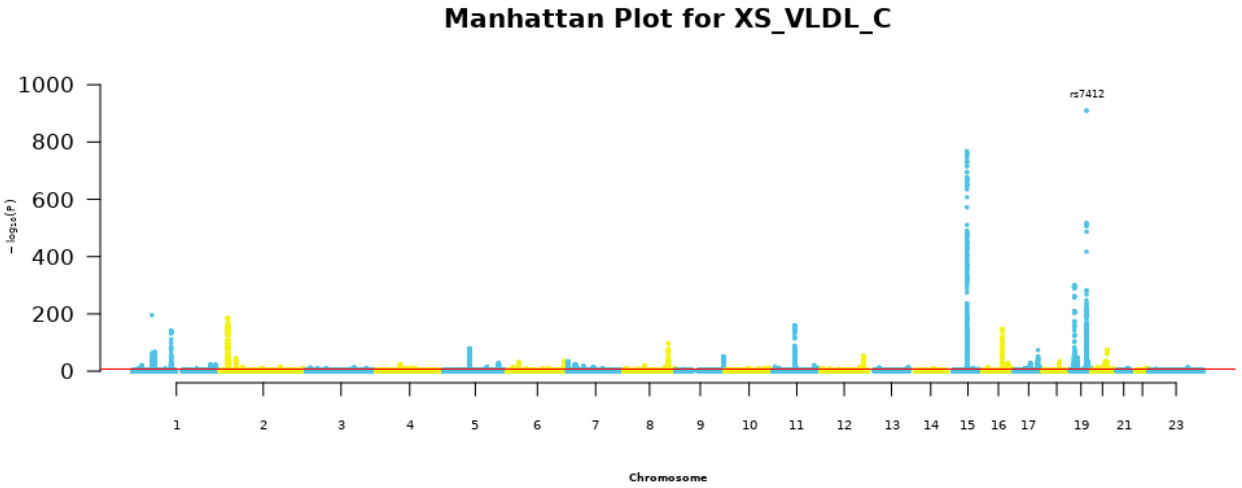
11 XXL_VLDL_L



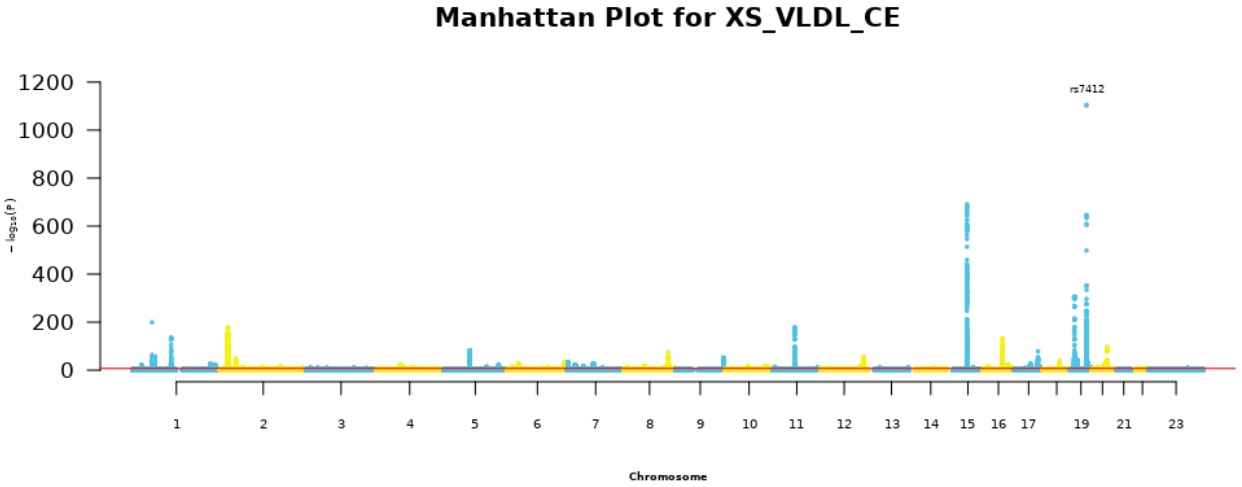
12 XXL_VLDL_FC



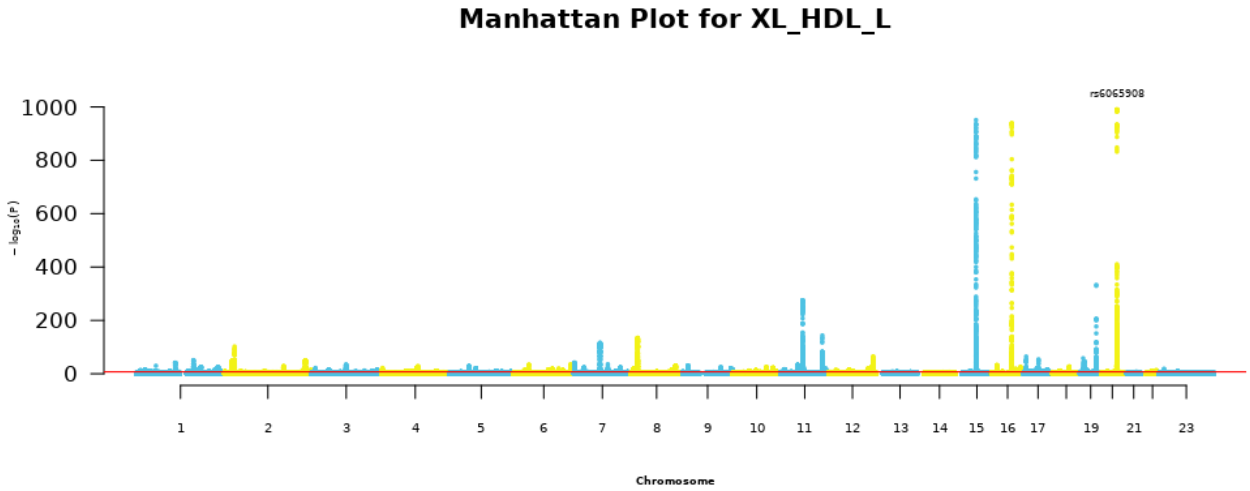
13 XS_VLDL_C



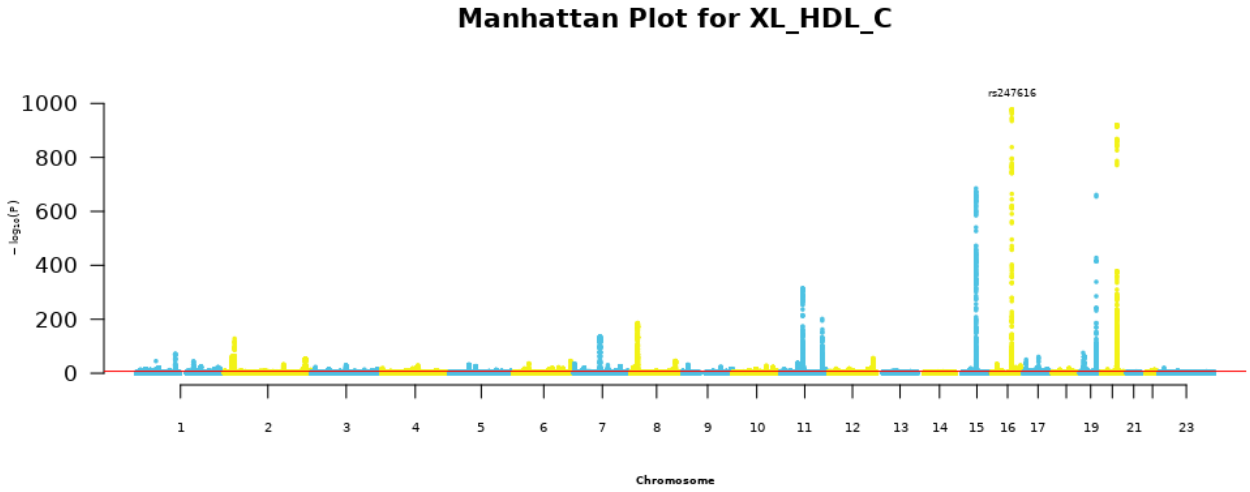
14 XS_VLDL_CE



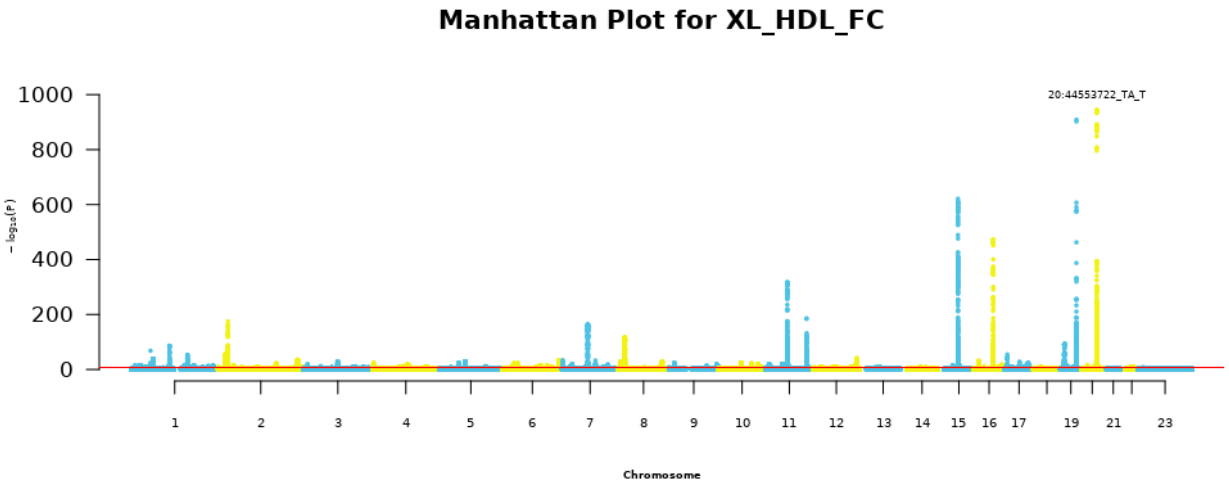
15 XL_HDL_L



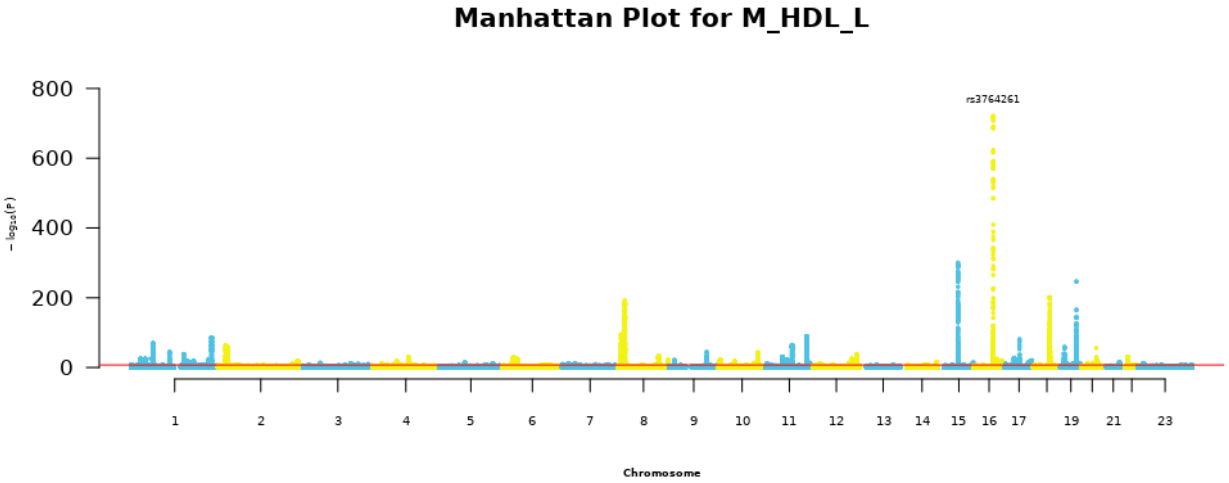
16 XL_HDL_C



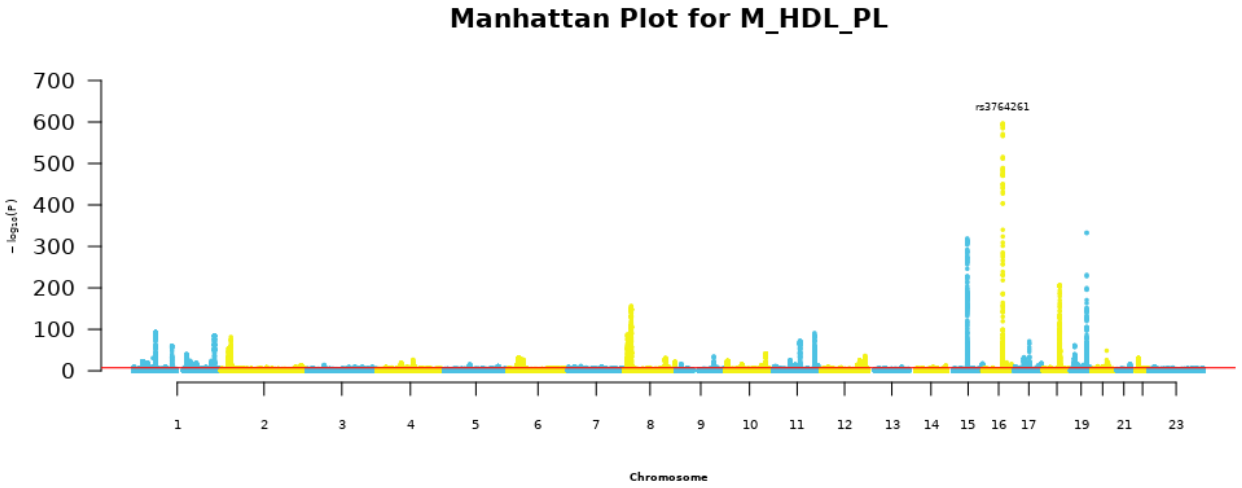
17 XL_HDL_FC



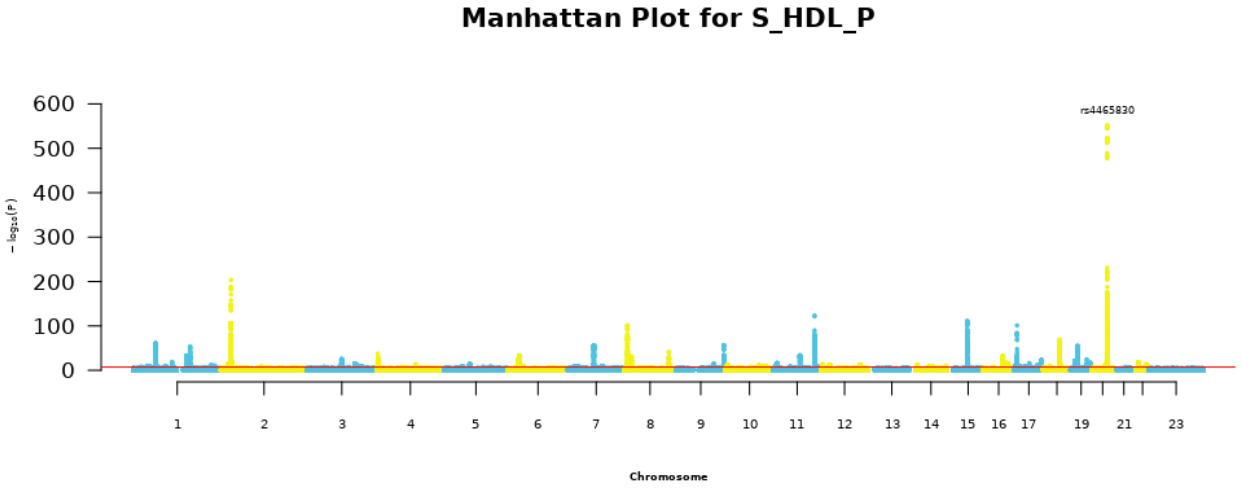
18 M_HDL_L



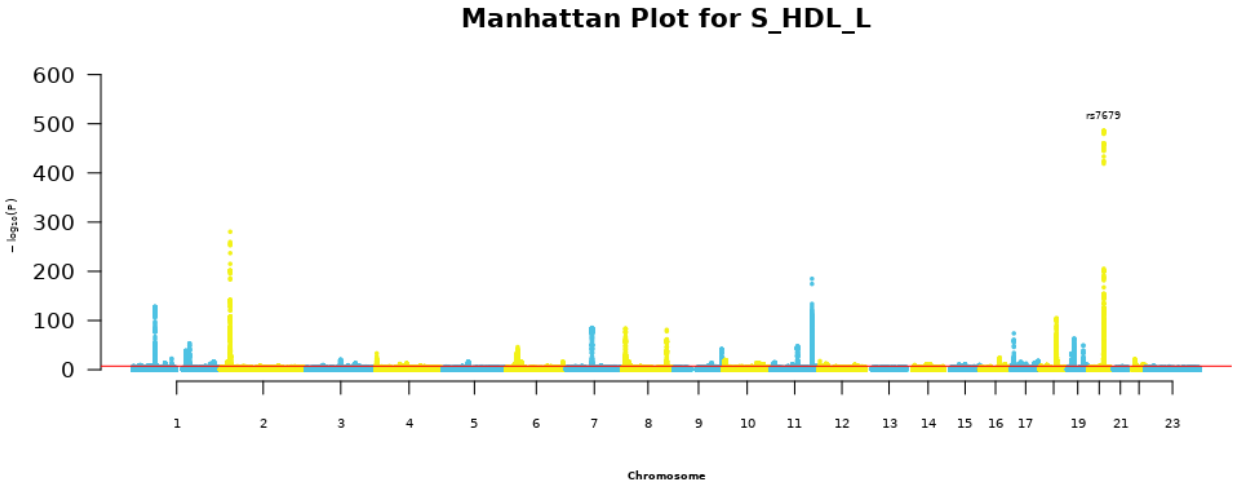
19 M_HDL_PL



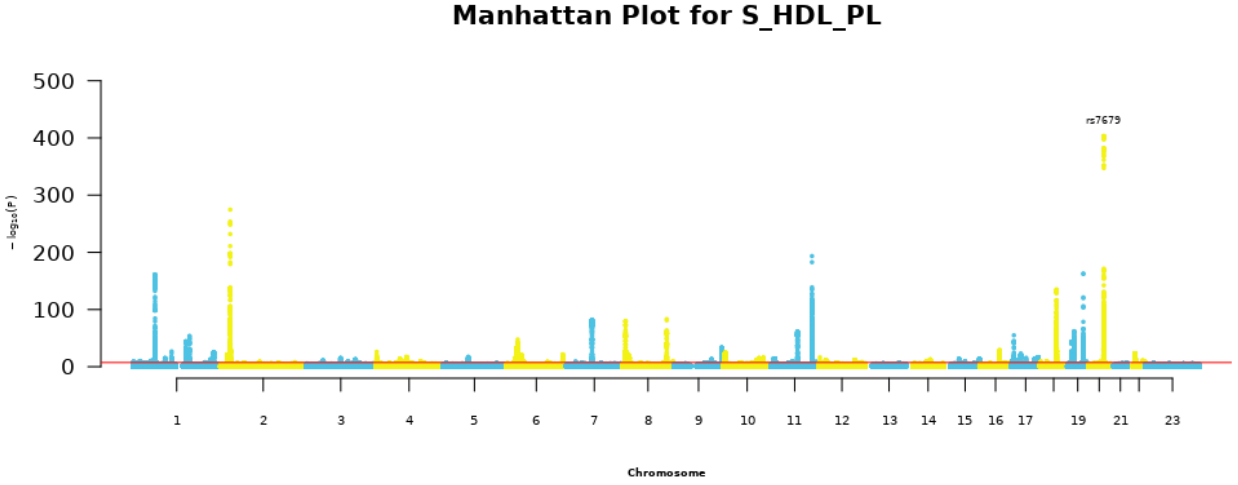
20 S_HDL_P



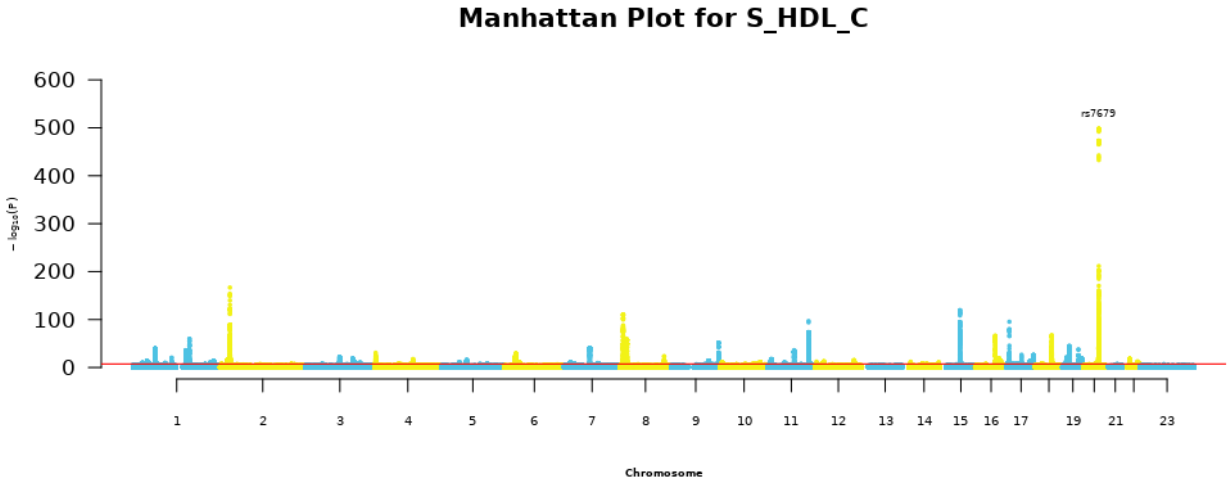
21 S_HDL_L



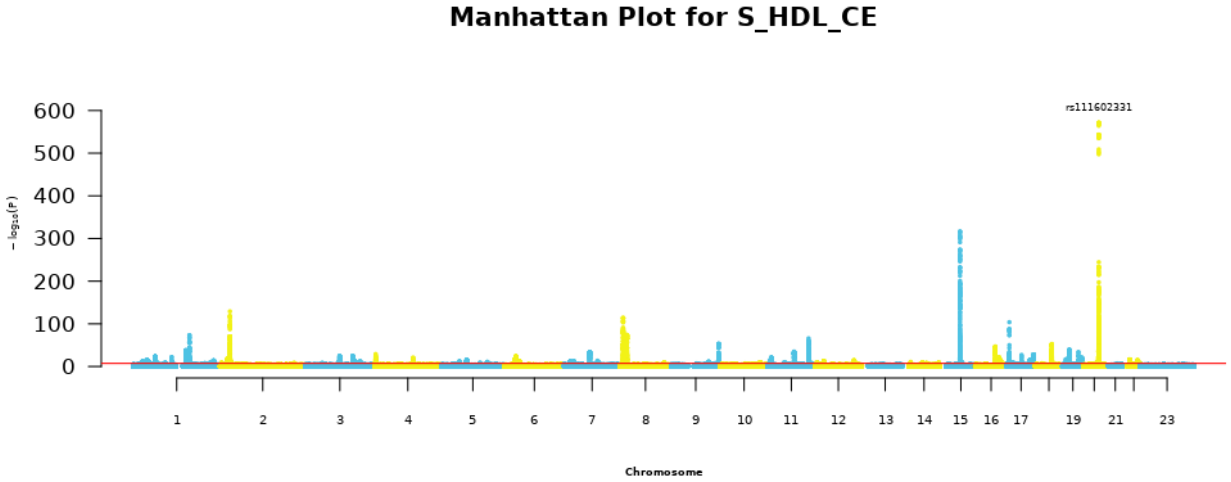
22 S_HDL_PL



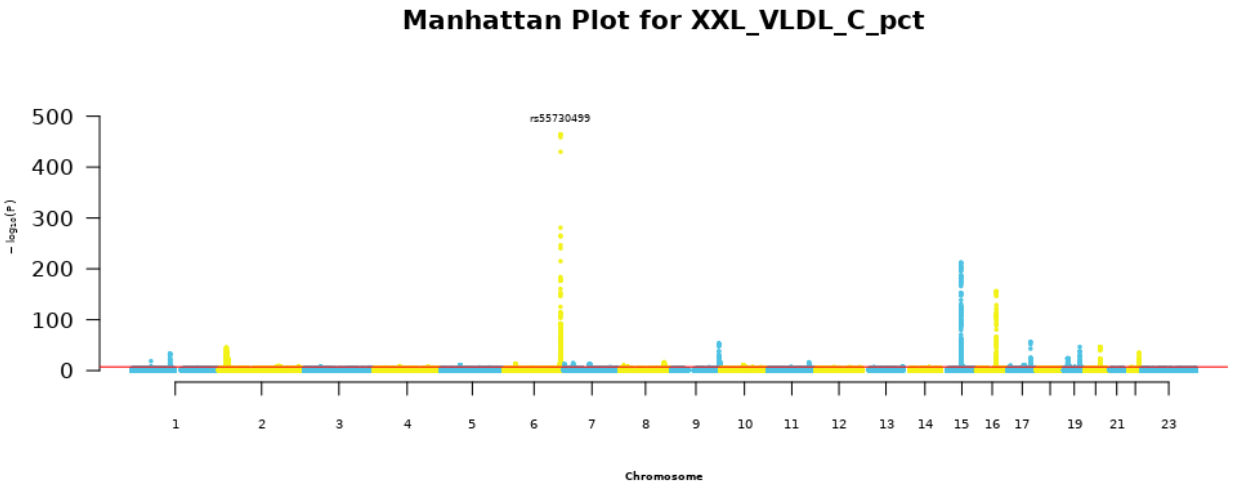
23 S_HDL_C



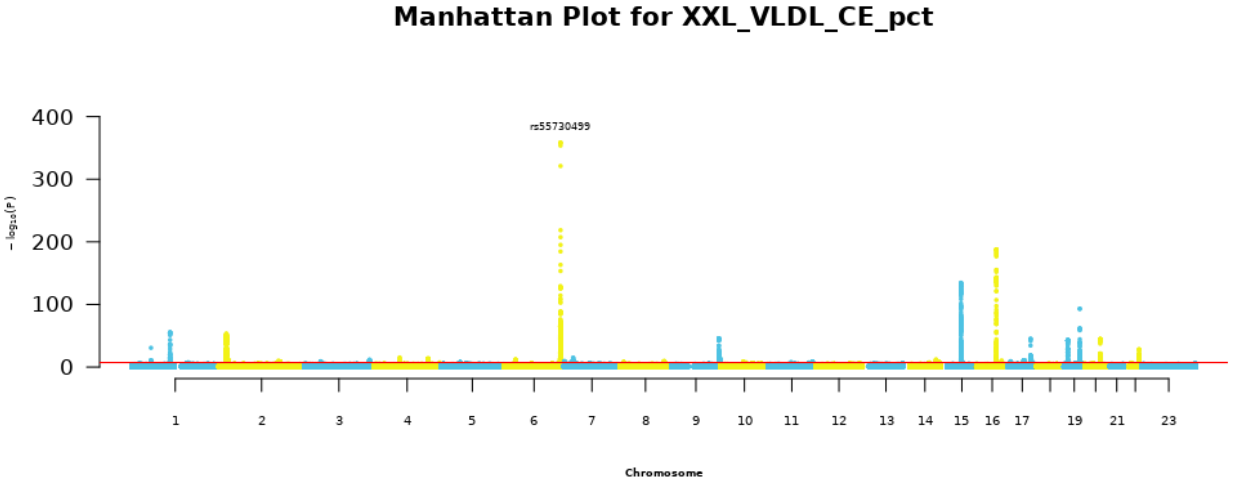
24 S_HDL_CE



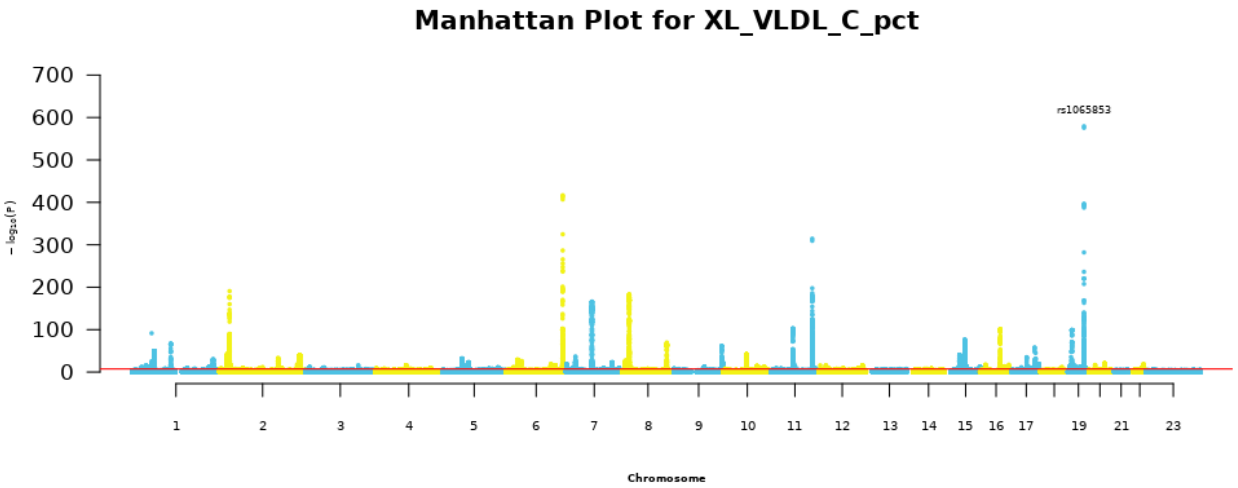
25 XXL_VLDL_C_pct



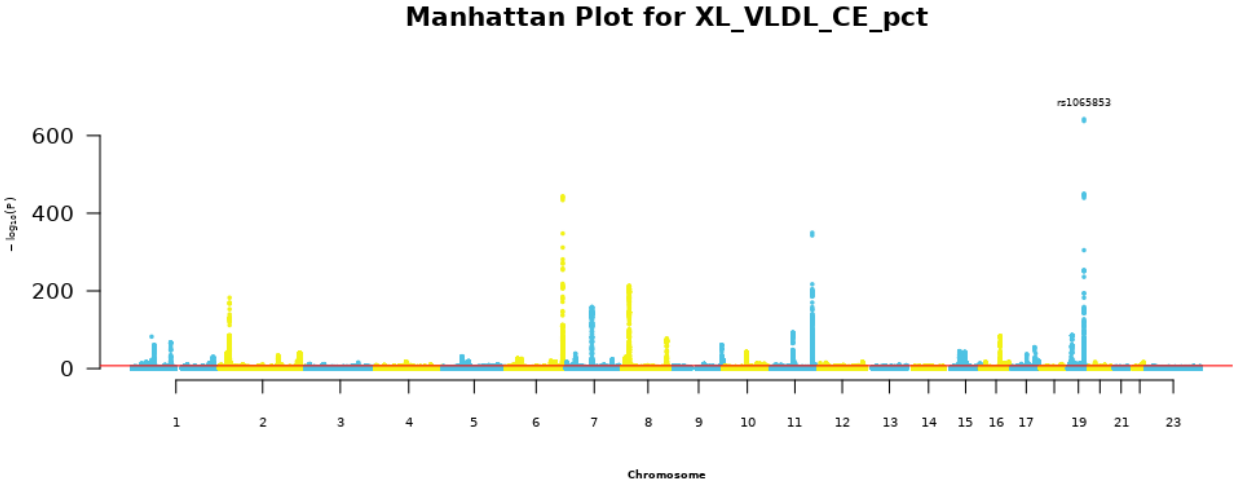
26 XXL_VLDL_CE_pct



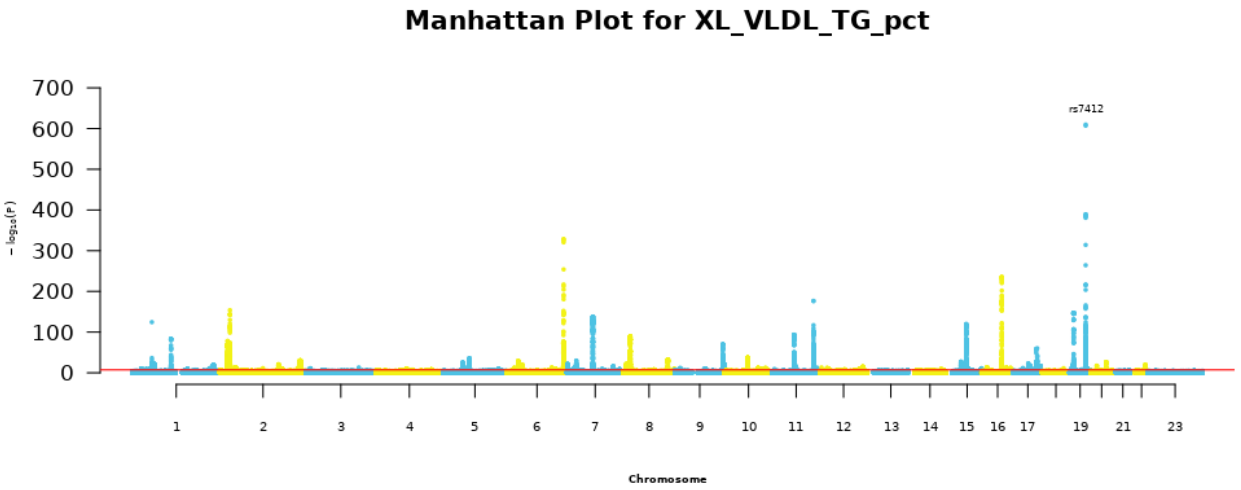
27 XL_VLDL_C_pct



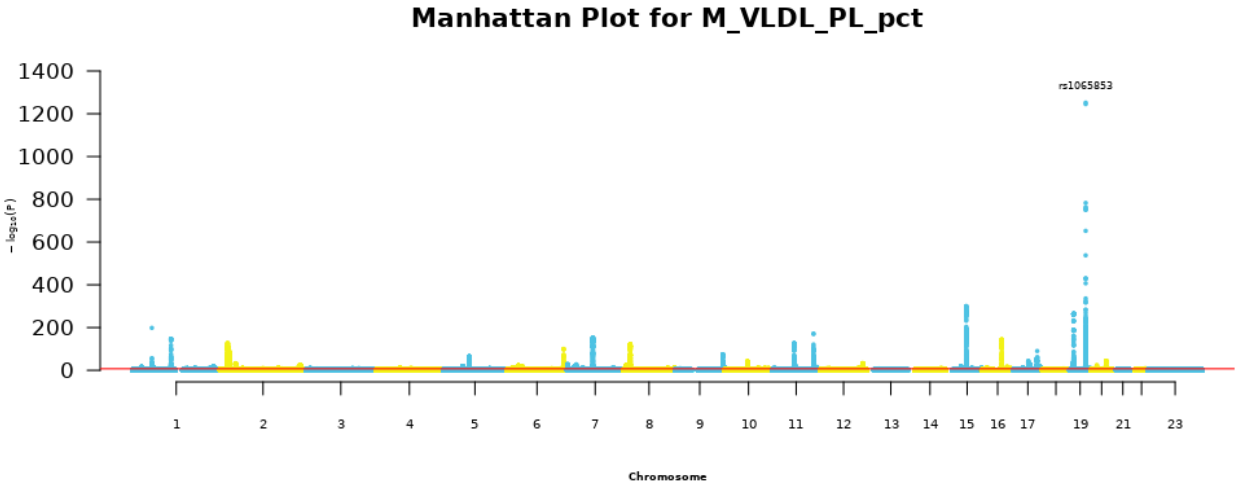
28 XL_VLDL_CE_pct



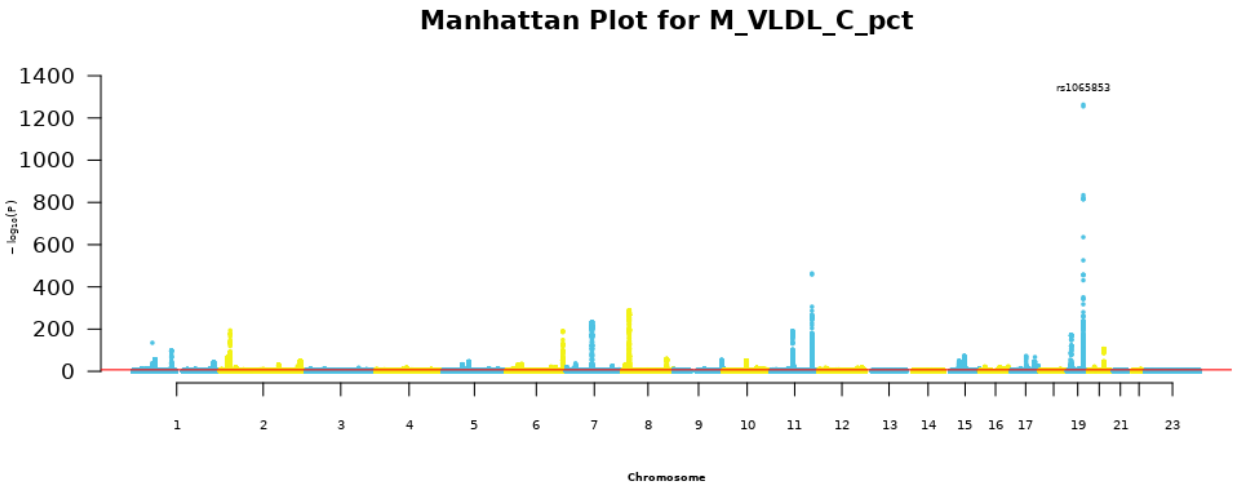
29 XL_VLDL_TG_pct



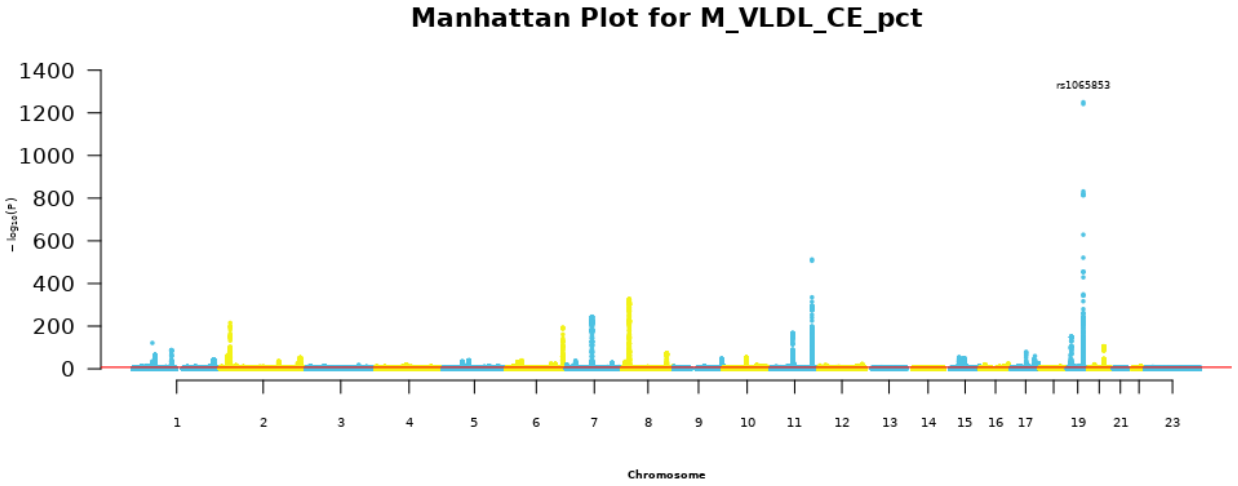
30 M_VLDL_PL_pct



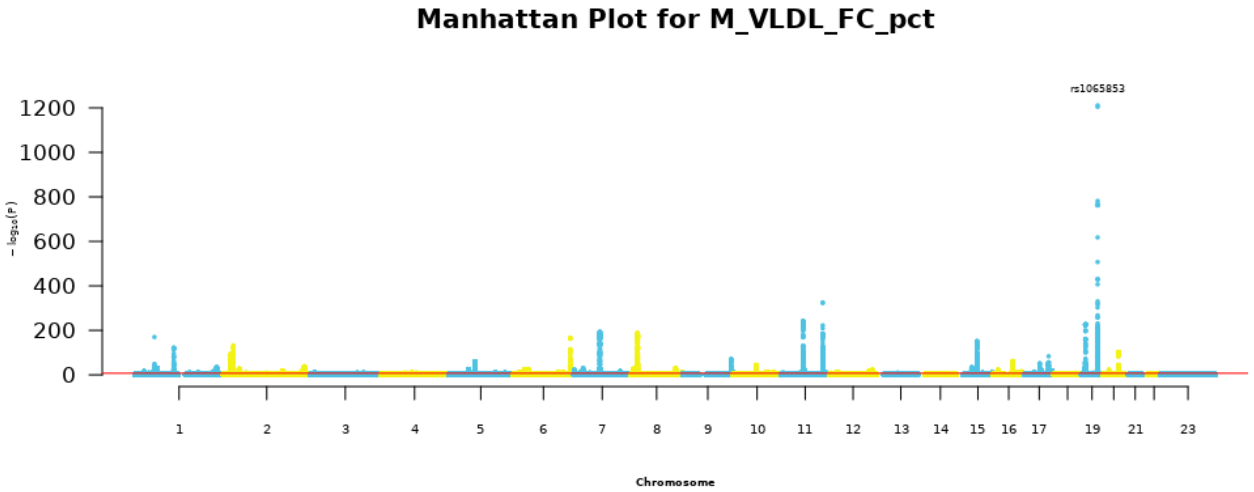
31 M_VLDL_C_pct



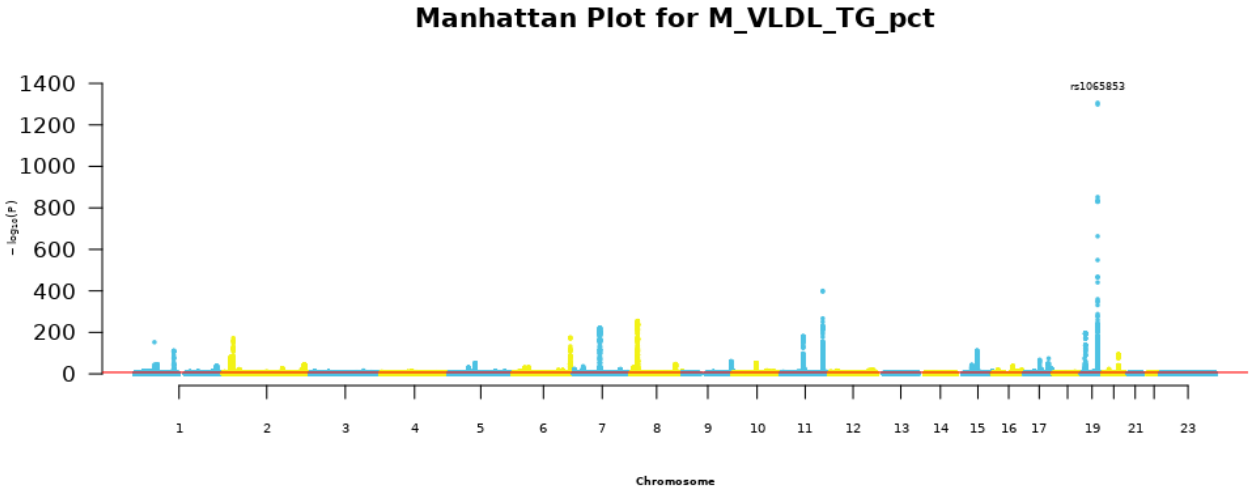
32 M_VLDL_CE_pct



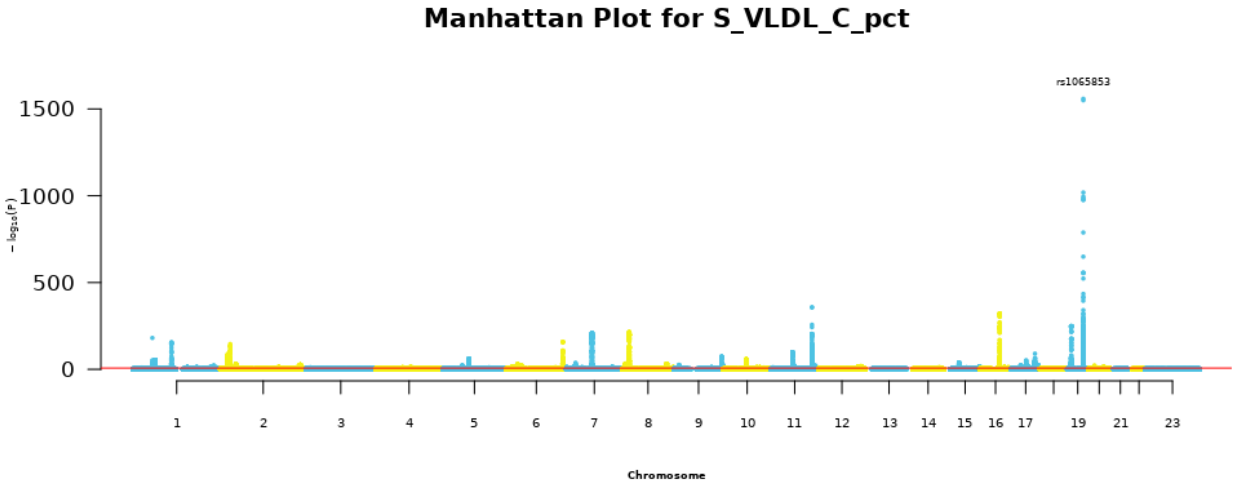
33 M_VLDL_FC_pct



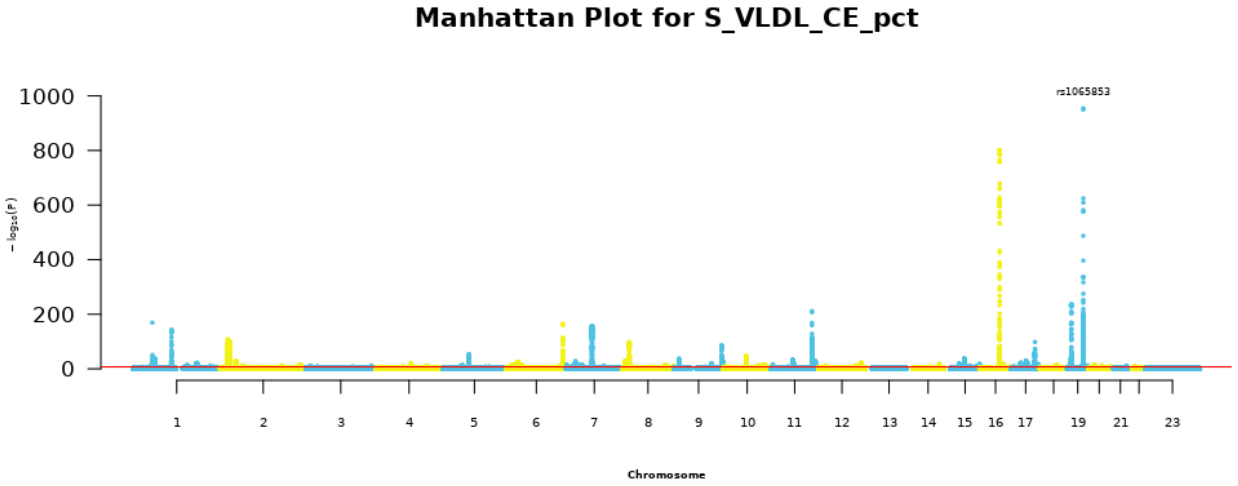
34 M_VLDL_TG_pct



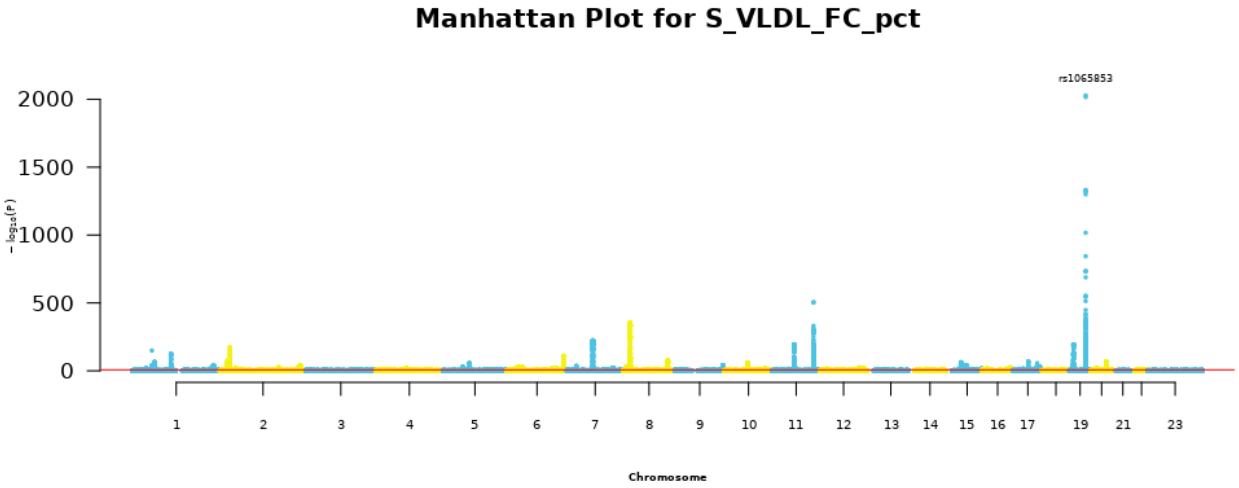
35 S_VLDL_C_pct



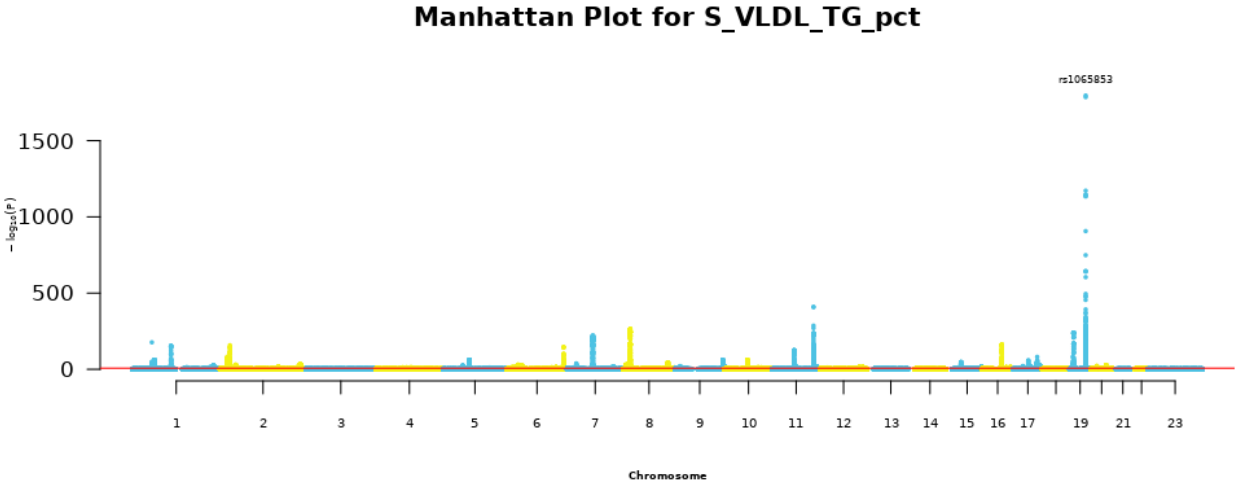
36 S_VLDL_CE_pct



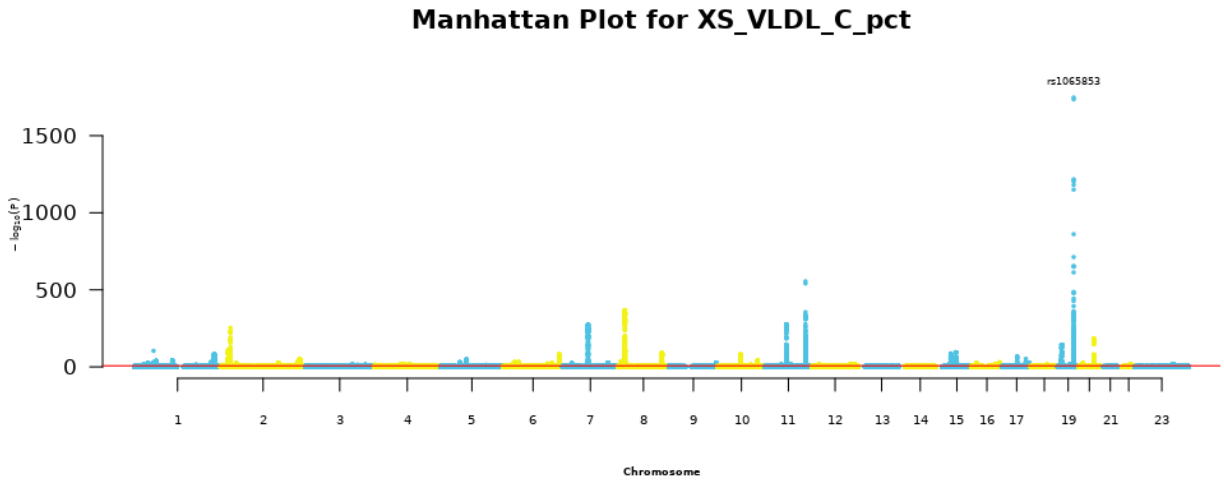
37 S_VLDL_FC_pct



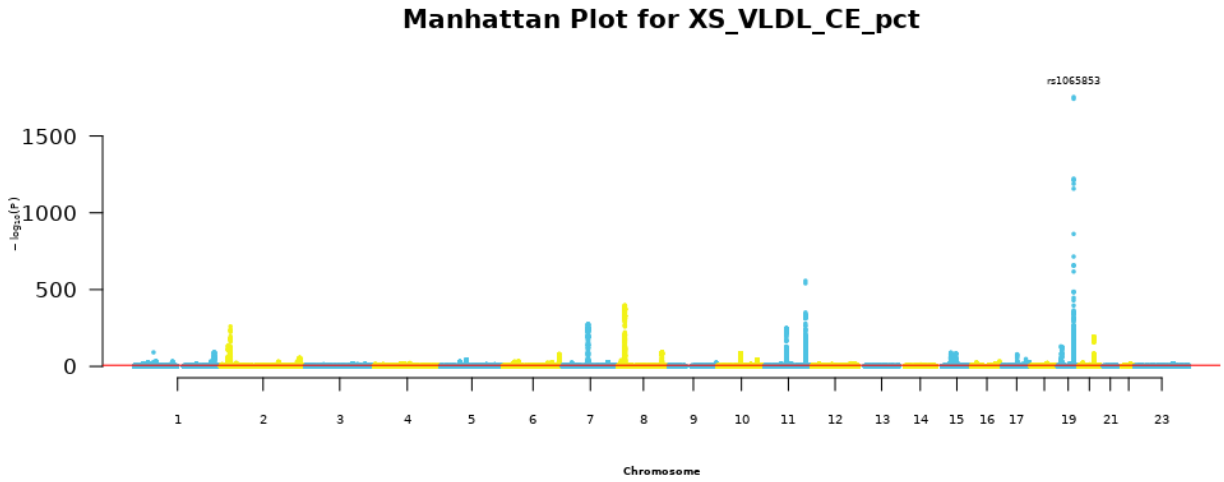
38 S_VLDL_TG_pct



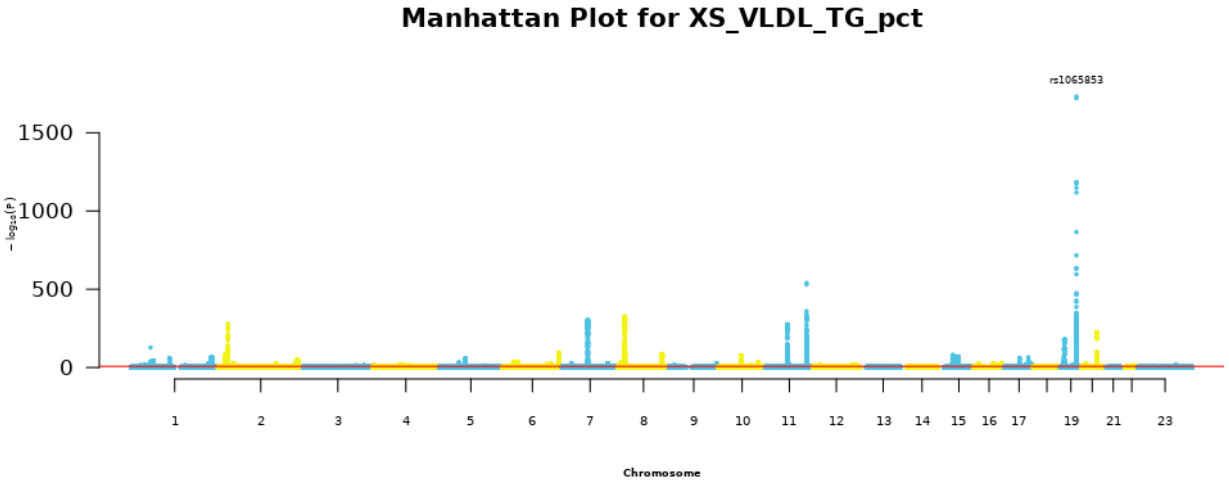
39 XS_VLDL_C_pct



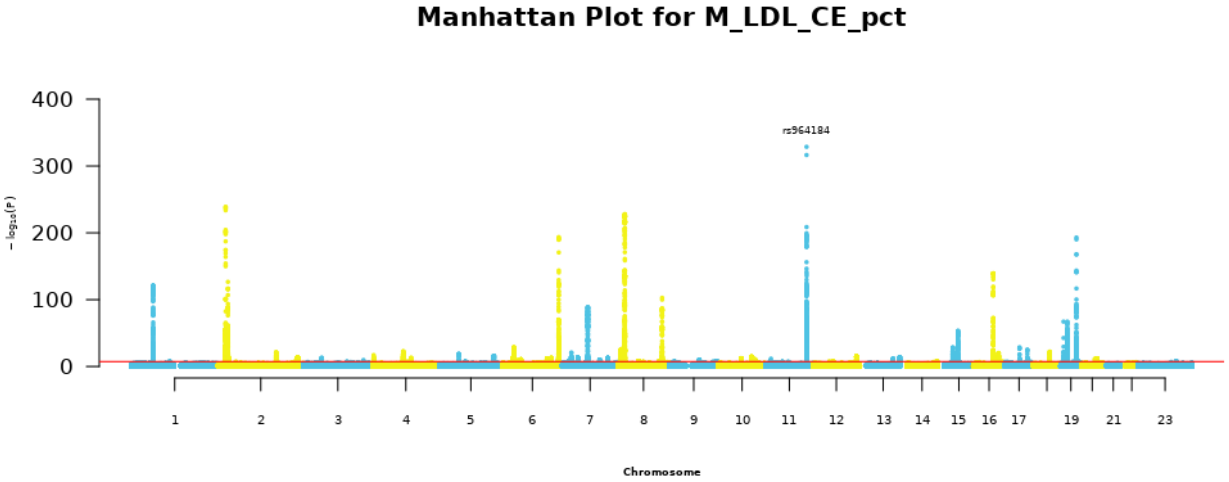
40 XS_VLDL_CE_pct



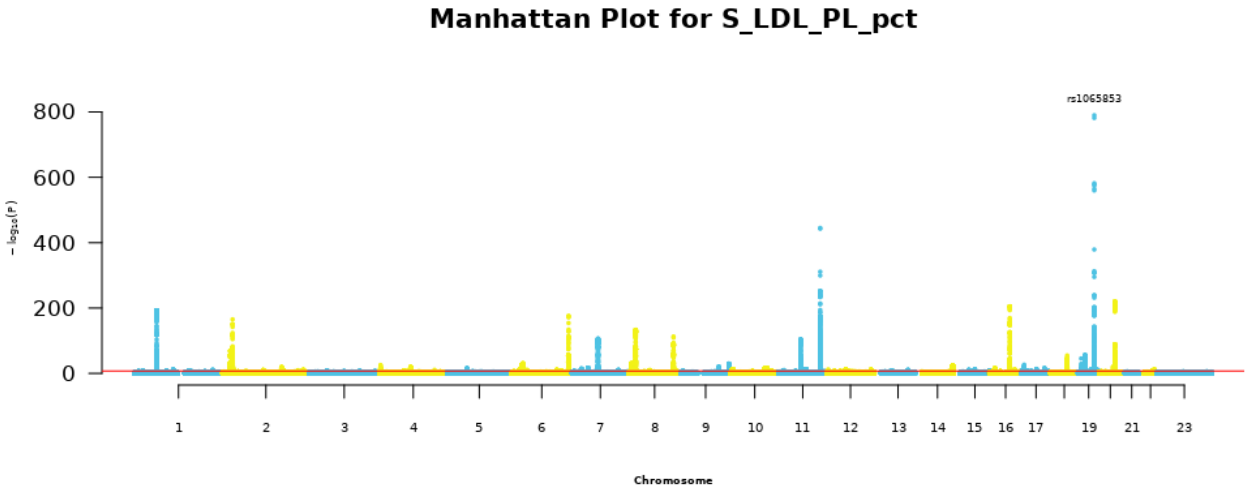
41 XS_VLDL_TG_pct



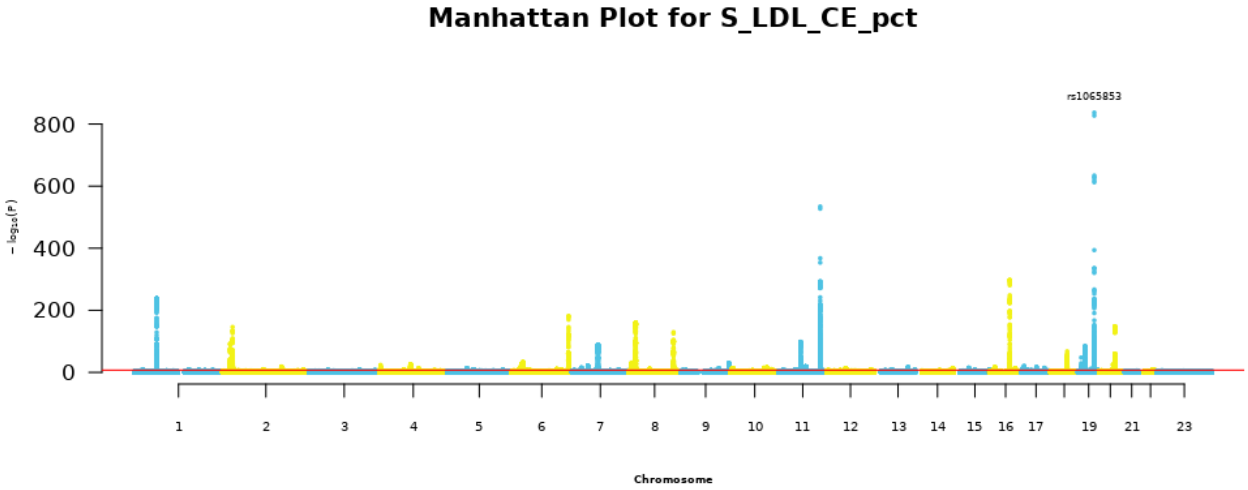
42 M_LDL_CE_pct



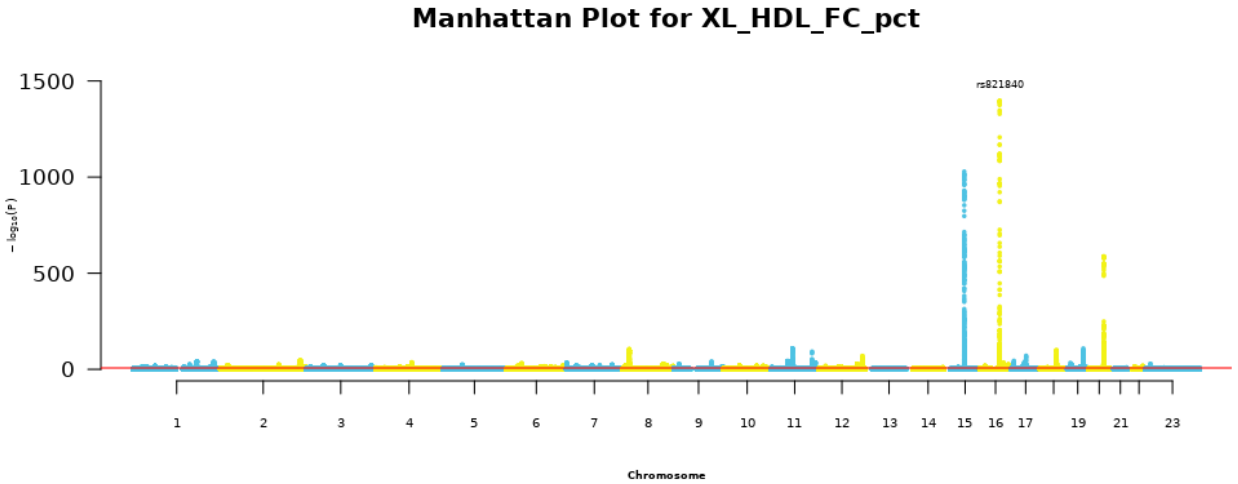
43 S_LDL_PL_pct



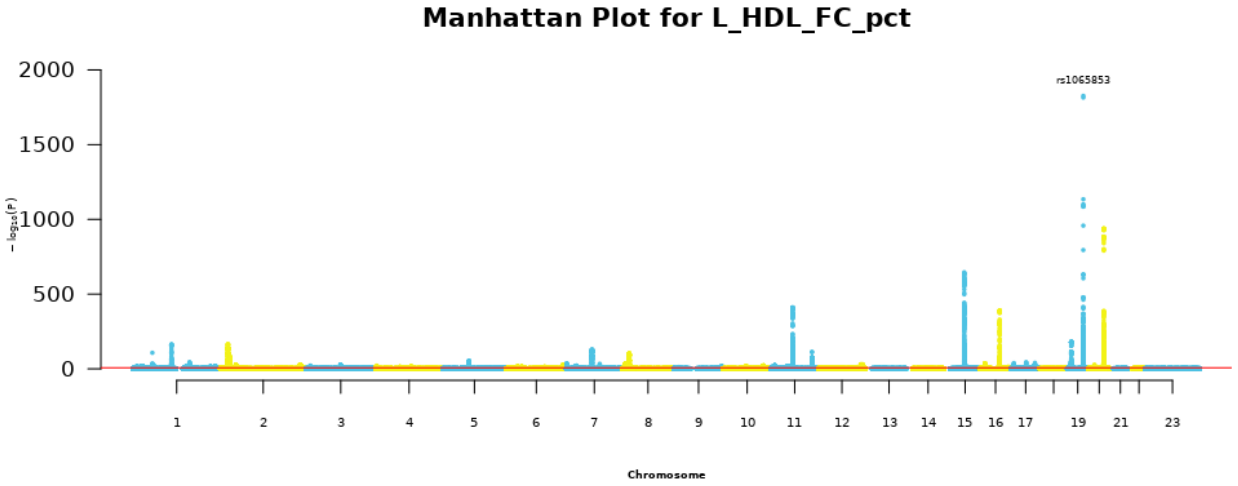
44 S_LDL_CE_pct



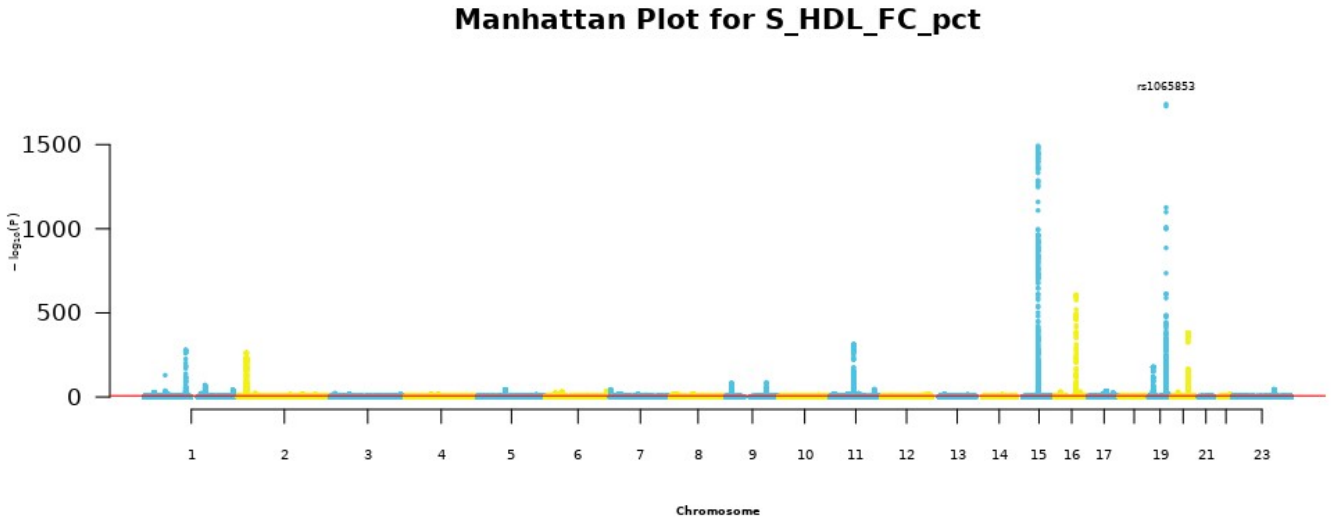
45 XL_HDL_FC_pct



46 L_HDL_FC_pct

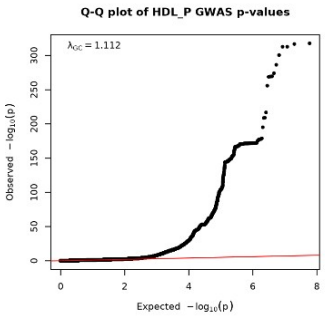
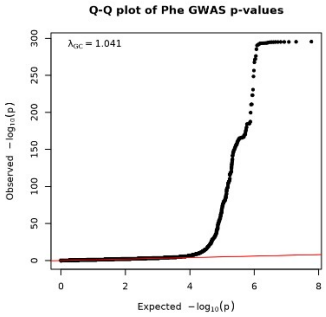
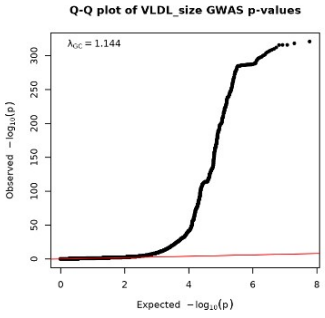
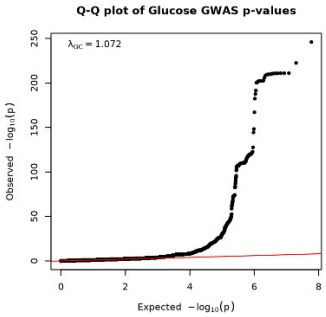
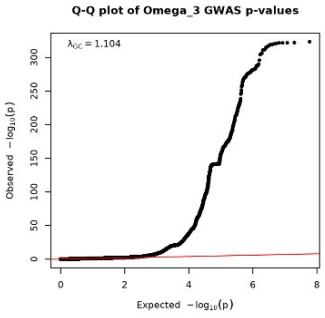
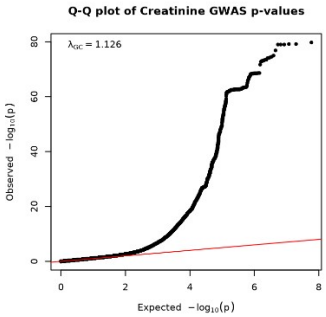
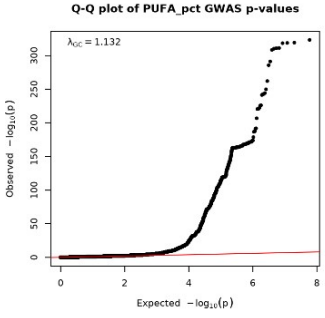
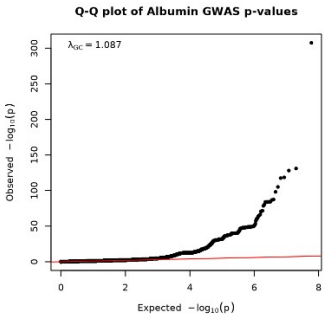
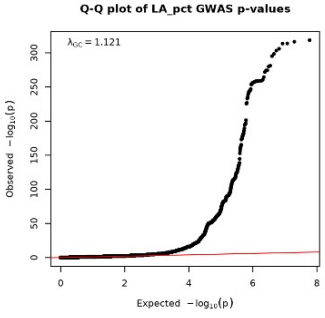
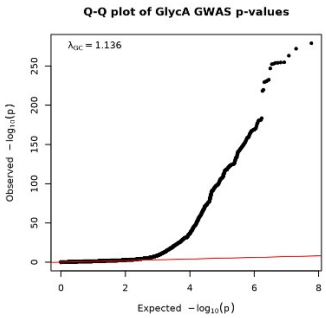


47 S_HDL_FC_pct



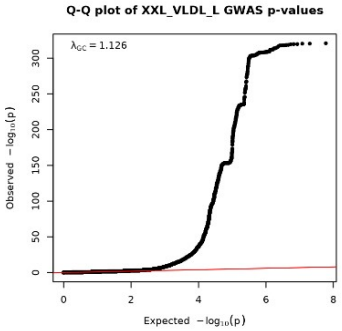
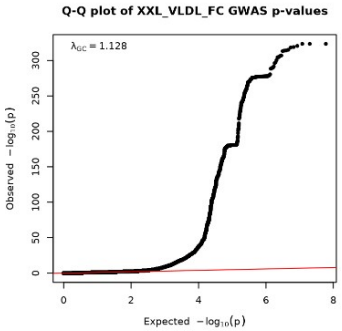
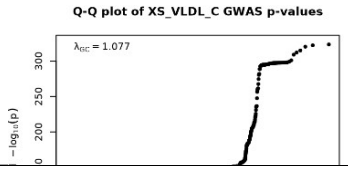
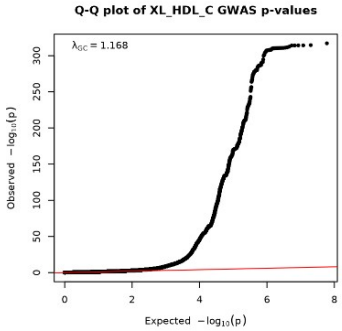
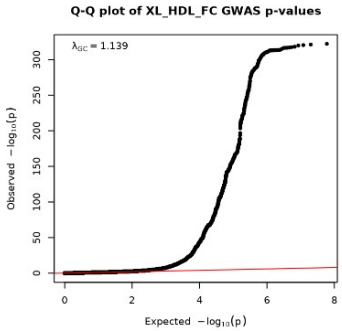
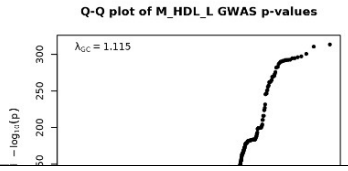
Supplementary Figure 6.2 Manhattan plots of 47 metabolites and single nucleotide polymorphisms (N = 230,197)

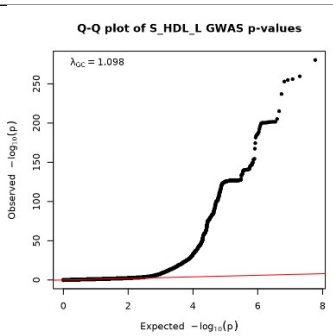
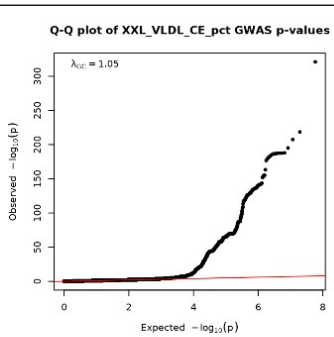
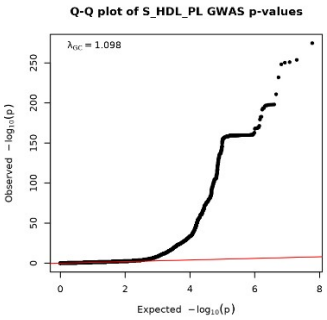
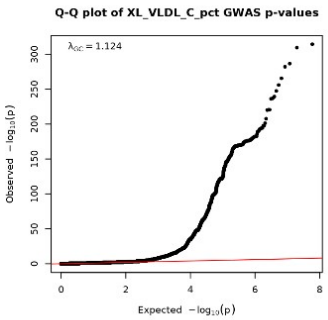
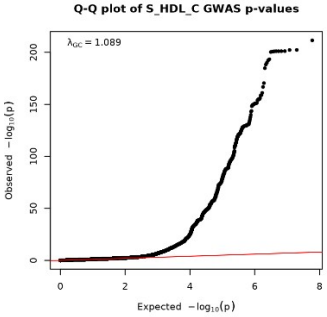
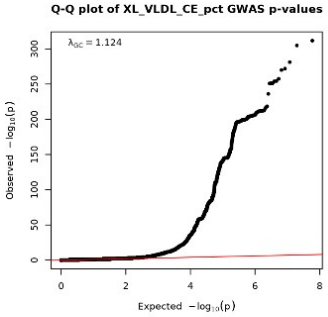
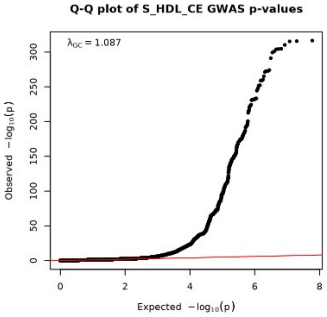
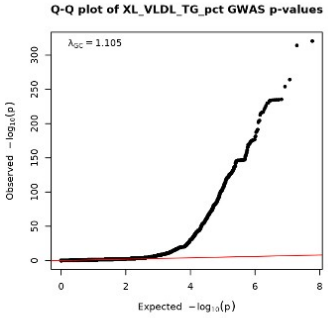
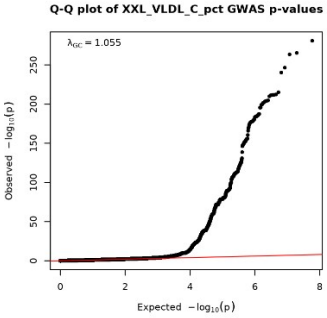
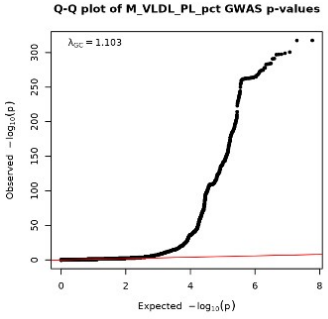
Manhattan plots display chromosomal positions (x-axis) of associations (the negative logarithm of the genome-wide association studies p values, y-axis). P values were obtained from REGENIE regression genome-wide association studies (two-sided), adjusting for age, sex, fasting time, genotype measurement batch, and the first twenty genetic principal components. Red horizontal lines represent genome-wide significance threshold ($p = 5.0 \times 10^{-8}$). Abbreviations are in line with **Supplementary Table 3.1**.

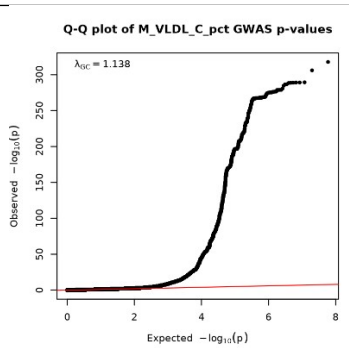
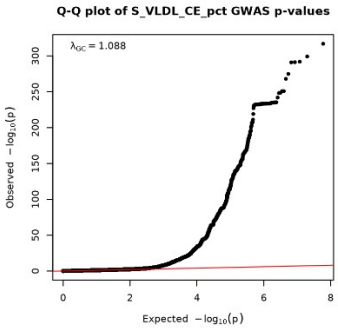
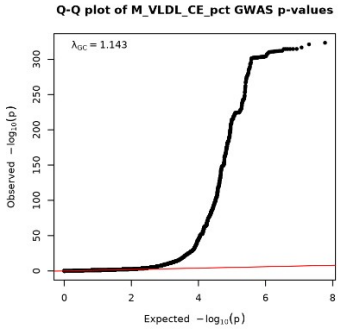
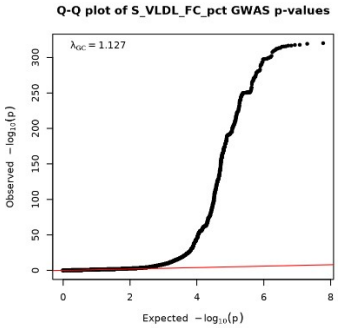
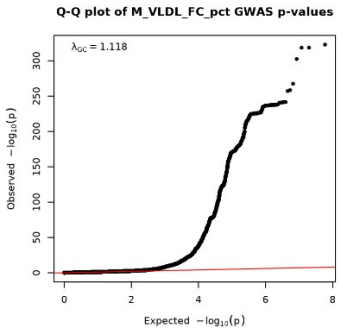
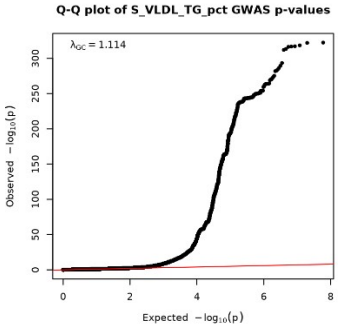
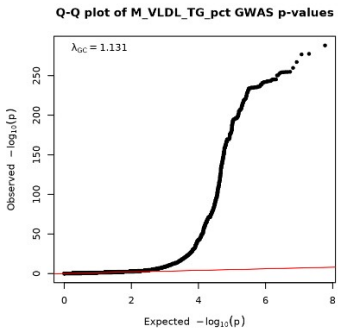
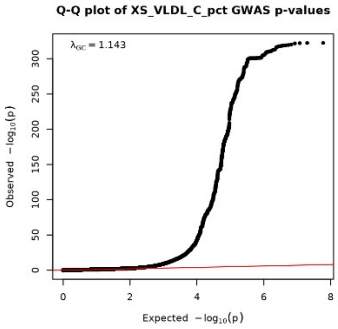
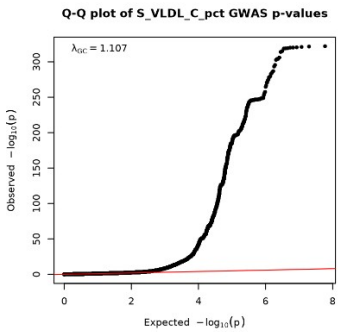
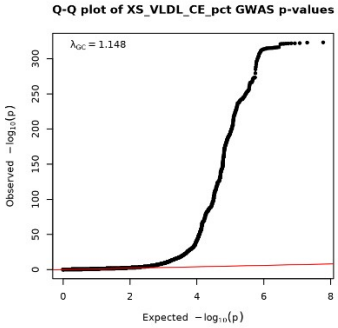
ID 1		ID 6	
ID 2		ID 7	
ID 3		ID 8	
ID 4		ID 9	
ID 5		ID 10	

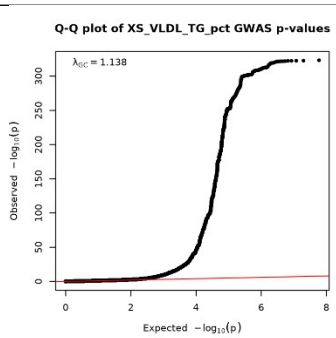
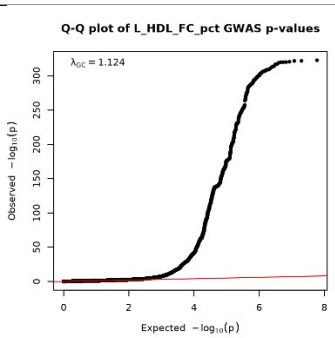
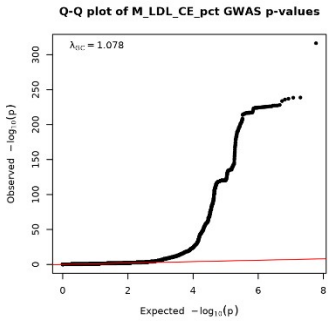
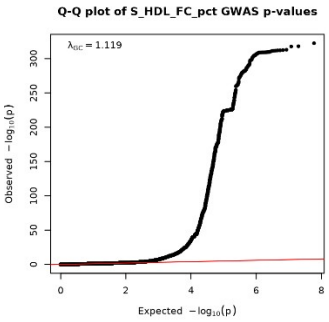
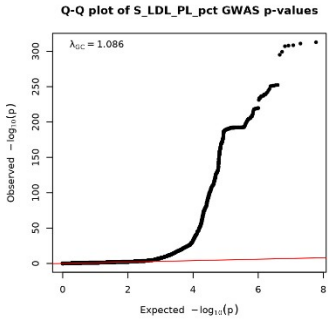
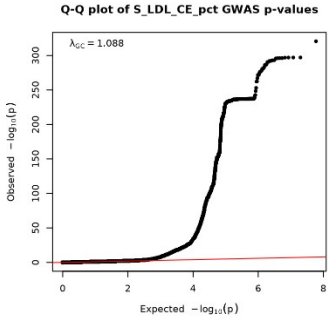
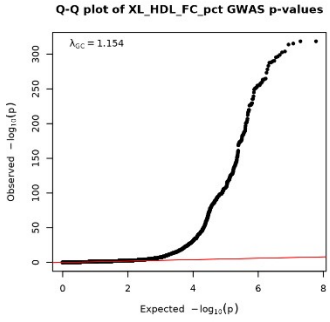
Supplementary Figure 6.3 Quantile-Quantile (QQ) Plot of p values from Genome-Wide Association Study (GWAS) with genomic control lambda (λ_{GC})

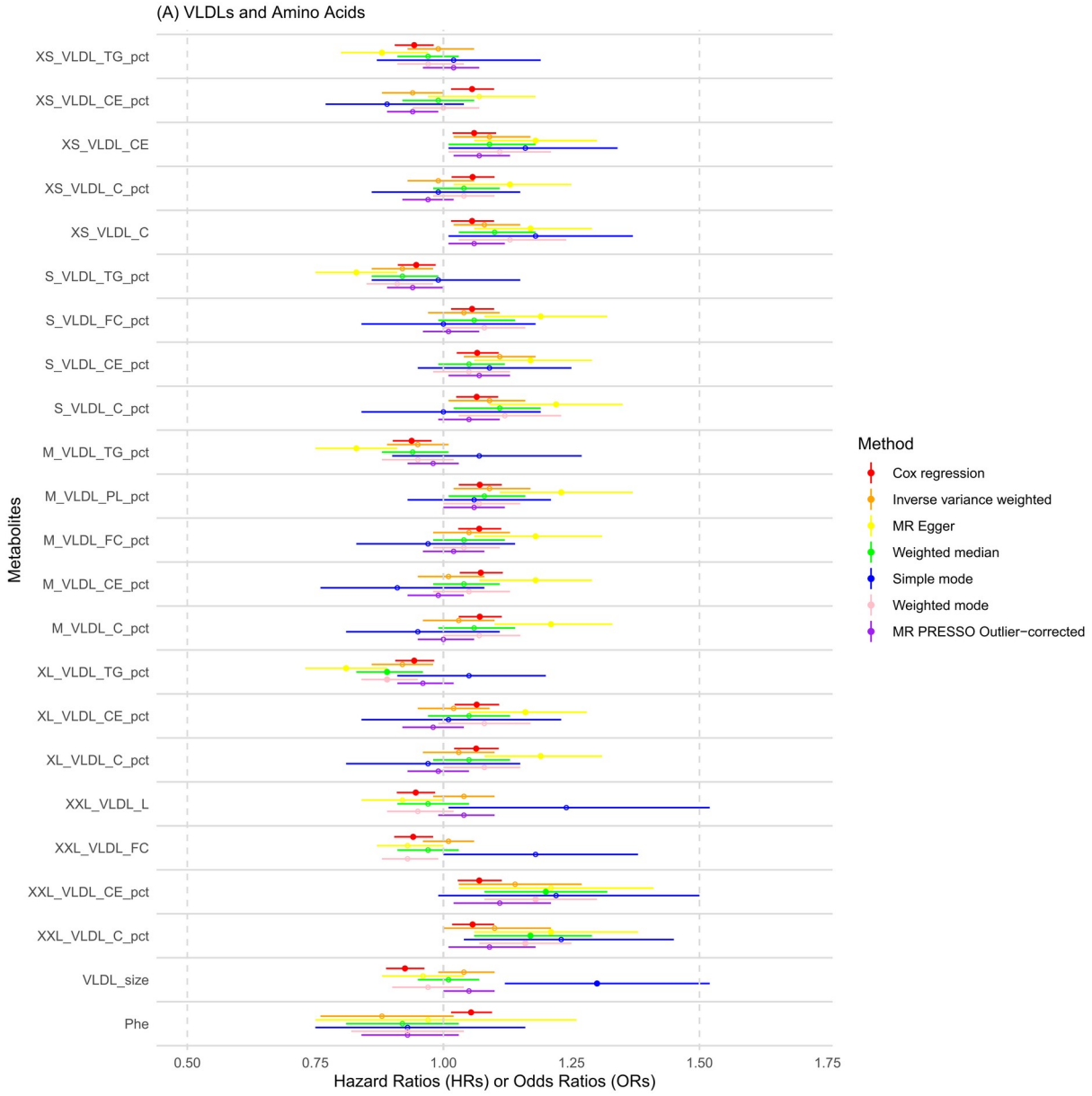
ID 11		ID 16	
ID 12		ID 17	
ID 13		ID 18	
ID 14		ID 19	

ID 15	  	ID 20	  

ID 21		ID 26	
ID 22		ID 27	
ID 23		ID 28	
ID 24		ID 29	
ID 25		ID 30	

ID 31		ID 36	
ID 32		ID 37	
ID 33		ID 38	
ID 34		ID 39	
ID 35		ID 40	

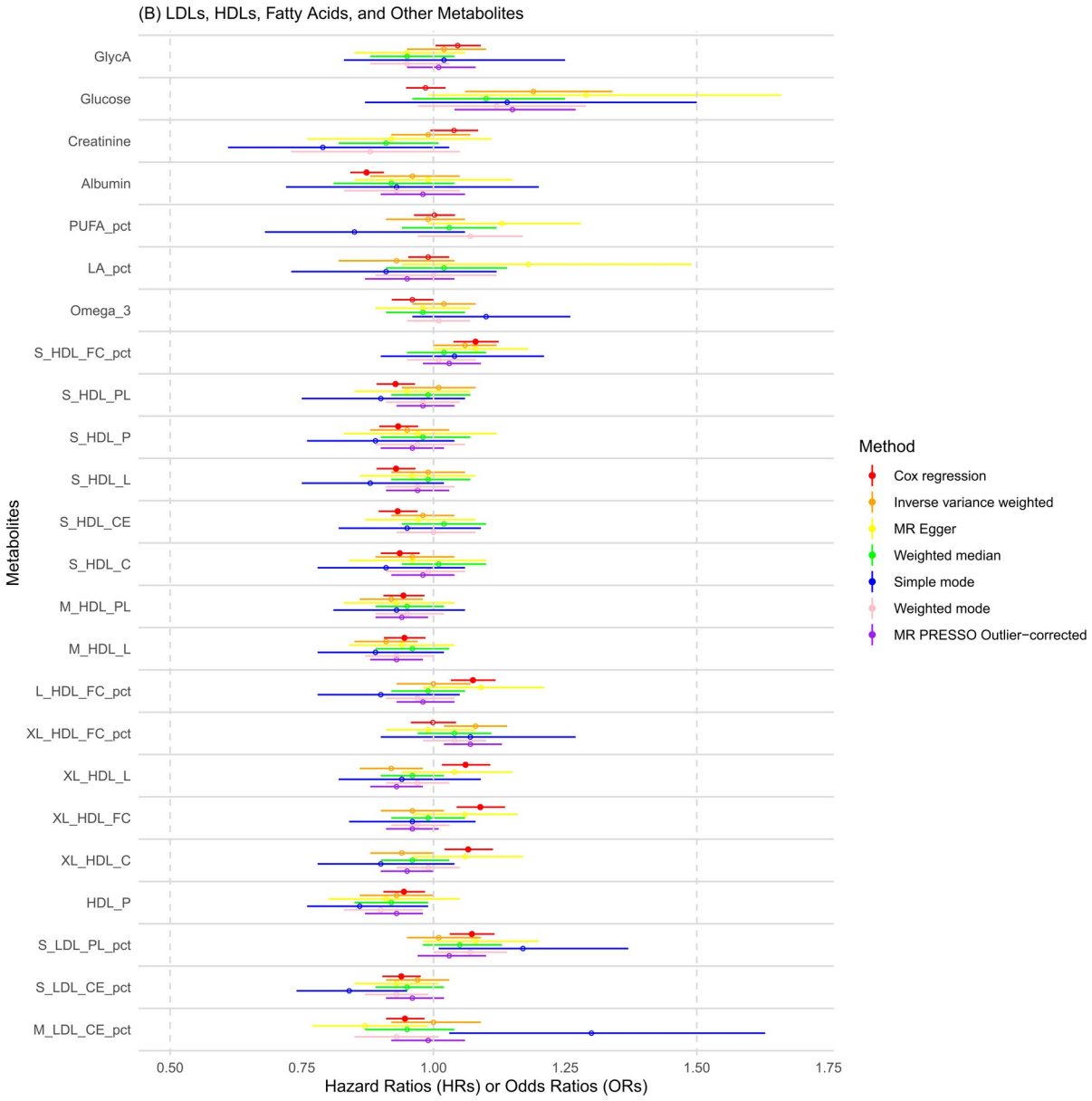
ID 41		ID 46	
ID 42		ID 47	
ID 43			
ID 44			
ID 45			



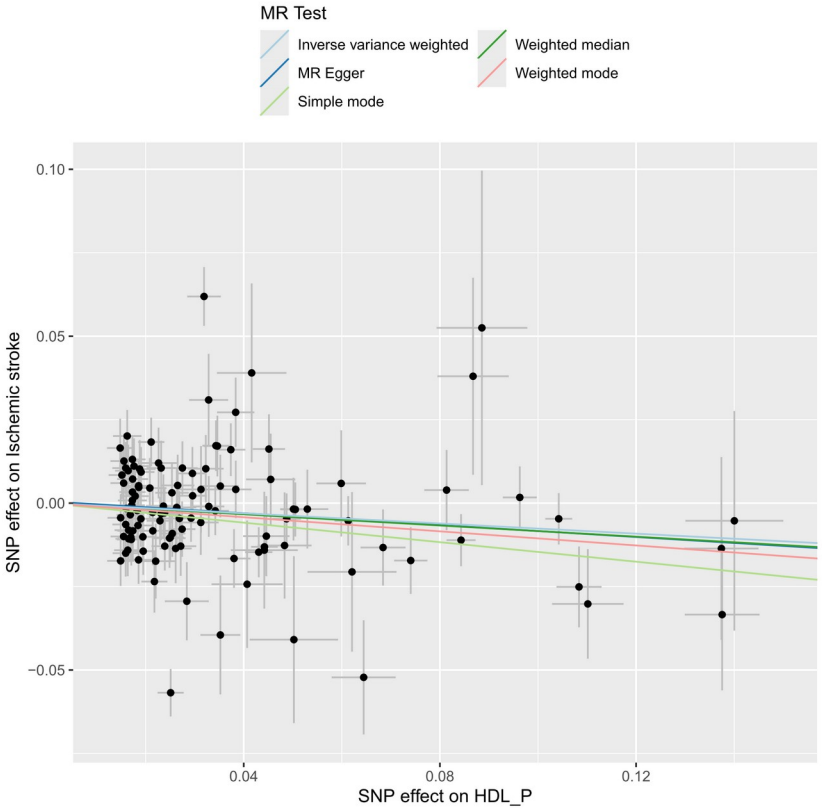
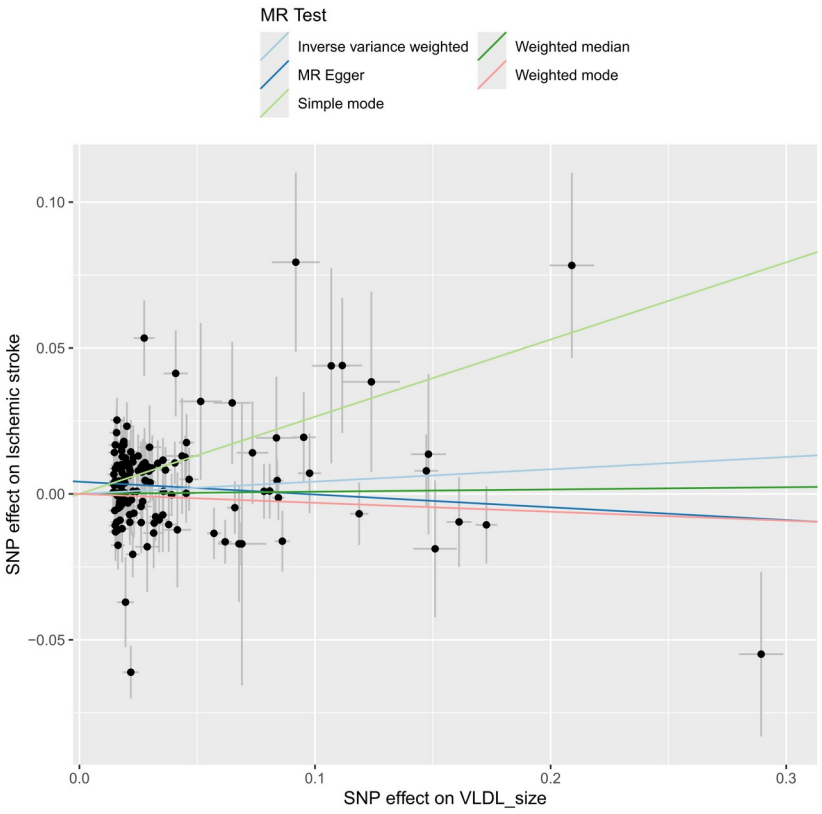
Supplementary Figure 6.4 Comparison of association of ischaemic stroke with metabolites from observational study and Mendelian randomisation

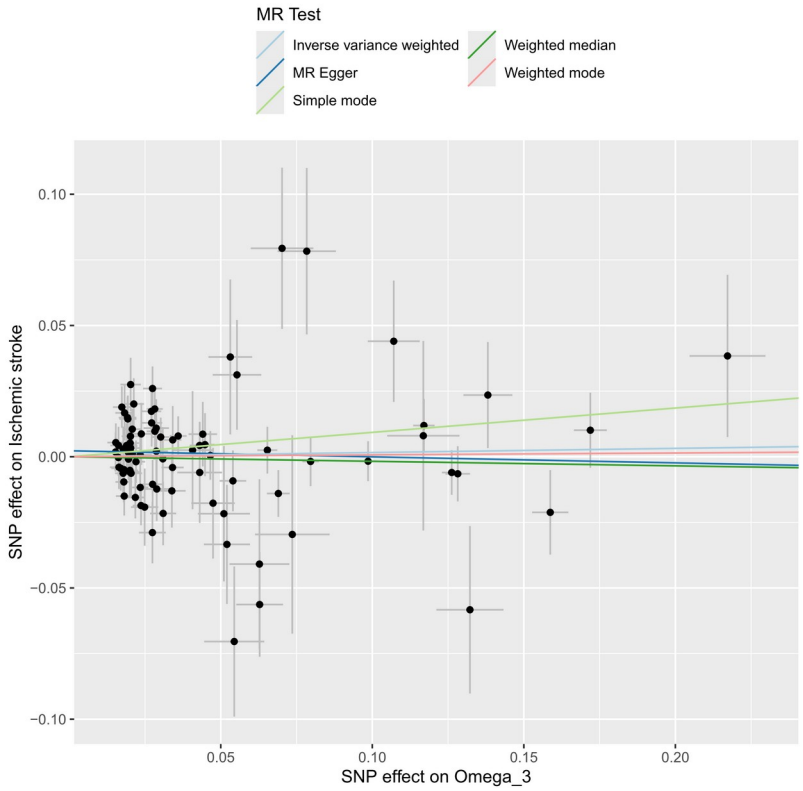
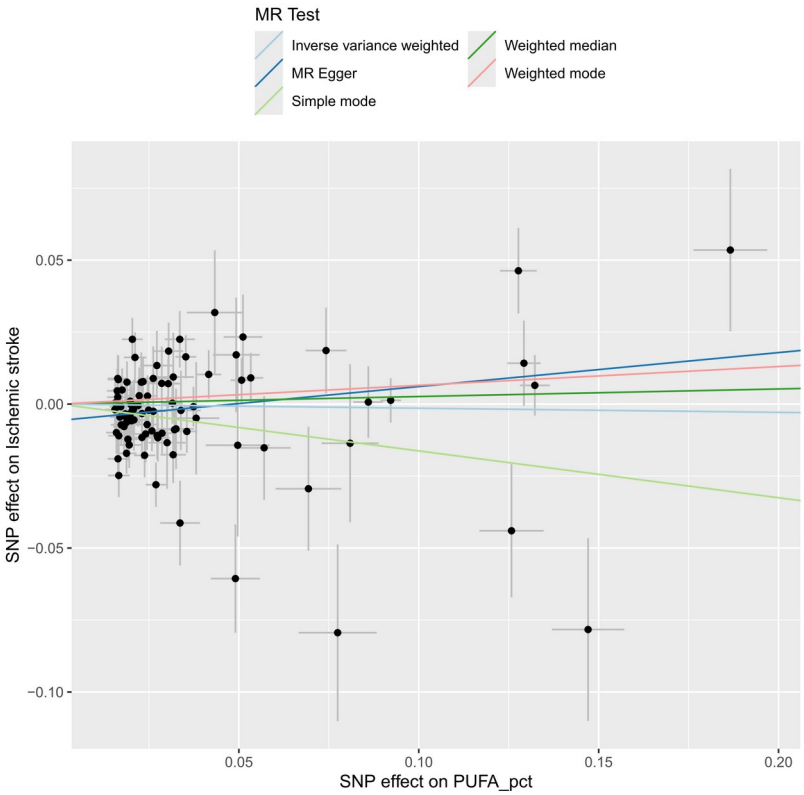
Cox regression model was adjusted for age, sex, ethnicity, lipid modifying medication, fasting time, spectrometers, assessment centres, smoking status, alcohol intake frequency, dietary supplement use and 13 dietary variables (fresh fruit, whole and refined grain food, caffeinated and decaffeinated coffee, tea, vegetables, cheese, unprocessed red meat, processed meat, oily and non-oily fish, and type of spread), physical activity, body mass

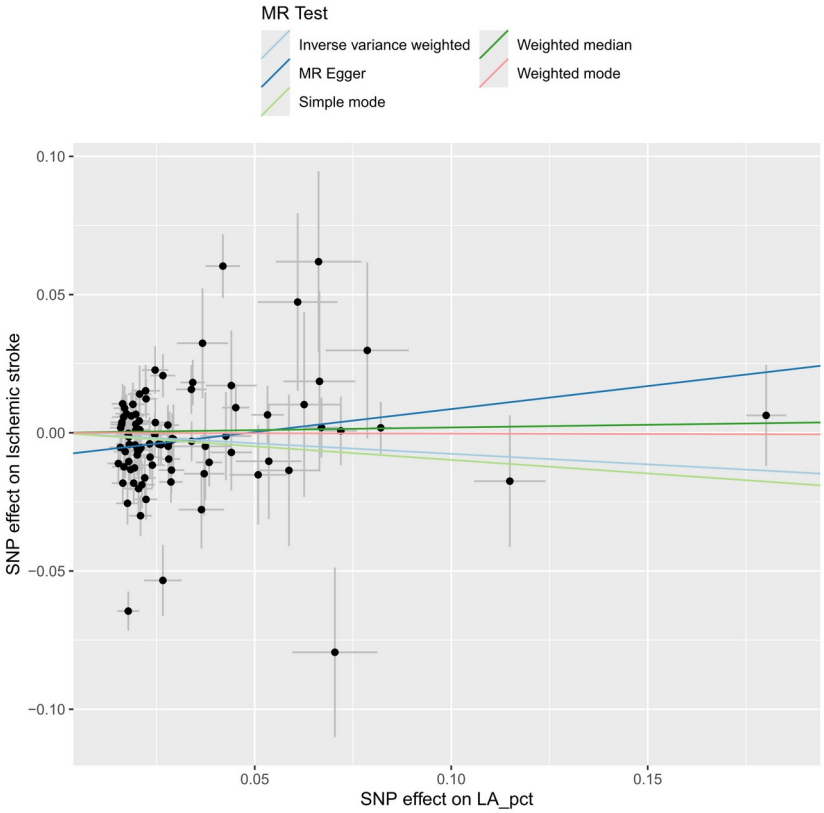
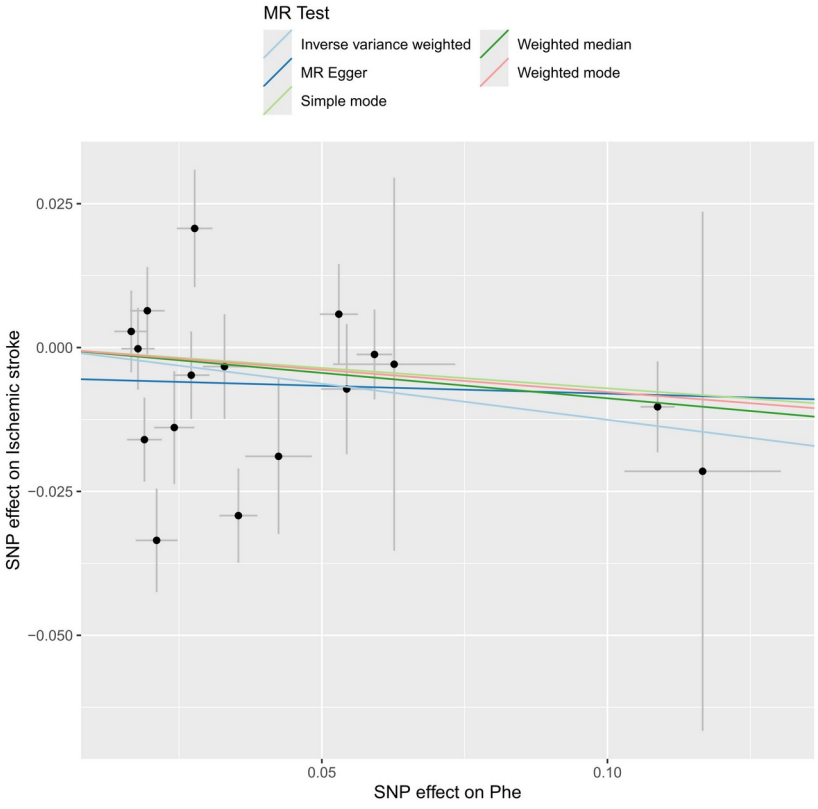
index, prevalent diabetes, and systolic blood pressure. Circle dots and lines denote the hazard ratios or odds ratios with 95% confidence intervals per one standard deviation higher metabolic levels between incident ischaemic stroke and individual metabolites. Filled circle dots represent significance after multiple testing correction. (A) VLDLs and amino acids, (B) LDLs, HDLs, fatty acids, and other metabolites. MR-PRESSO, MR pleiotropy residual sum and outlier. Abbreviations are in line with **Supplementary Table 3.1**.

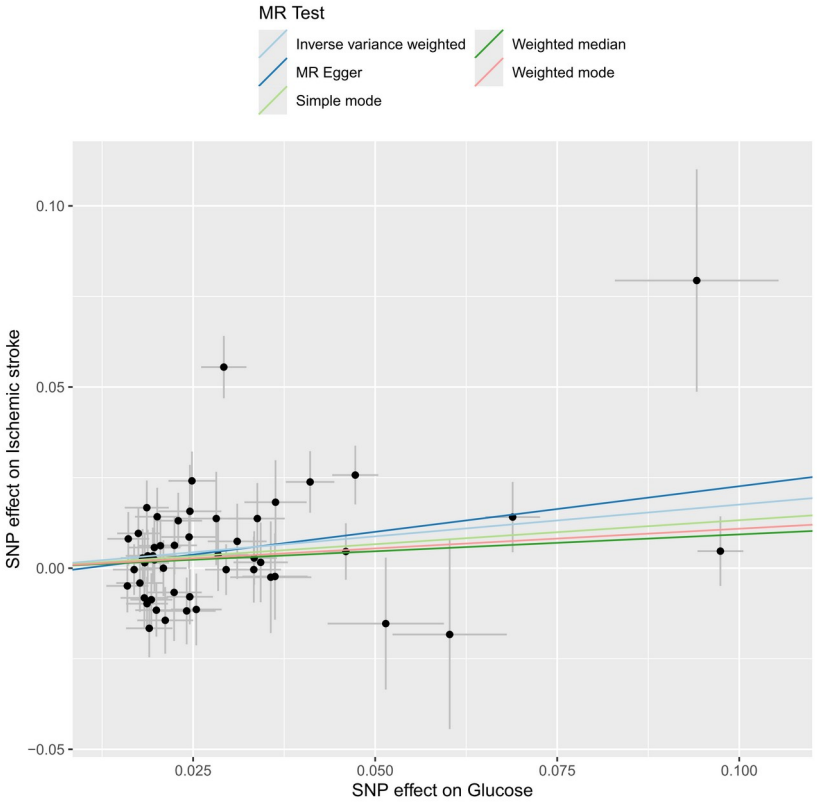
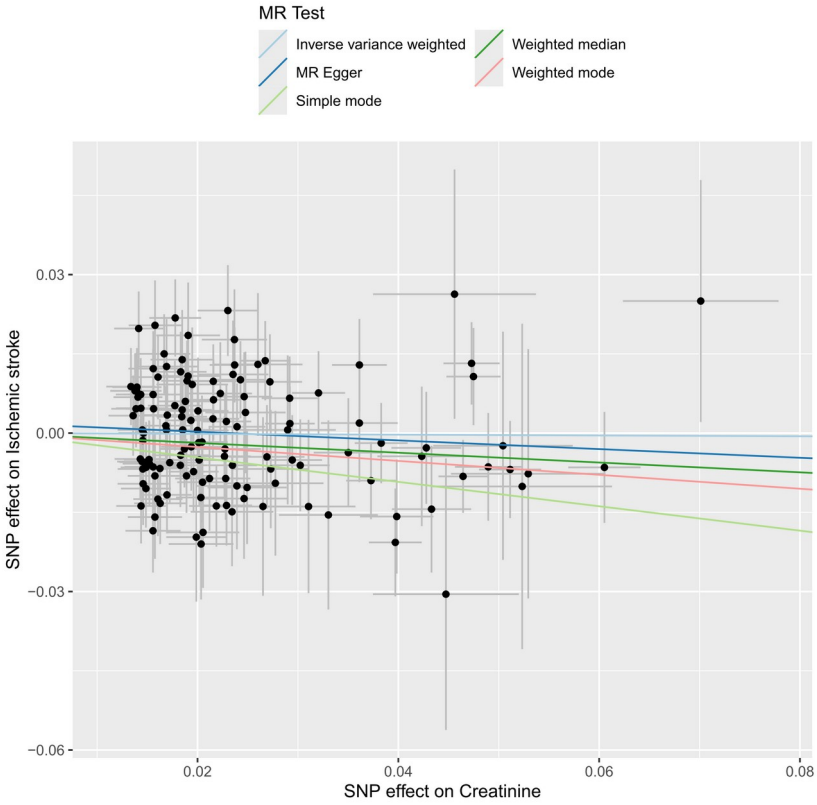


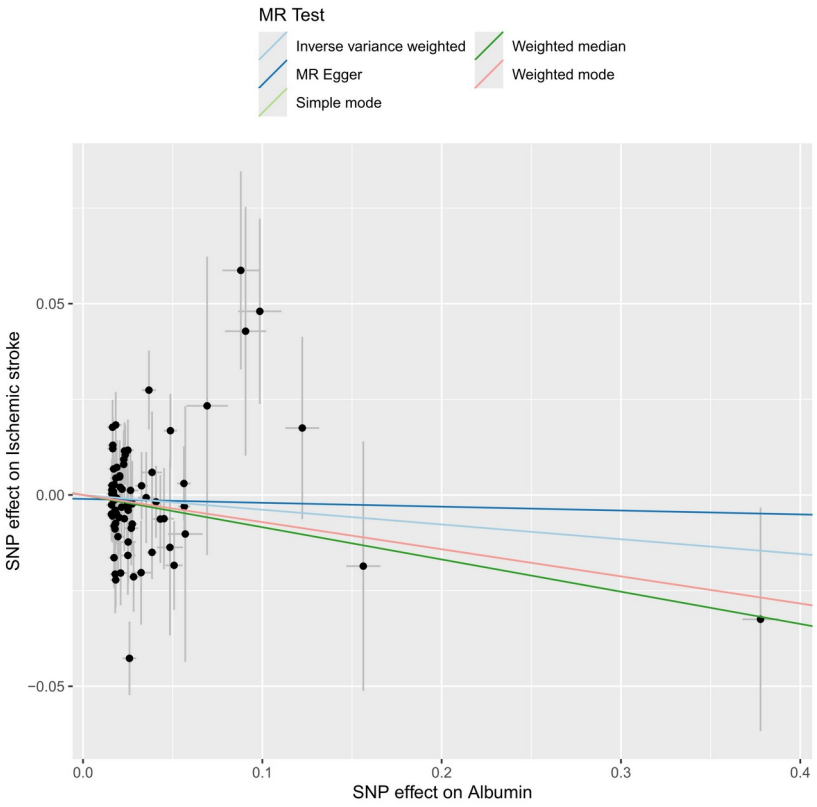
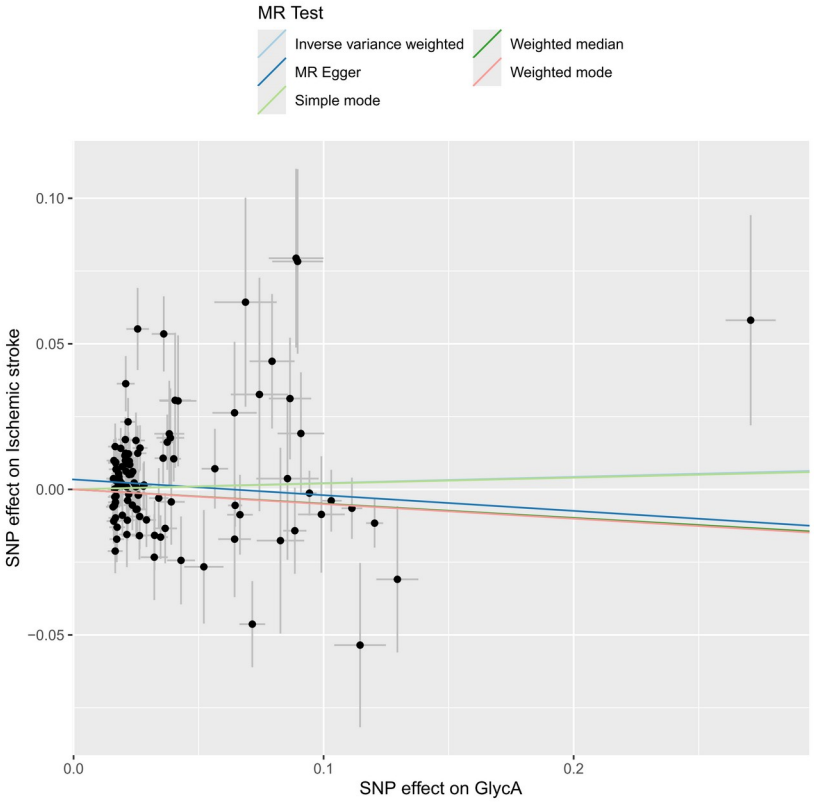
Supplementary Figure 6.2 (Continued) Comparison of association of ischaemic stroke with metabolites from observational study and Mendelian randomisation

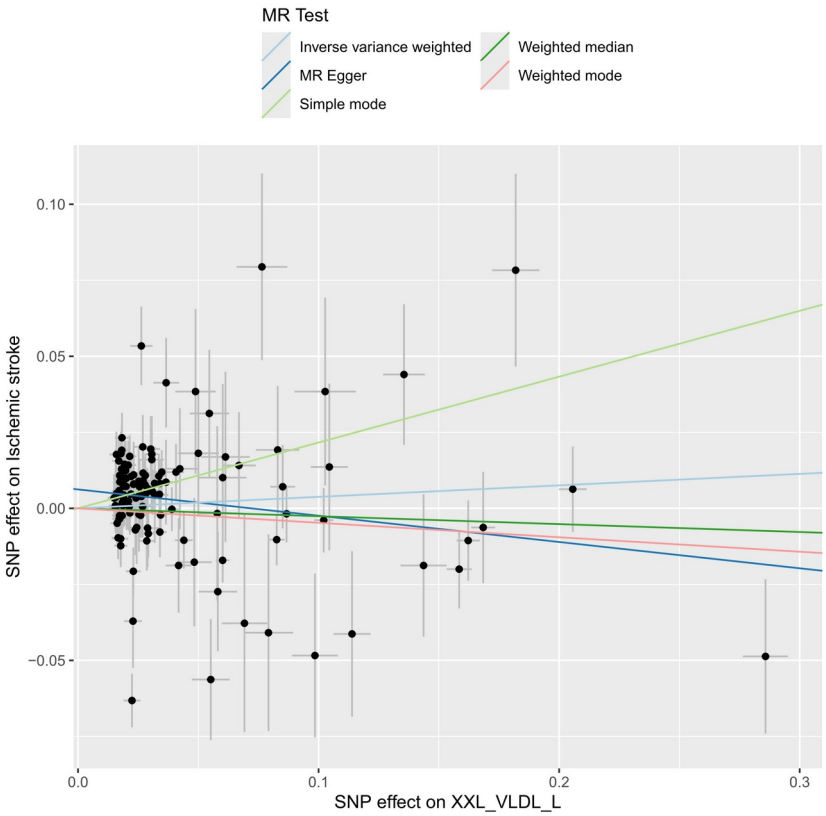
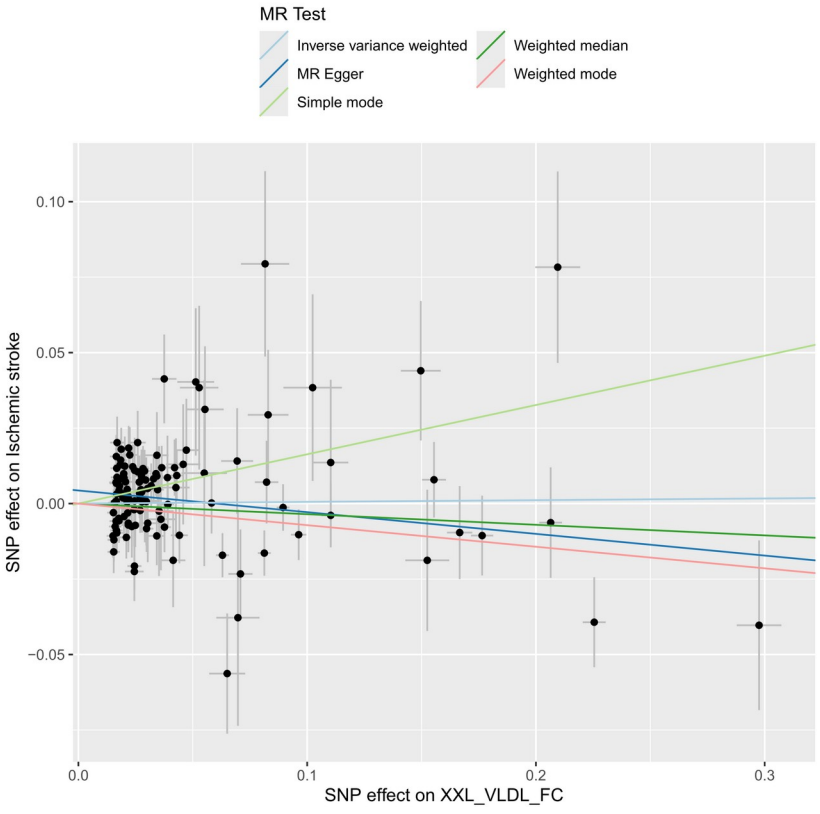
ID	Metabolite	Scatter plot
1	HDL_P	
2	VLDL_size	

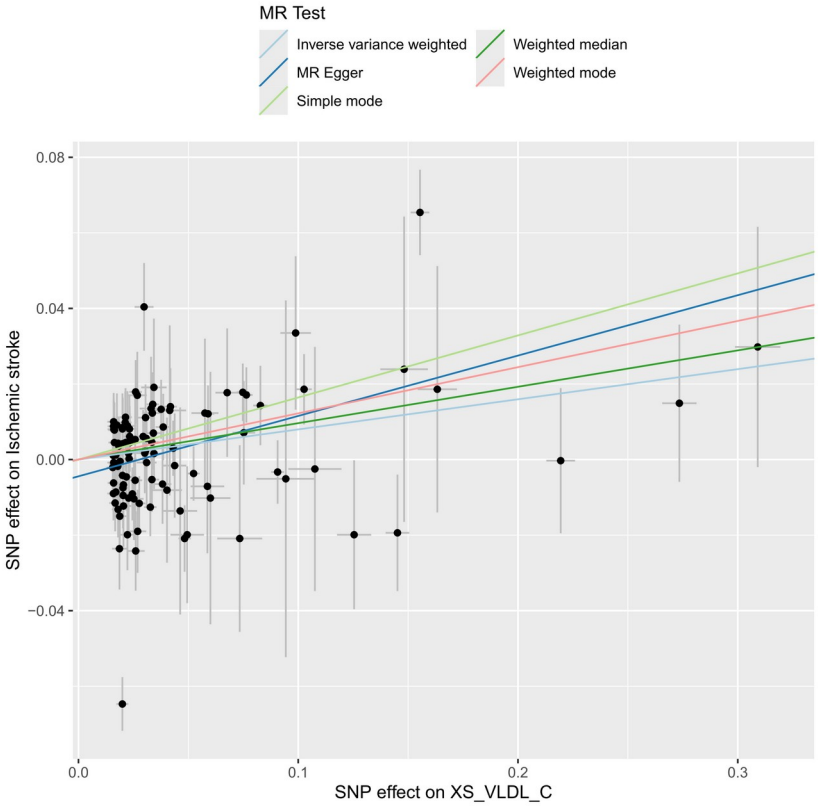
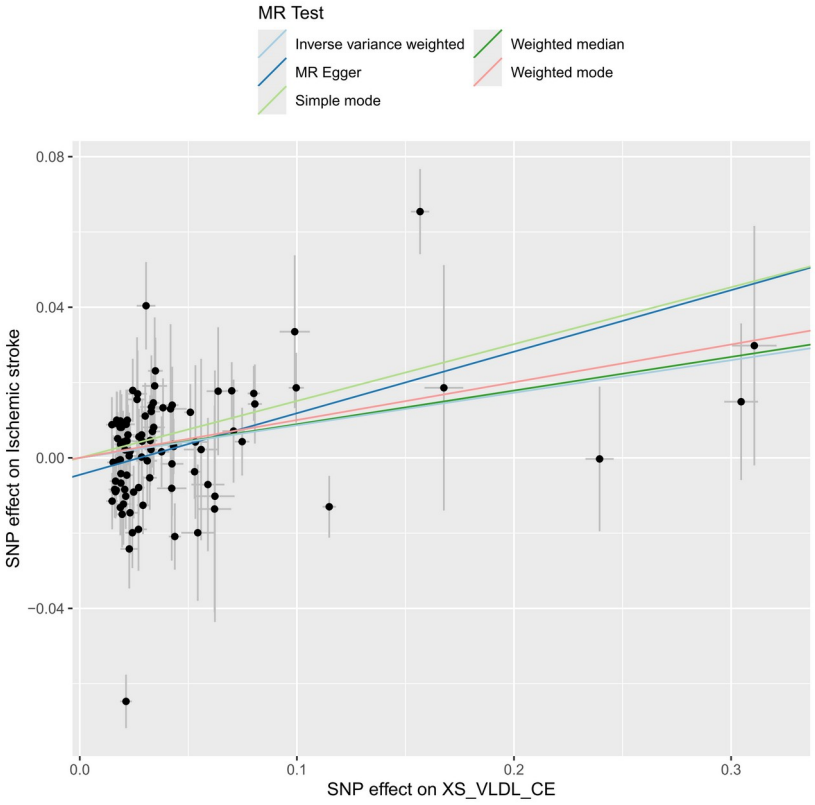
ID	Metabolite	Scatter plot
3	Omega_3	
4	PUFA_pct	

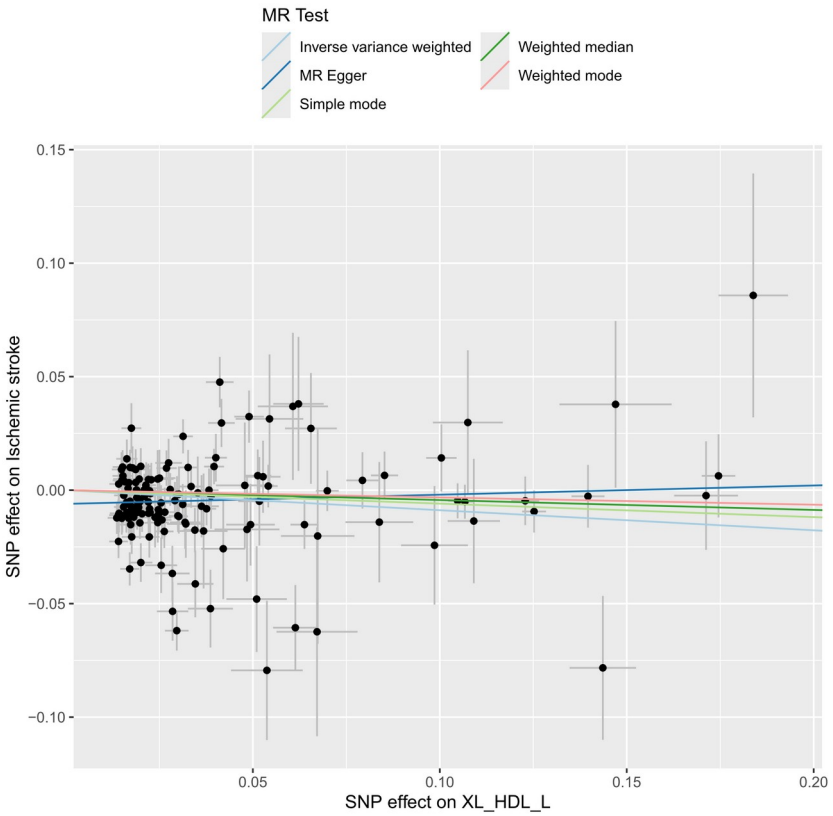
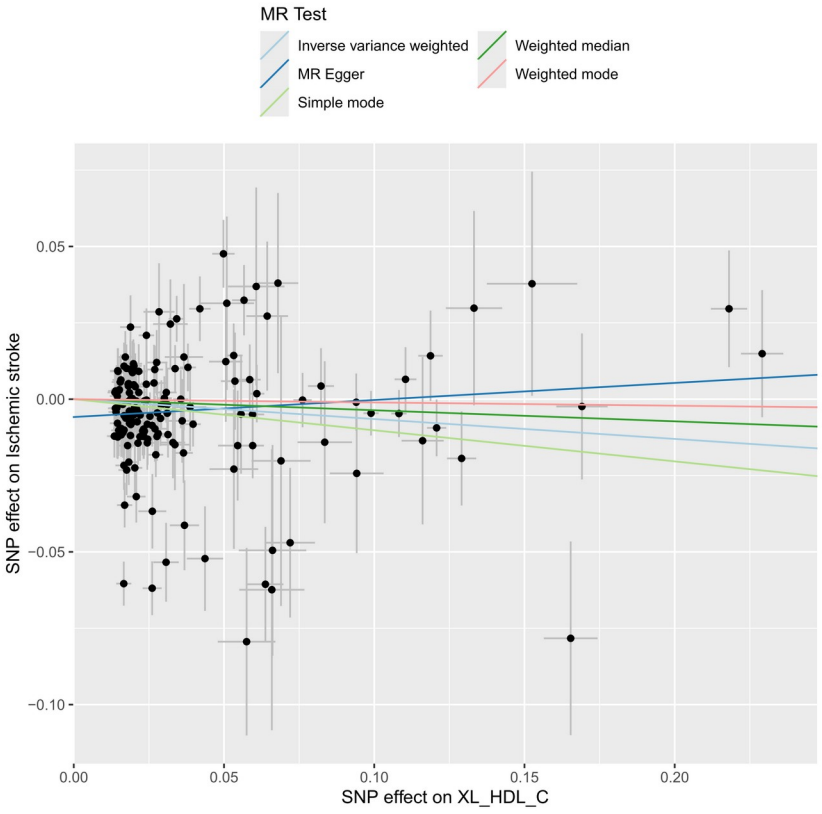
ID	Metabolite	Scatter plot
5	LA_pct	
6	Phe	

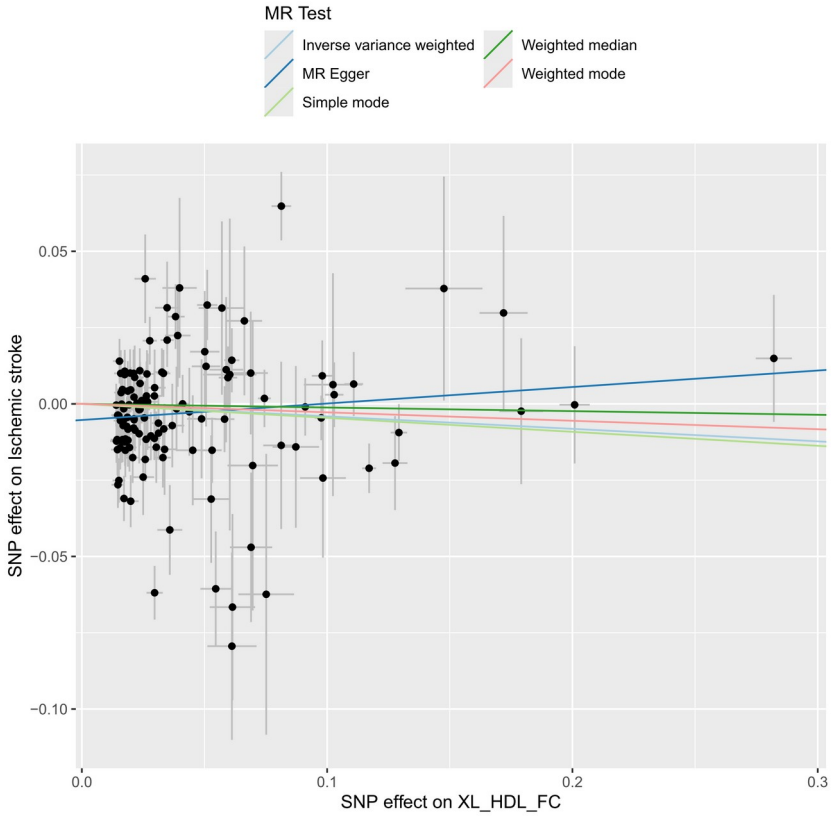
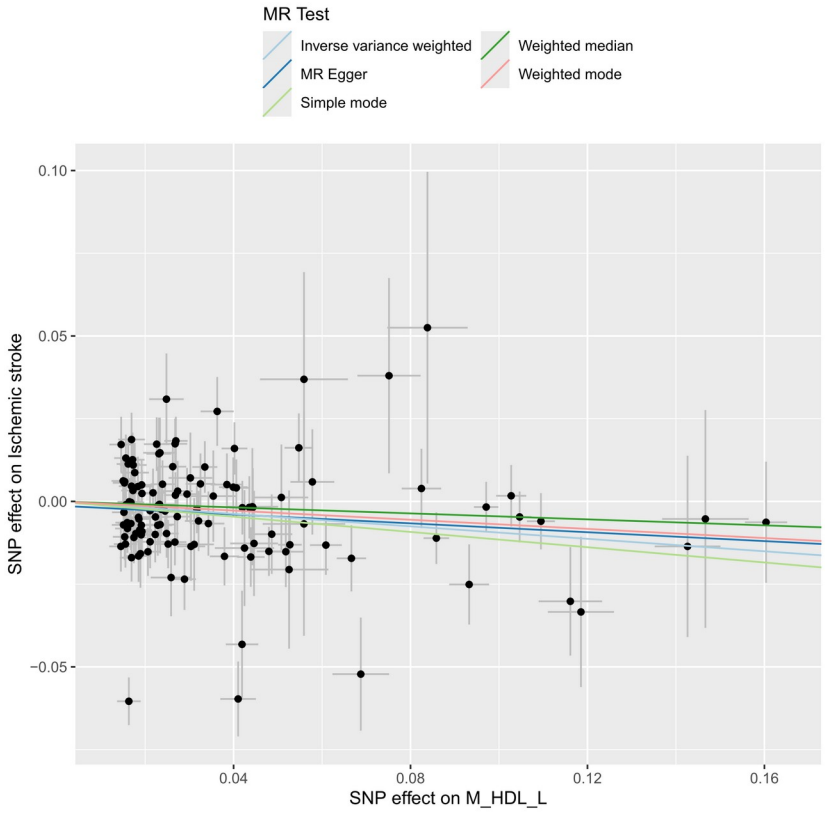
ID	Metabolite	Scatter plot
7	Glucose	
8	Creatinine	

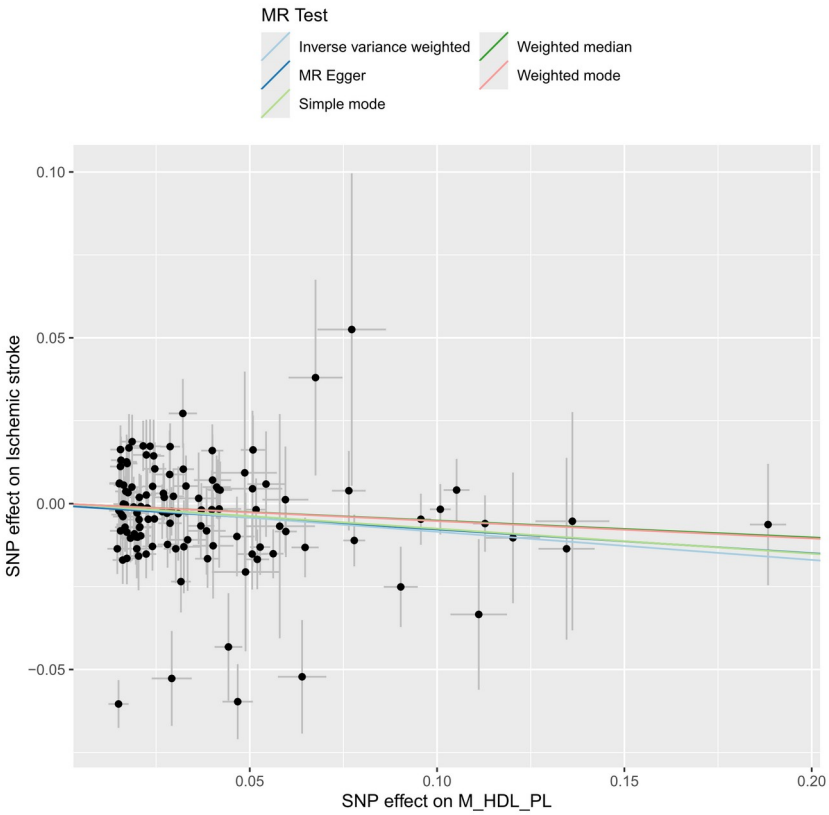
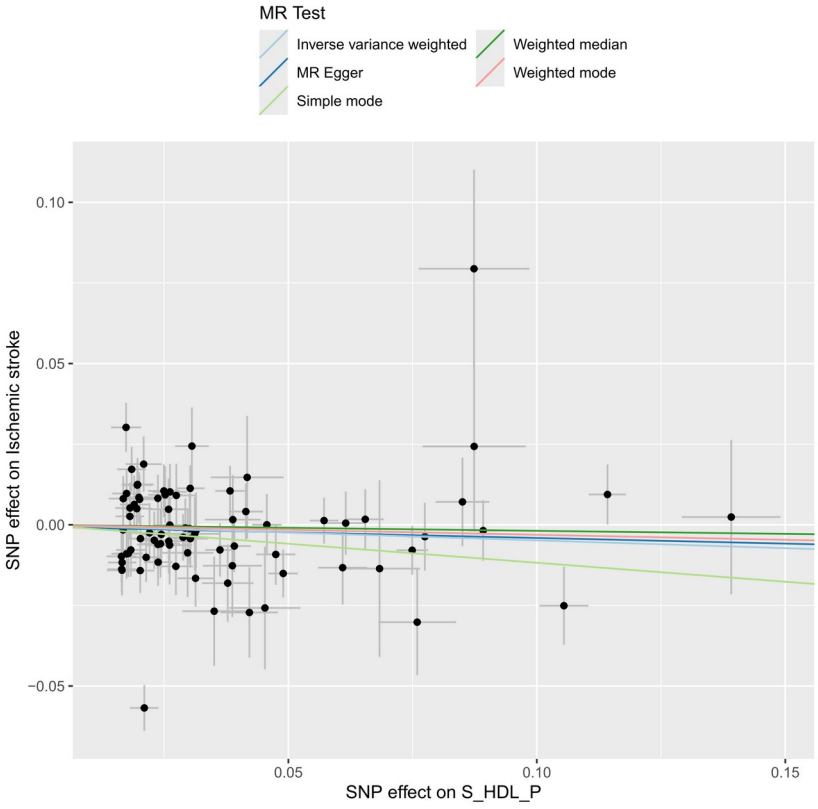
ID	Metabolite	Scatter plot
9	Albumin	
10	GlycA	

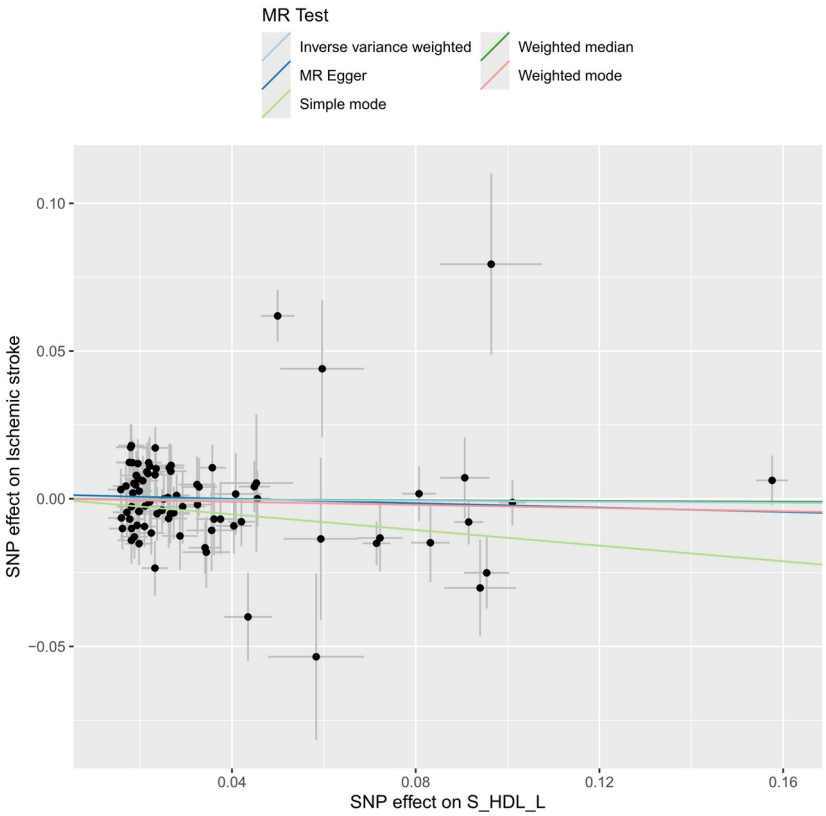
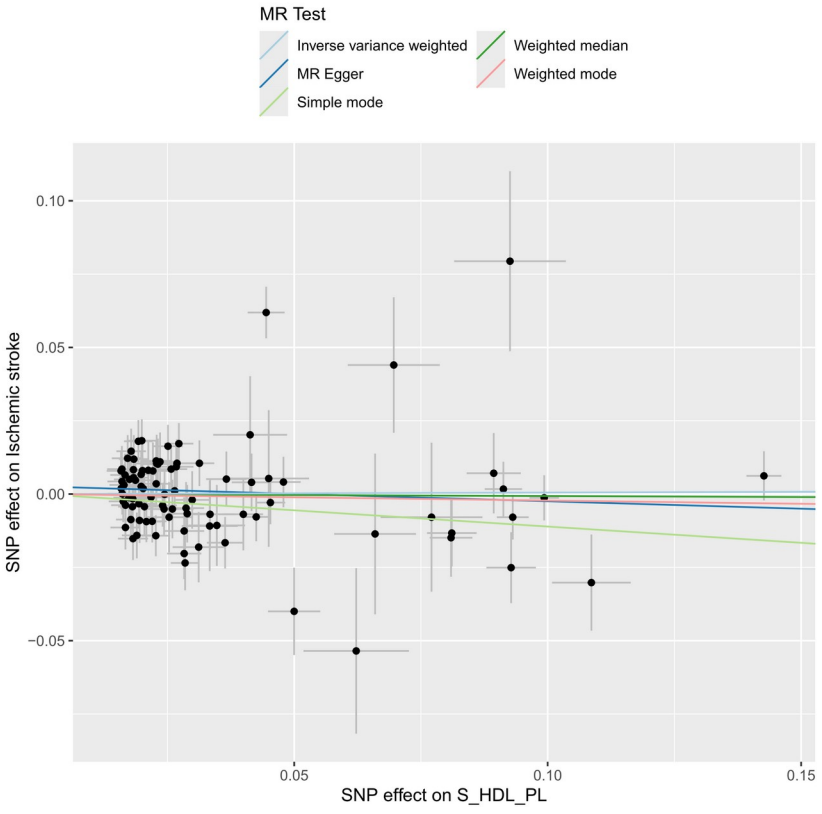
ID	Metabolite	Scatter plot
11	XXL_VLDL_L	
12	XXL_VLDL_FC	

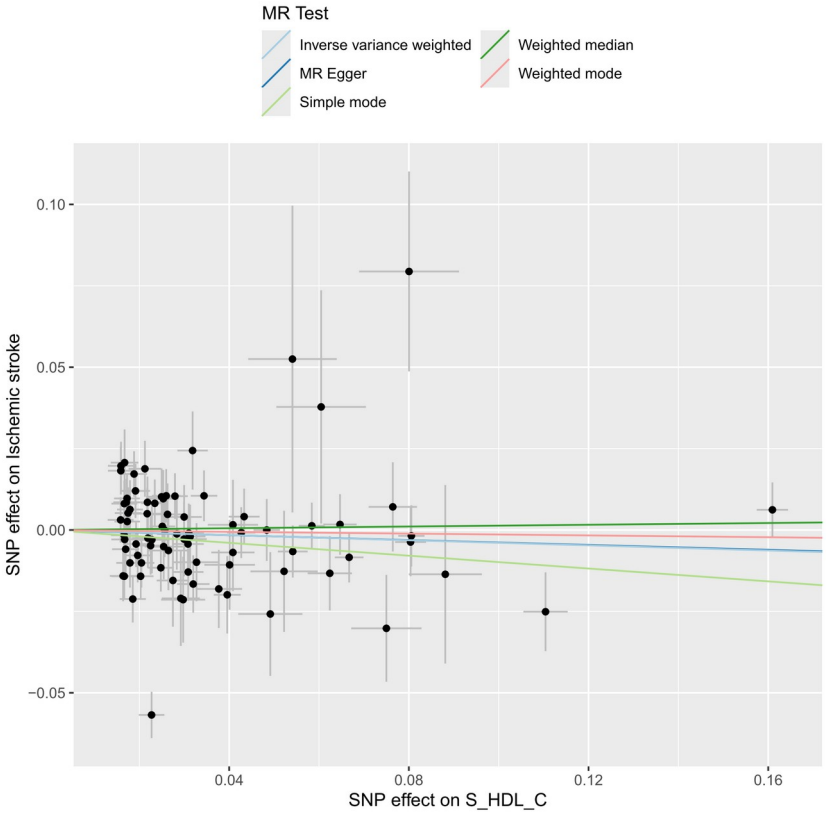
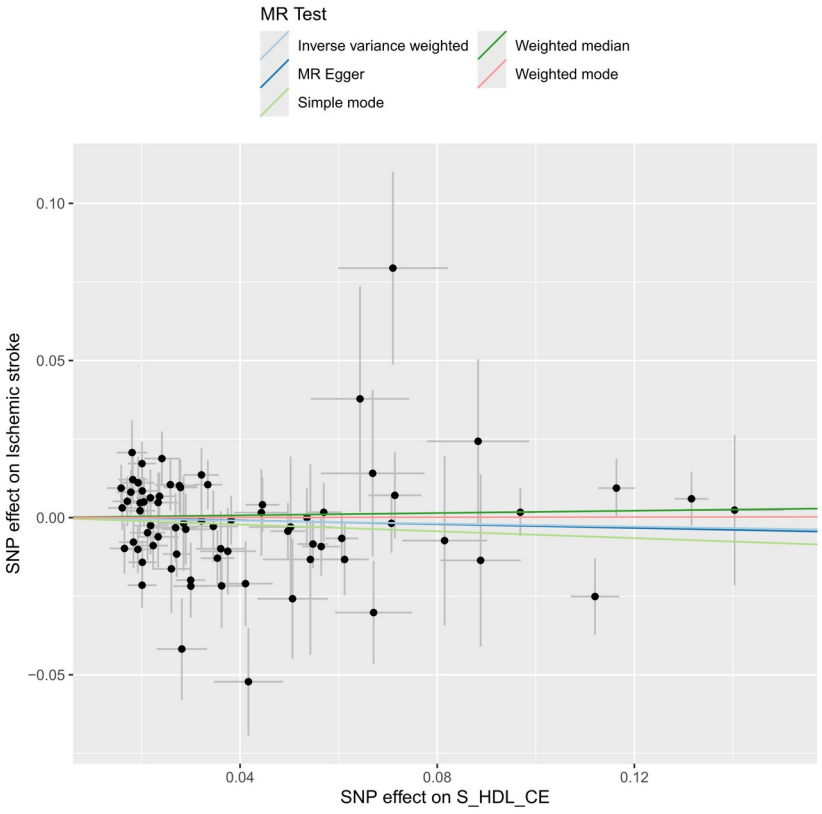
ID	Metabolite	Scatter plot
13	XS_VLDL_C	
14	XS_VLDL_CE	

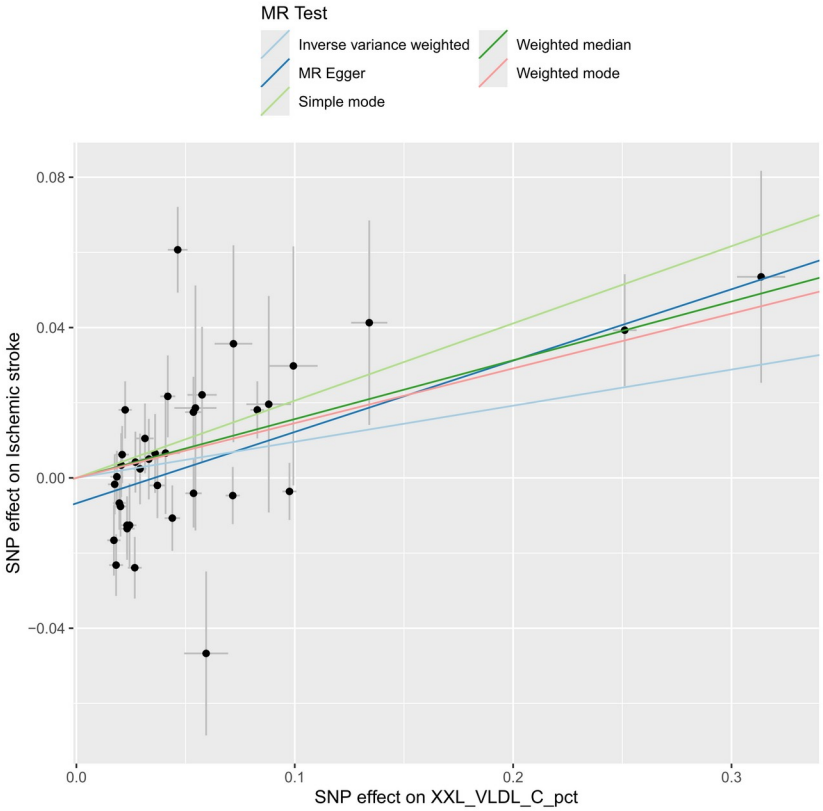
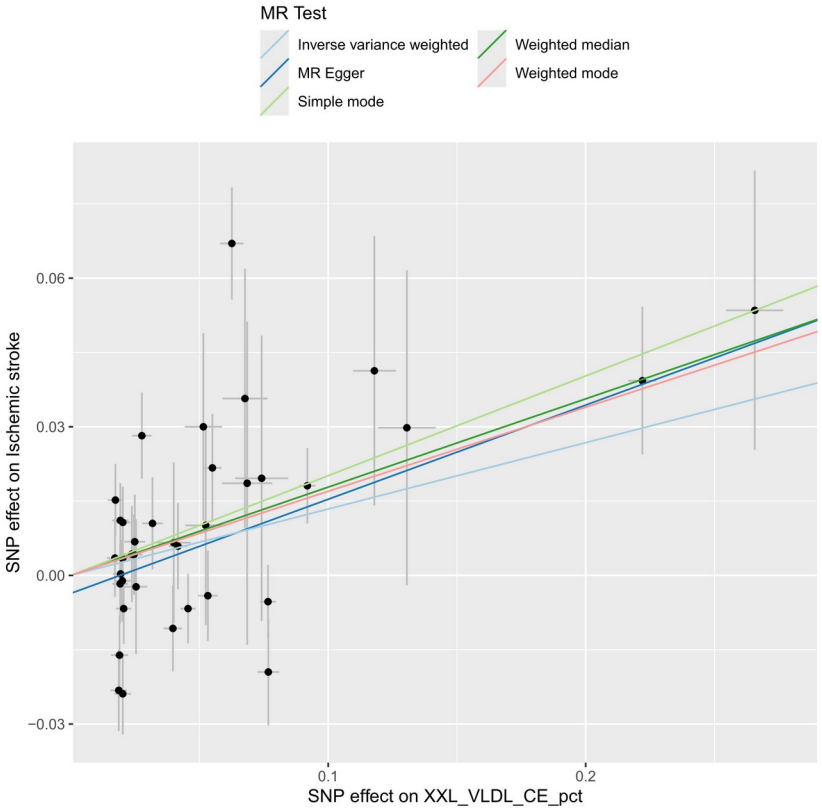
ID	Metabolite	Scatter plot
15	XL_HDL_L	
16	XL_HDL_C	

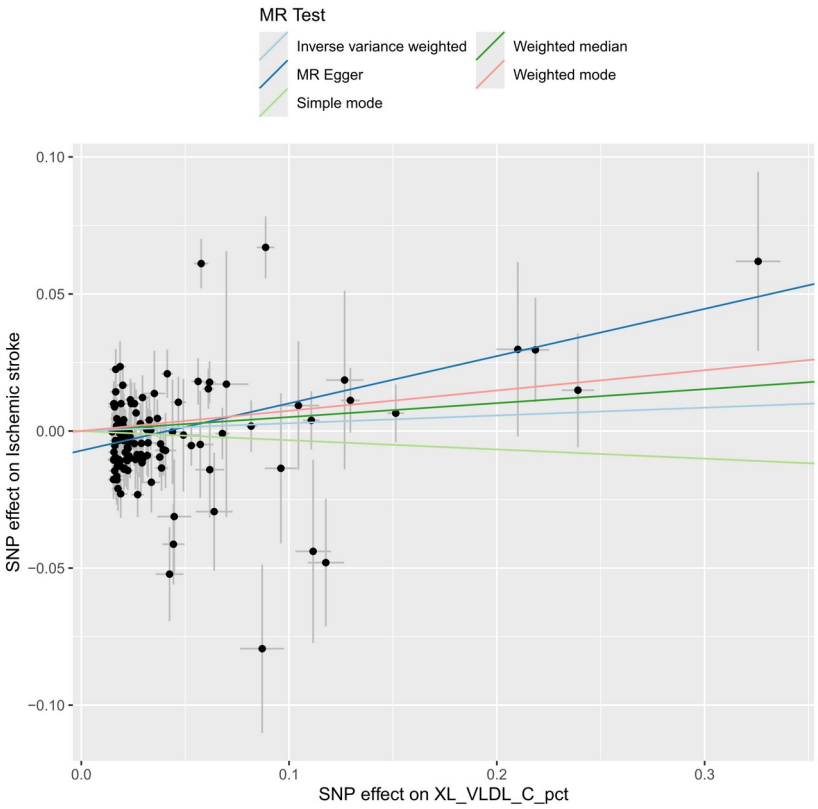
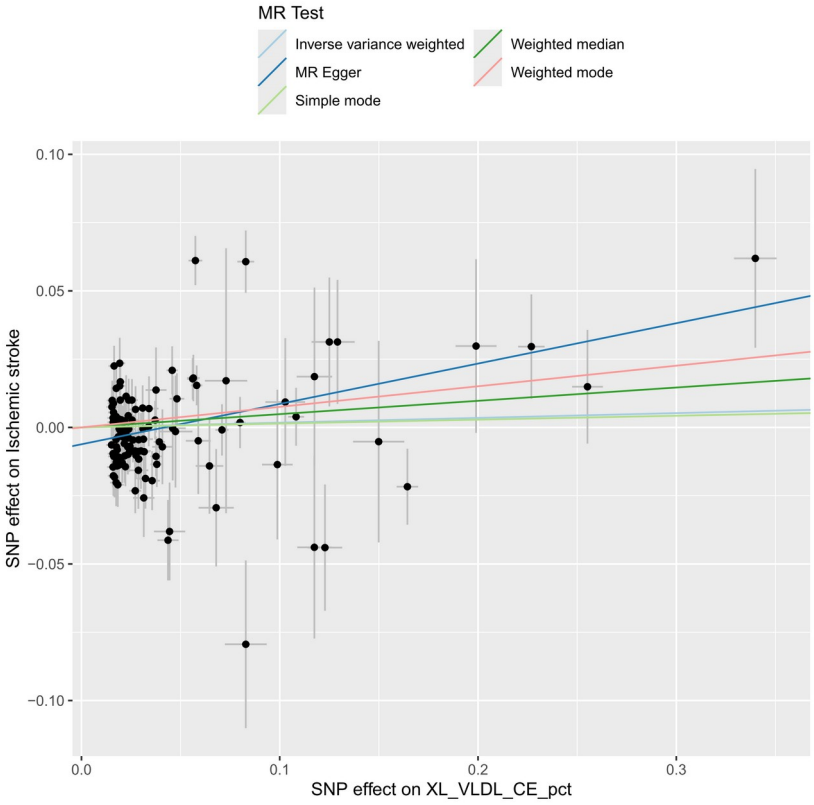
ID	Metabolite	Scatter plot
17	XL_HDL_FC	
18	M_HDL_L	

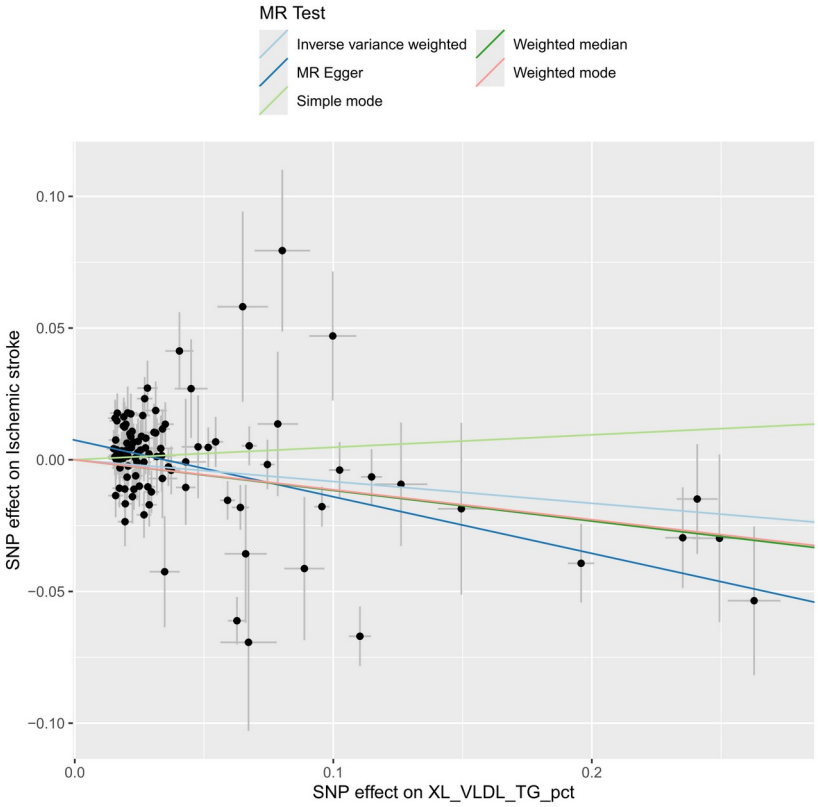
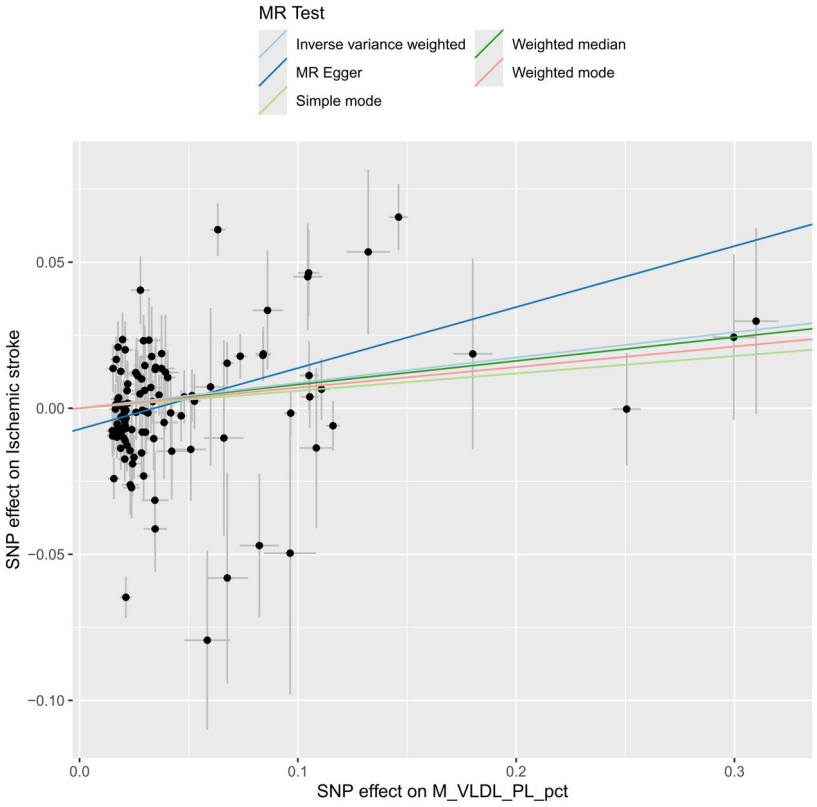
ID	Metabolite	Scatter plot
19	M_HDL_PL	
20	S_HDL_P	

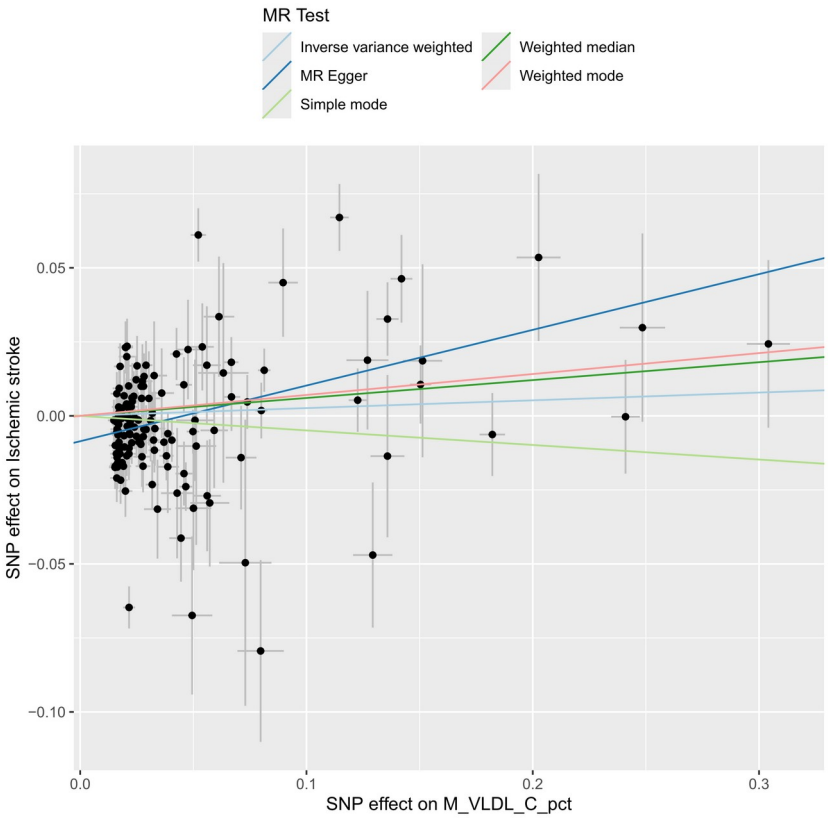
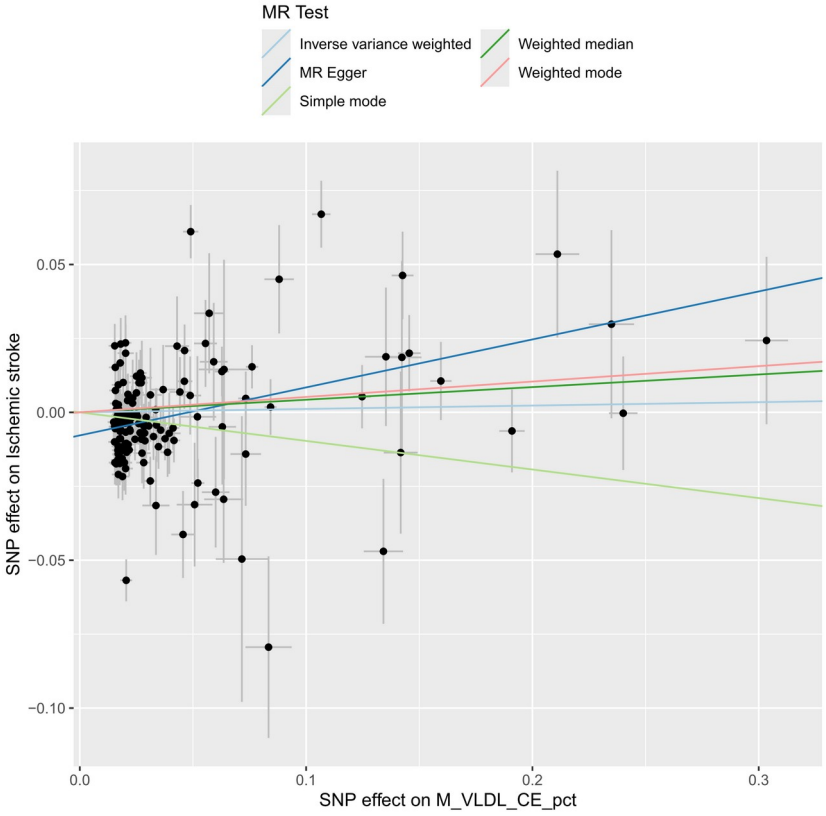
ID	Metabolite	Scatter plot
21	S_HDL_L	
22	S_HDL_PL	

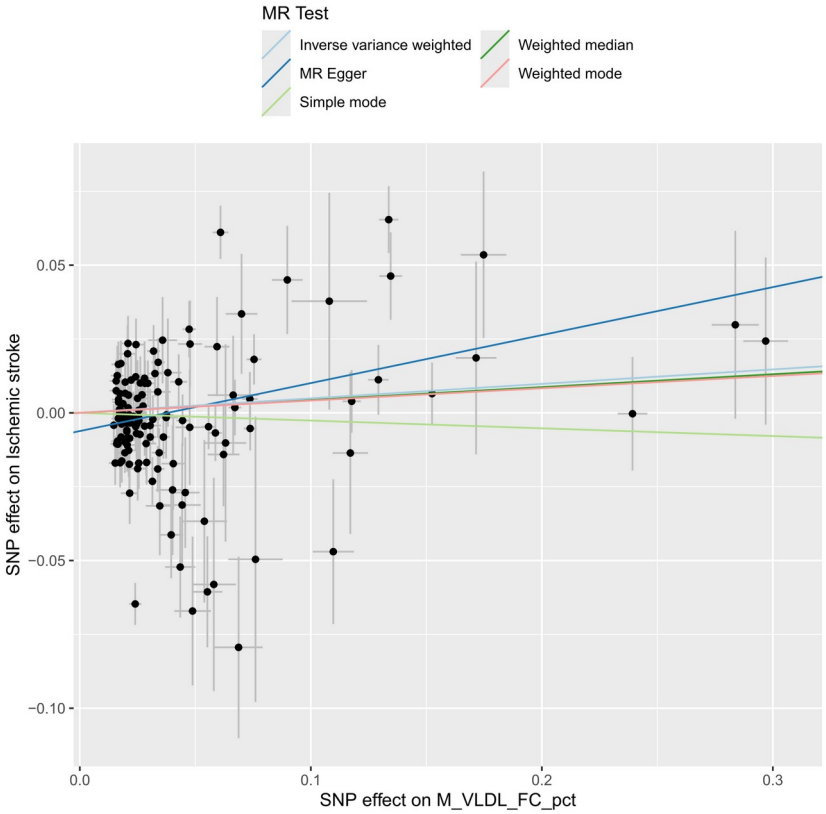
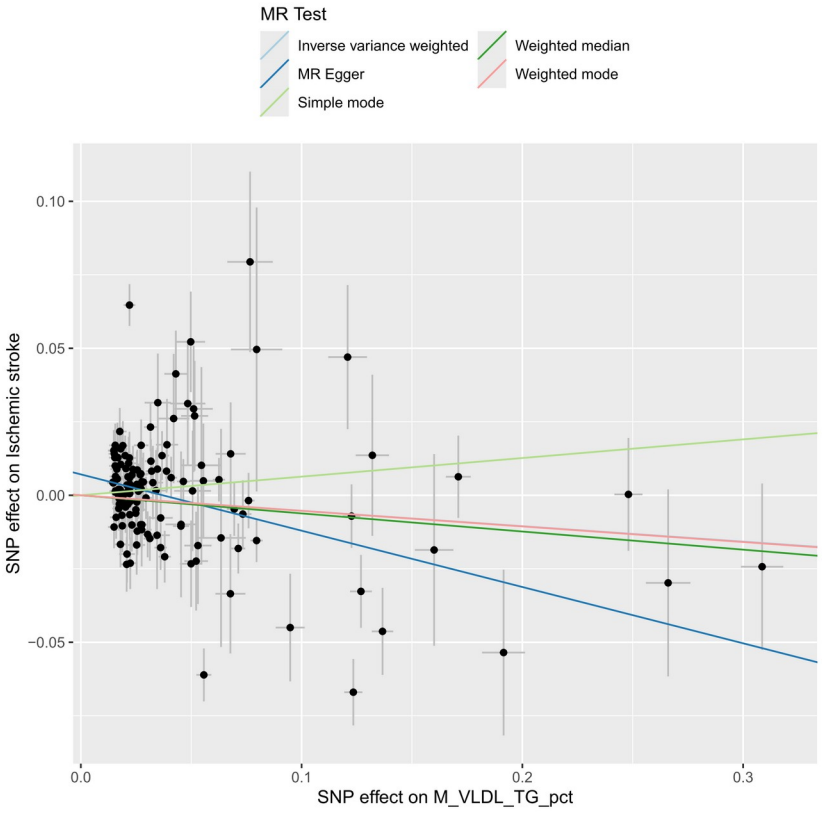
ID	Metabolite	Scatter plot
23	S_HDL_C	
24	S_HDL_CE	

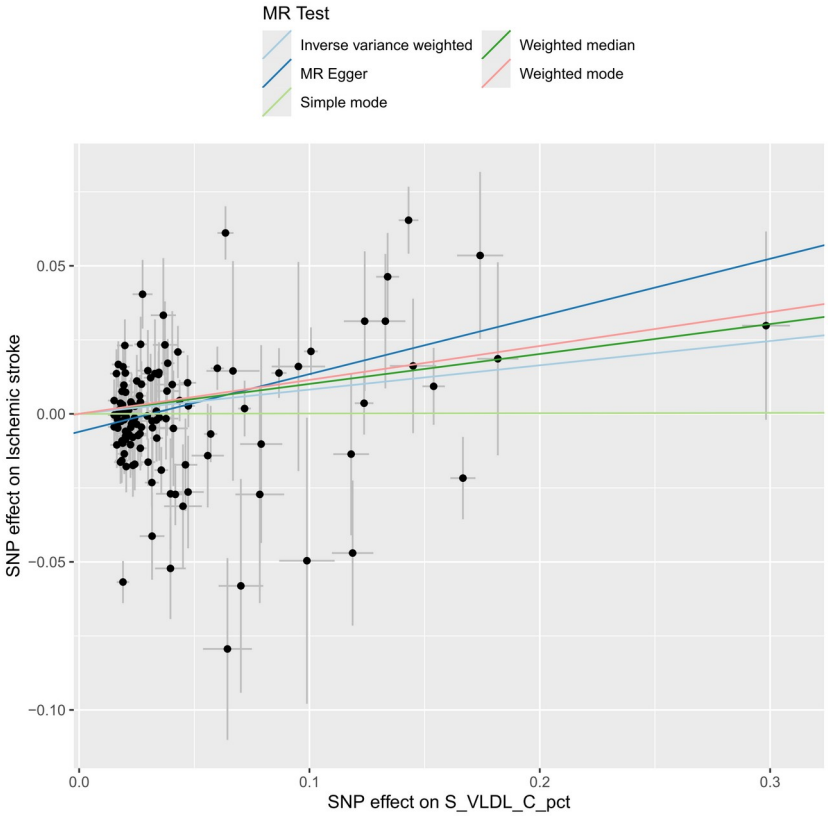
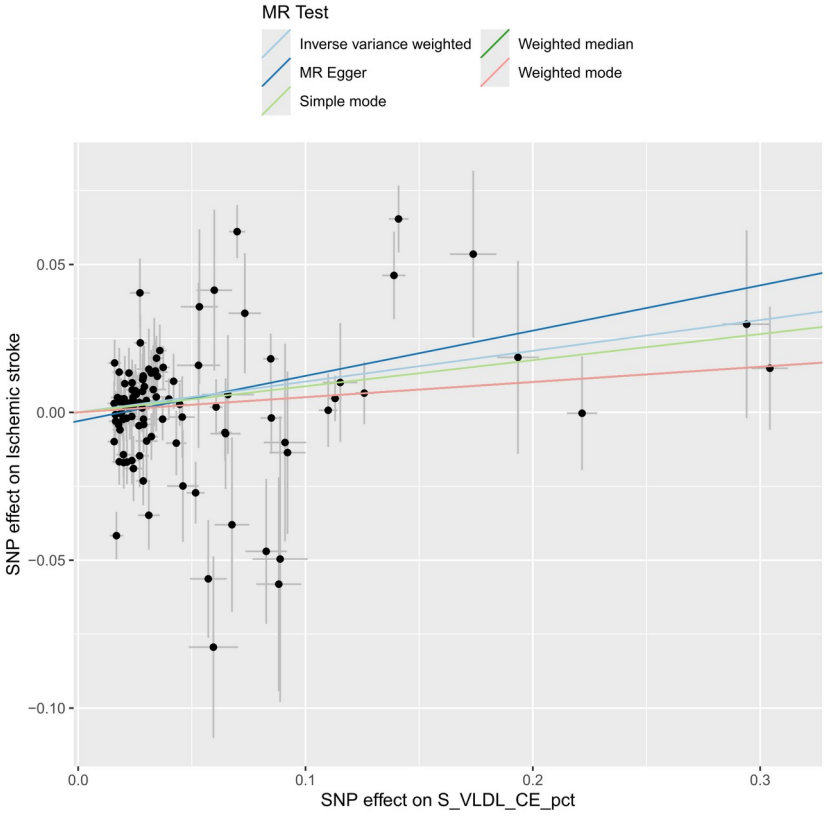
ID	Metabolite	Scatter plot
25	XXL_VLDL_C_pct	
26	XXL_VLDL_CE_pct	

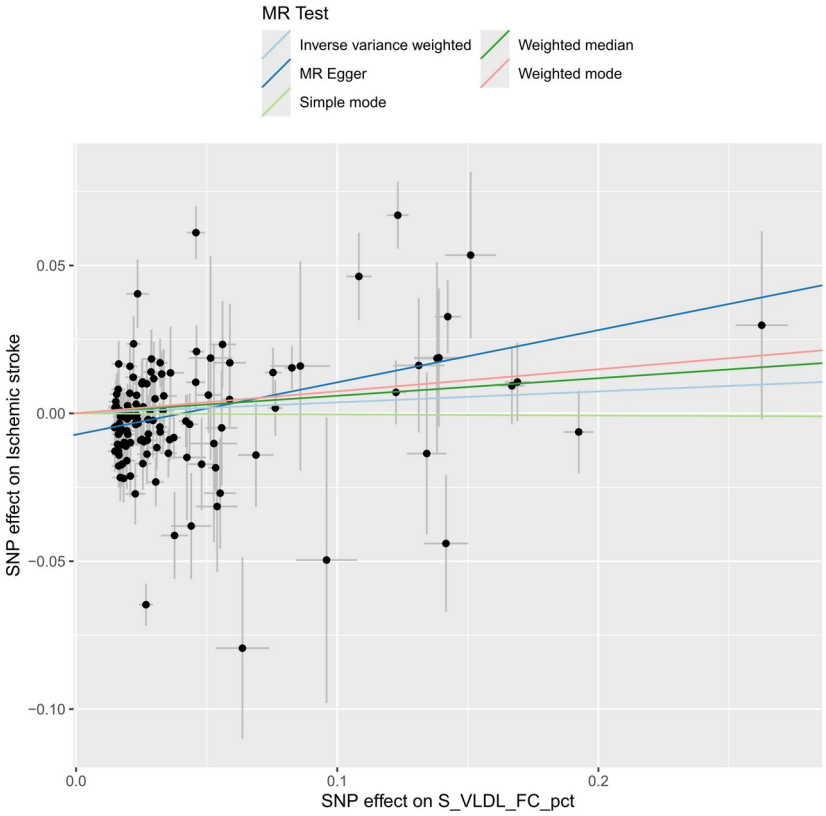
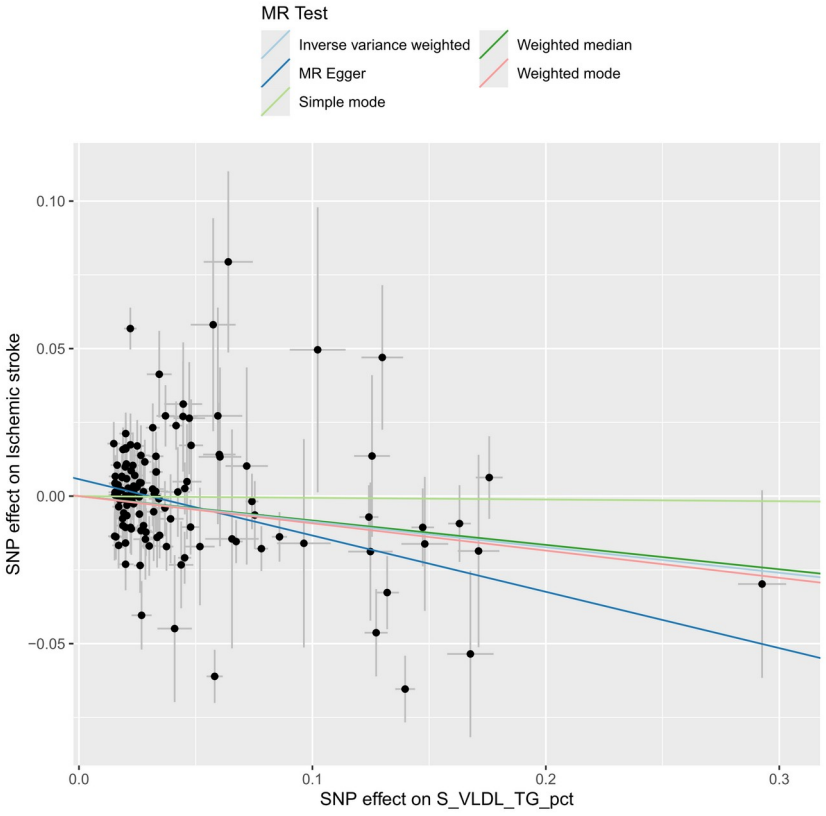
ID	Metabolite	Scatter plot
27	XL_VLDL_C_pct	
28	XL_VLDL_CE_pct	

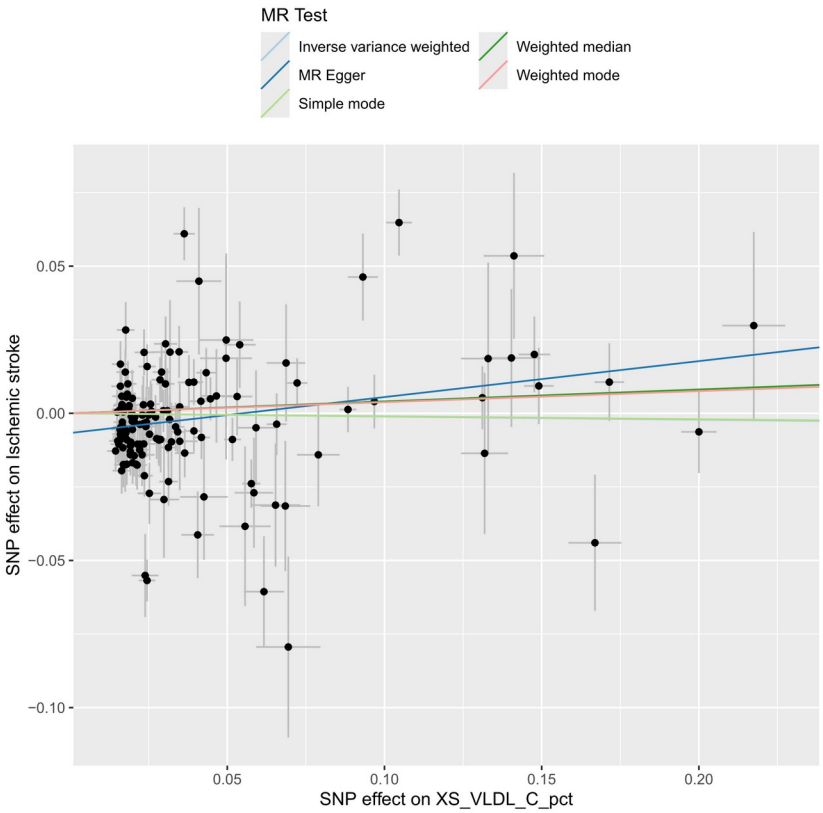
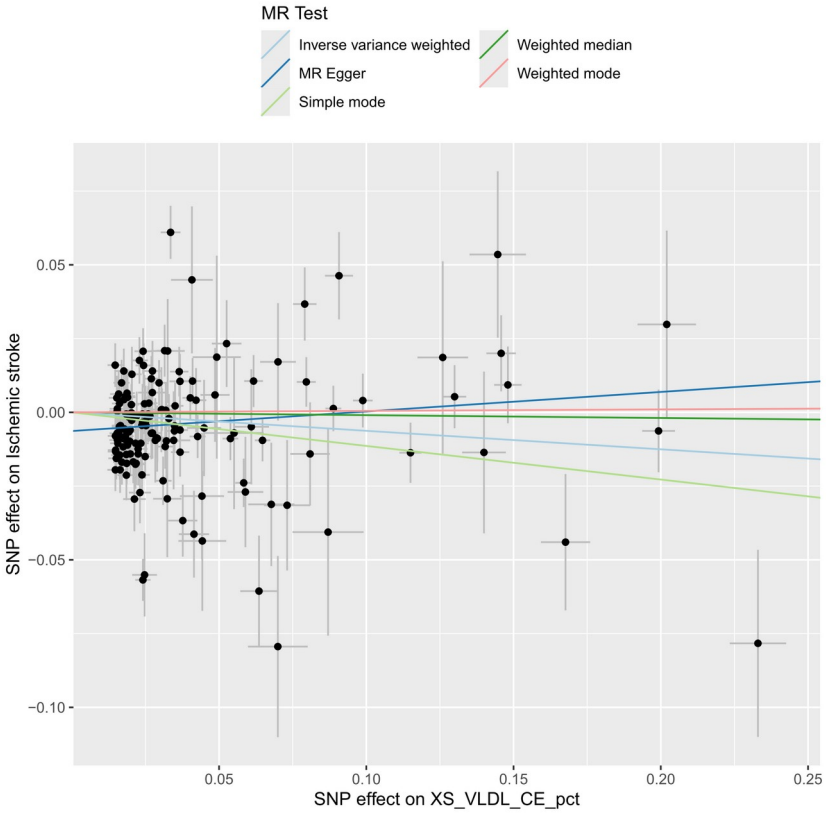
ID	Metabolite	Scatter plot
29	XL_VLDL_TG_pct	
30	M_VLDL_PL_pct	

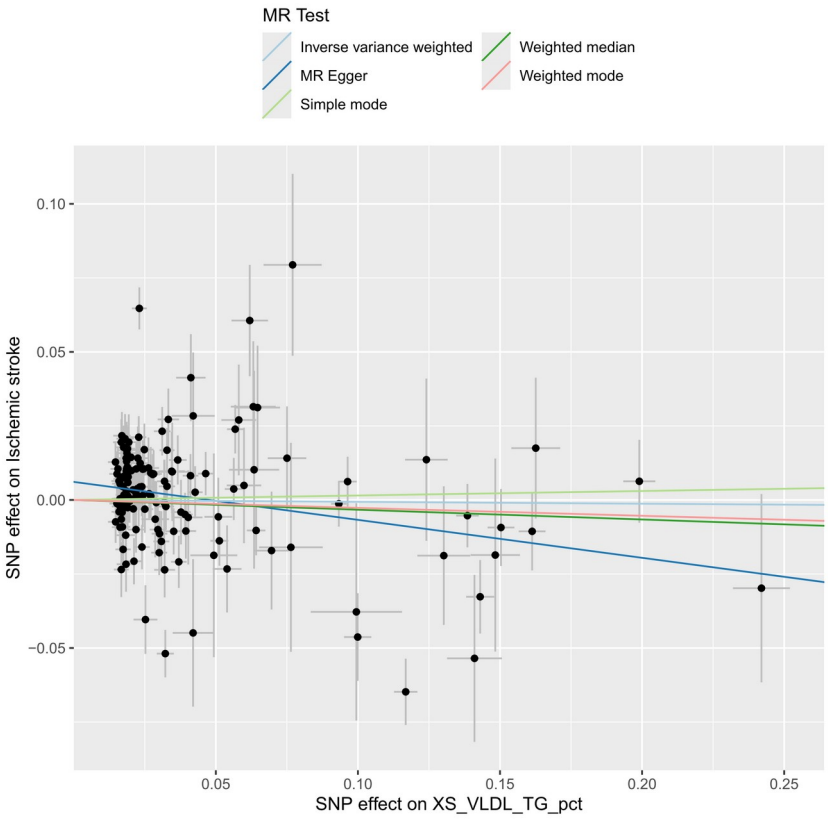
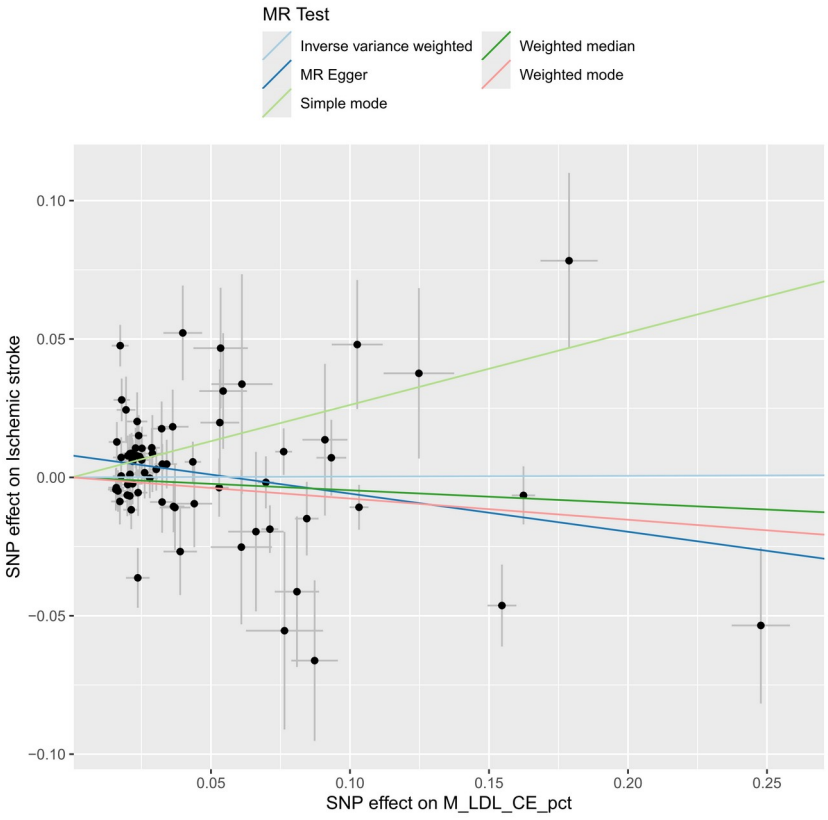
ID	Metabolite	Scatter plot
31	M_VLDL_C_pct	
32	M_VLDL_CE_pct	

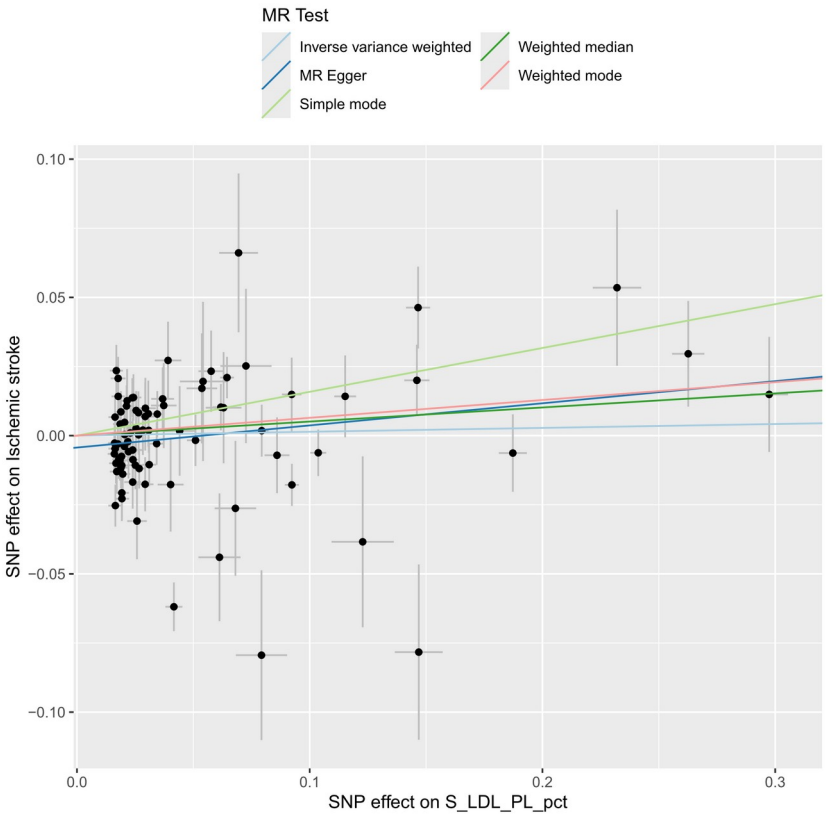
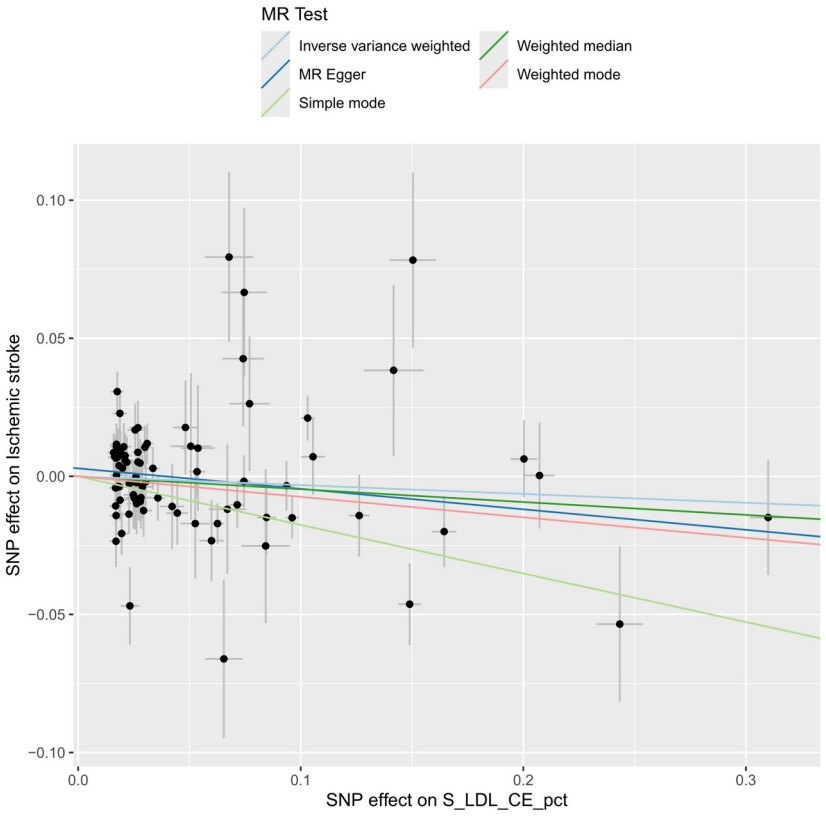
ID	Metabolite	Scatter plot
33	M_VLDL_FC_pct	
34	M_VLDL_TG_pct	

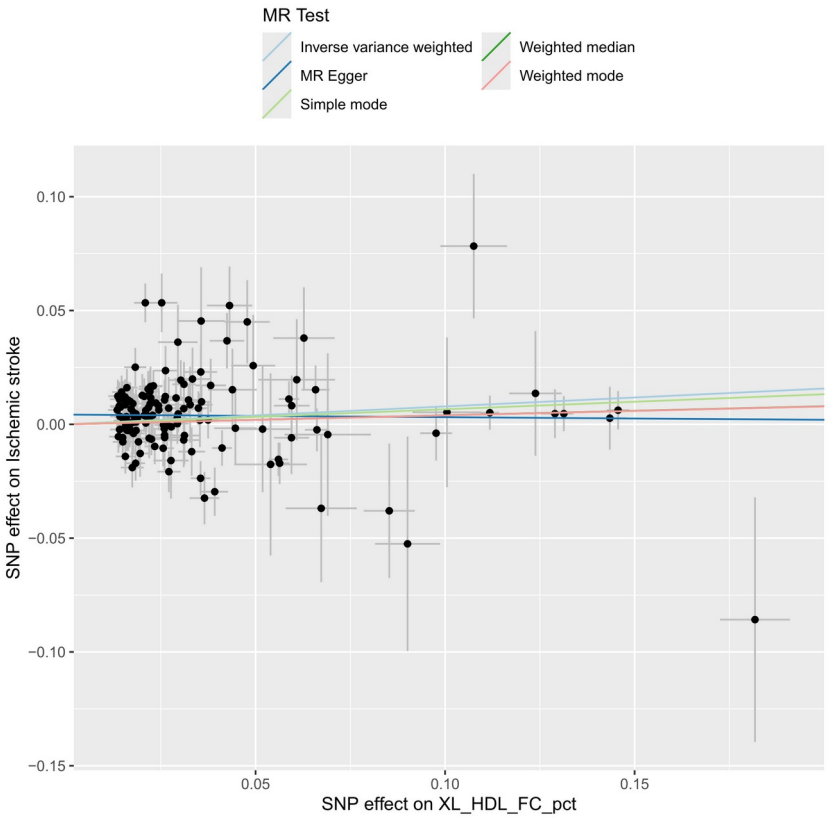
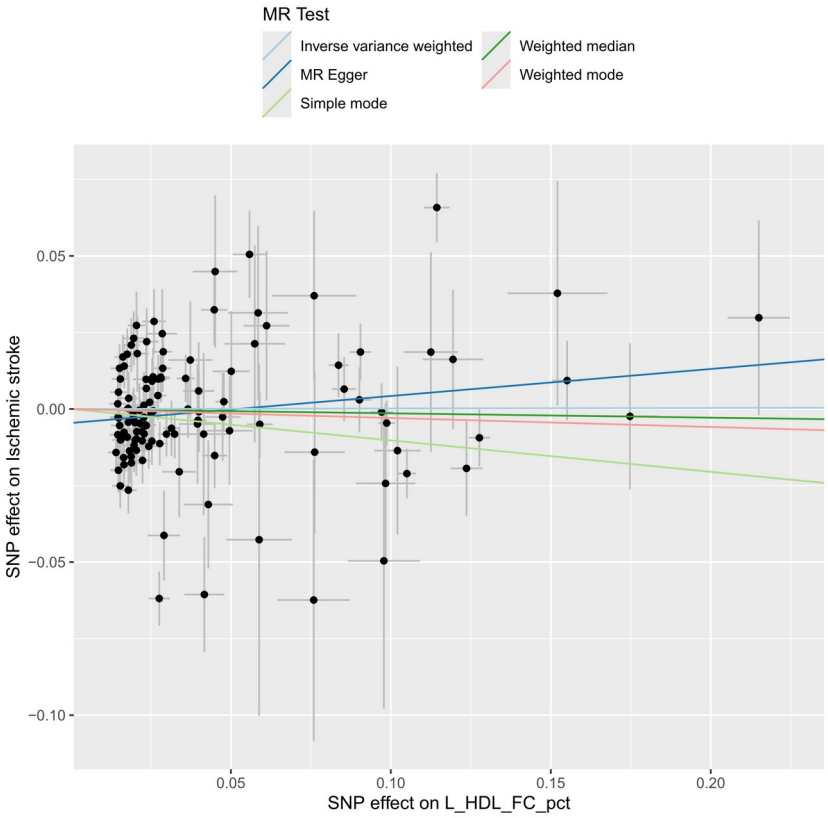
ID	Metabolite	Scatter plot
35	S_VLDL_C_pct	
36	S_VLDL_CE_pct	

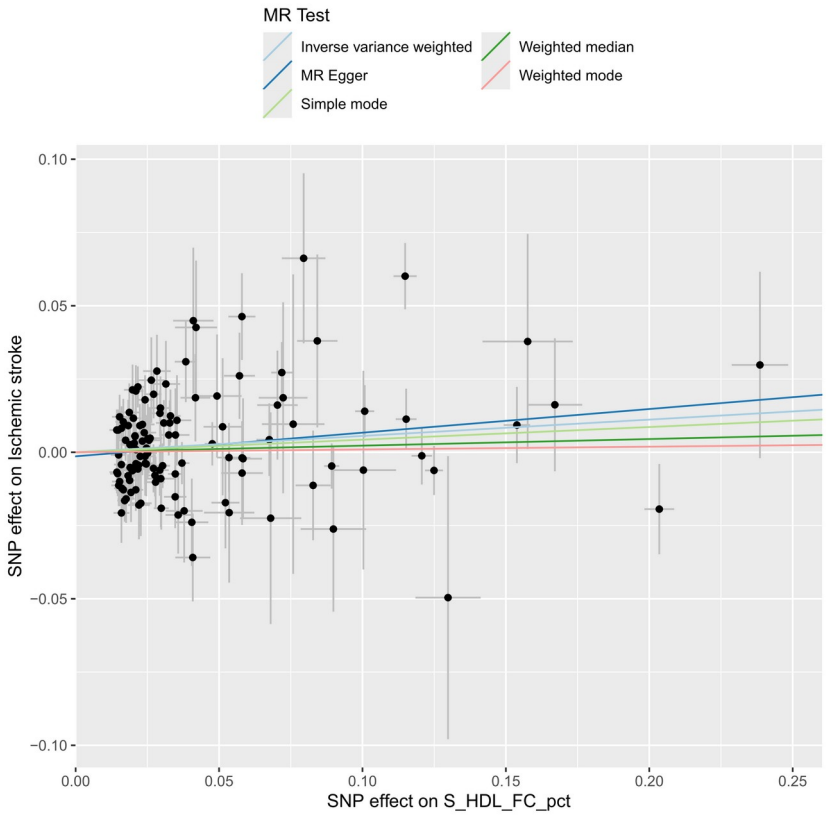
ID	Metabolite	Scatter plot
37	S_VLDL_FC_pct	
38	S_VLDL_TG_pct	

ID	Metabolite	Scatter plot
39	XS_VLDL_C_pct	
40	XS_VLDL_CE_pct	

ID	Metabolite	Scatter plot
41	XS_VLDL_TG_pct	
42	M_LDL_CE_pct	

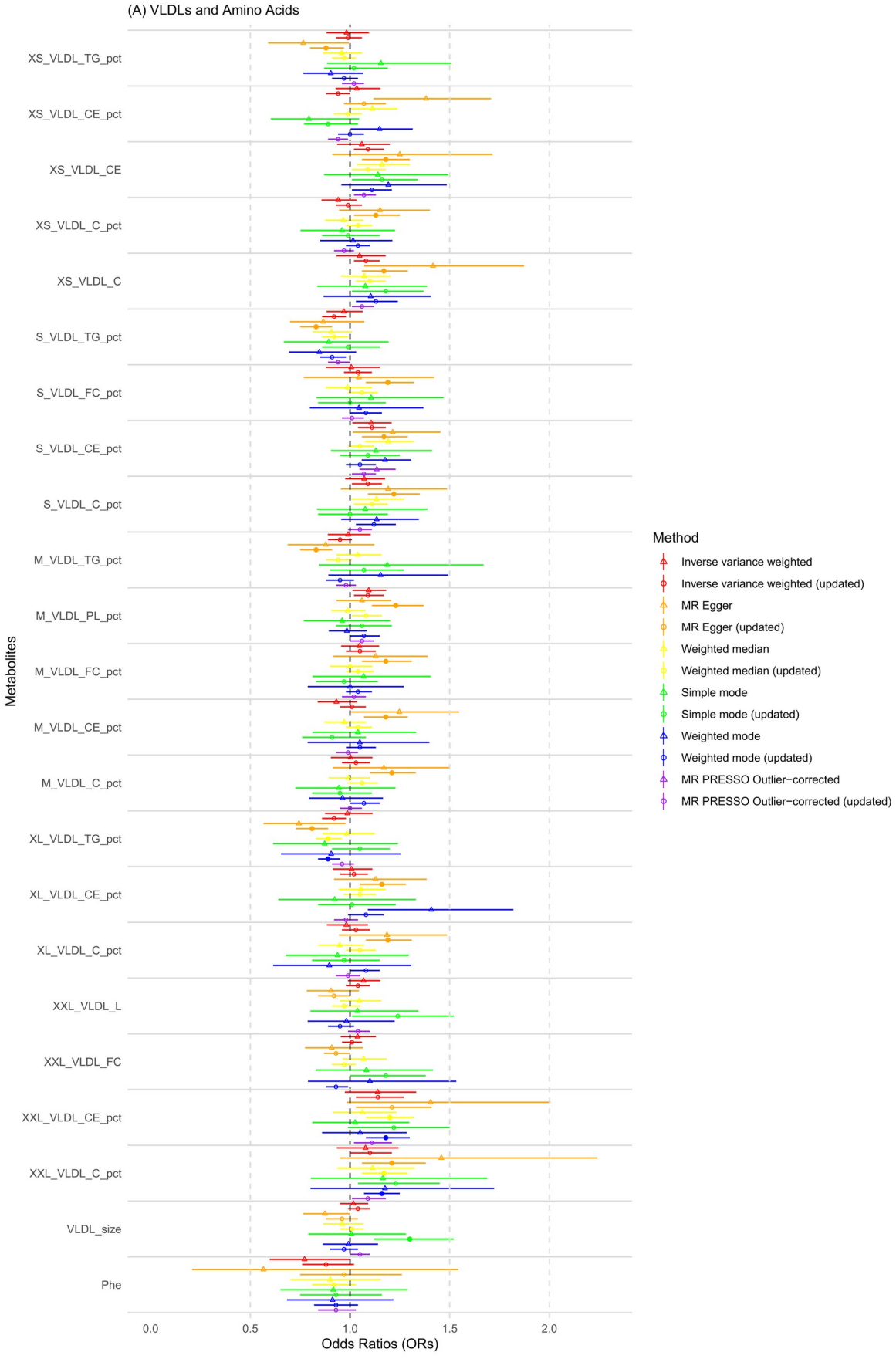
ID	Metabolite	Scatter plot
43	S_LDL_PL_pct	
44	S_LDL_CE_pct	

ID	Metabolite	Scatter plot
45	XL_HDL_FC_pct	
46	L_HDL_FC_pct	

ID	Metabolite	Scatter plot
47	S_HDL_FC_pct	

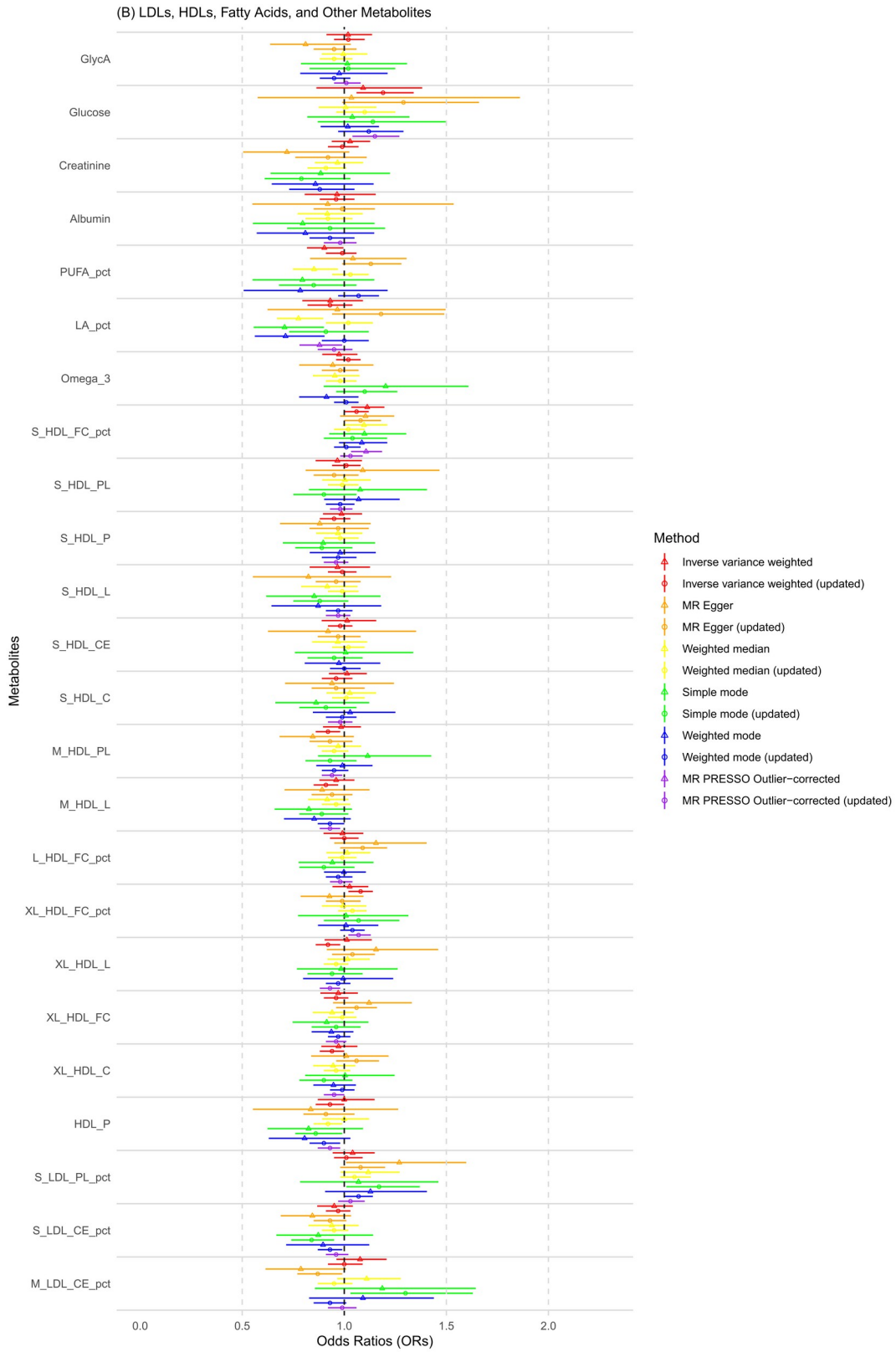
Supplementary Figure 6.5 Scatter plots of Mendelian randomisation results

Scatter plots with lines display the causal estimate for different Mendelian randomisation (MR) results with (1) single nucleotide polymorphisms effect on ischaemic stroke (log odds ratio) at the y-axis, and (2) single nucleotide polymorphisms effect on the metabolite (log odds ratio) at the x-axis. MR-PRESSO, MR pleiotropy residual sum and outlier. Abbreviations are in line with **Supplementary Table 3.1**.

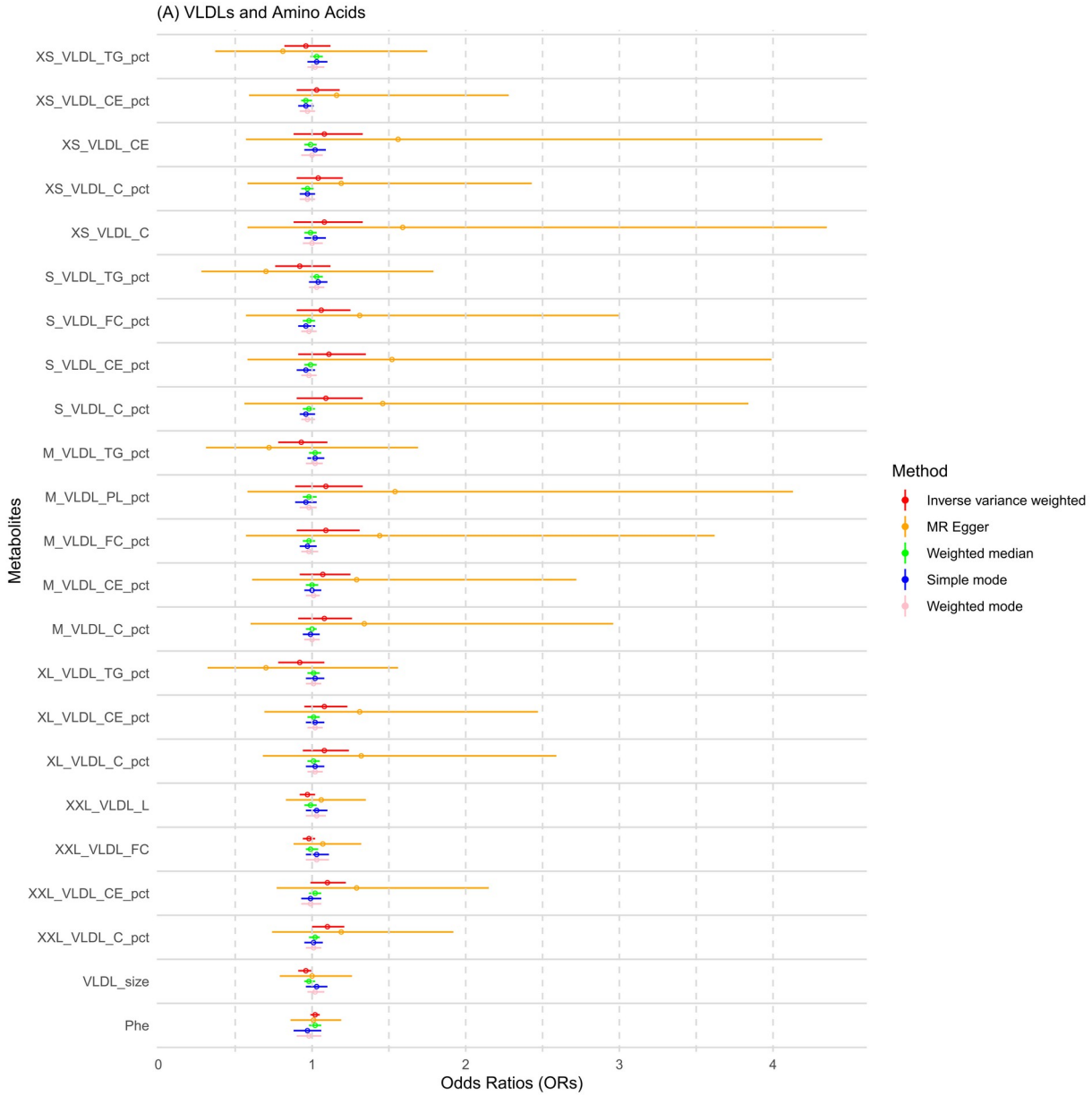

(A) VLDLs and Amino Acids
XS_VLDL_TG_pct
XS_VLDL_CE_pct
XS_VLDL_CE
XS_VLDL_C_pct
XS_VLDL_C
S_VLDL_TG_pct
S_VLDL_FC_pct
S_VLDL_CE_pct
S_VLDL_C_pct
M_VLDL_TG_pct
M_VLDL_PL_pct
M_VLDL_FC_pct
M_VLDL_CE_pct
M_VLDL_C_pct
XL_VLDL_TG_pct
XL_VLDL_CE_pct
XL_VLDL_C_pct
XXL_VLDL_L
XXL_VLDL_FC
XXL_VLDL_CE_pct
XXL_VLDL_C_pct
VLDL_size
Phe
Metabolites
0.0
0.5
1.0
1.5
2.0
Odds Ratios (ORs)
Method
Inverse variance weighted
Inverse variance weighted (updated)
MR Egger
MR Egger (updated)
Weighted median
Weighted median (updated)
Simple mode
Simple mode (updated)
Weighted mode
Weighted mode (updated)
MR PRESSO Outlier−corrected
MR PRESSO Outlier−corrected (updated)

Supplementary Figure 6.6 Comparison of Mendelian randomisation results from the current thesis (updated) and previous study

The first release of the nuclear magnetic resonance data (N~120,000) was used in the previous study by Harshfield et al. (2023). Circle dots, triangles, and lines denote the hazard ratios or odds ratios with 95% confidence intervals per one standard deviation higher metabolic levels between incident ischaemic stroke and individual metabolites. Filled circle dots and triangles represent significance after multiple testing correction. (A) VLDLs and amino acids, (B) LDLs, HDLs, fatty acids, and other metabolites. MR-PRESSO, MR pleiotropy residual sum and outlier. Abbreviations are in line with **Supplementary Table 3.1**.

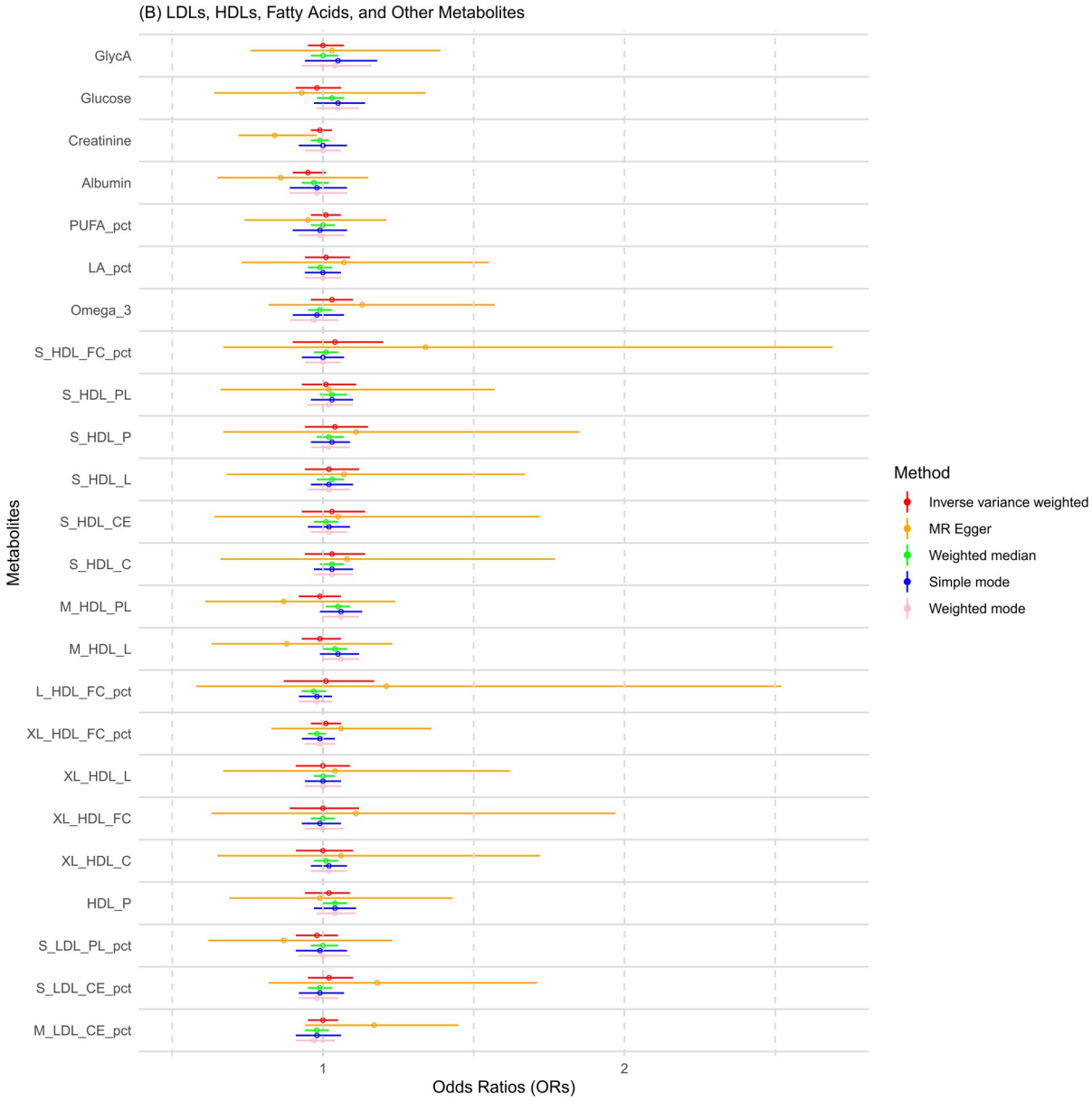


Supplementary Figure 6.6 (Continued) Comparison of Mendelian randomisation results from the current thesis (updated) and previous study



Supplementary Figure 6.7 Reverse Mendelian randomisation of ischaemic stroke with metabolites

Circle dots and lines denote the odds ratios with 95% confidence intervals per one standard deviation higher metabolic levels between incident ischaemic stroke and individual metabolites. (A) VLDLs and amino acids, (B) LDLs, HDLs, fatty acids, and other metabolites. Abbreviations are in line with **Supplementary Table 3.1**.



Supplementary Figure 6.7 (Continued) Reverse Mendelian randomisation of ischaemic stroke with metabolites