Metabolic Drivers of Dysglycemia in Pregnancy: Ethnic-Specific GWAS of 146 Metabolites and 1-Sample Mendelian Randomisation Analyses in a UK Multi-Ethnic Birth Cohort

Short title: Metabolic drivers of dysglycemia in pregnancy.

Authors: Harriett Fuller[1], Mark M Iles[2,3], J. Bernadette Moore[1], Michael A. Zulyniak[1].

Affiliations:

1 Nutritional Epidemiology Group, School of Food Science and Nutrition, University of Leeds, Leeds, UK.

2 Leeds Institute of Medical Research, University of Leeds, Leeds, UK.

3 Leeds Institute for Data Analytics, University of Leeds, Leeds, UK

*Corresponding Author: Michael Zulyniak. University of Leeds, School of Food Science and Nutrition, Leeds, UK. Email. m.a.zulyniak@leeds.ac.uk.

Abstract

Gestational diabetes mellitus (GDM) is the most common pregnancy complication worldwide and is associated with short- and long-term health implications for both mother and child. Prevalence of GDM varies between ethnicities, with South Asians (SAs) experiencing up to three times the risk compared to white Europeans (WEs). This study aimed to evaluate the causal role of metabolic characteristics in the ethnic-associated differences in gestational dysglycemia.

A one-sample Mendelian Randomisation (MR) was performed separately on 3688 SA and 3354 WE women (<28th week of pregnancy) from the Born in Bradford (BiB) cohort for 146 metabolites exposures for the outcomes of fasting glucose and 2-hr post glucose ($P \leq 1 \times 10^{-5}$ was considered significant). Additional GWAS and MR analyses on 22 composite measures of metabolite classes were also conducted.

Through an extensive GWAS analysis this study identified 15 novel genome-wide significant (GWS) SNPs associated with tyrosine in the *FOXN* and *SLC13A2* genes and 1 novel GWS SNP (currently in no known gene) associated with acetate in SAs. Through the utilisation of a MR analysis, 14 metabolites were found to be associated with postprandial glucose in WEs, while in SAs a distinct panel of 11 metabolites were identified. Furthermore, in WEs, cholesterols were most the common metabolite associated with dysglycemia, while in SAs saturated fatty acids were most common. Furthermore, in SAs a composite measure of the fatty acid class was also found to associate with 2-hour post glucose.

The presence of ethnic-specific causal relationships between a comprehensive set of metabolites and postprandial glucose measures (fasting glucose and 2-hour post glucose) in mid-pregnancy has been established in a UK SA and WE population. Future work should aim to investigate the biological mechanisms of metabolites on GDM risk and inform ethnically tailored GDM prevention strategies are required.

Introduction

Pregnancy is accompanied by a period of intense maternal metabolic adaptation to meet the energy demands of the foetus [1-3]. Mild maternal insulin resistance (IR) is a natural adaption to prioritise adequate glucose for the growing foetus [4]. However, if gestational IR exceeds healthy levels and glycaemia is uncontrolled, moderate IR can progress to gestational diabetes mellitus (GDM) [1, 5]. GDM is characterised by persistent maternal and foetal exposure to elevated levels of glucose, and places the mother and offspring at risk during pregnancy (i.e., macrosomia and haemorrhaging) and in later life (i.e. from obesity , type 2 diabetes (T2D) and cardiovascular disease) [6].

Globally, GDM is the most common pregnancy complication, affecting up to one in every seven births; however, its prevalence varies between ethnic groups, with South Asian (SA) women at 3-fold greater risk compared to white European (WE) women, irrespective of BMI and country of residence [5, 7]. Furthermore, SA women are more likely to develop T2D in later life following a GDM diagnosis [8]. Factors driving this disparity in prevalence are not fully understood but metabolism is thought to play a key role [6], given emerging evidence demonstrating (i) differences in metabolic profiles between GDM and non-GDM pregnancies in an ethnic-specific manner [9, 10]; and (ii) that a single dietary strategy to manage GDM across all ethnic groups appears ineffective [11, 12]. However, heterogeneity of reported metabolite-GDM associations between studies (due to differences in quantification methods, GDM diagnostic criteria [13, 14], ethnic and cultural groups), as well as residual confounding in observational studies have prevented complete understanding and, moreover, advancement to improved equitable care. In short, the field requires a clear and accurate understanding of ethnic-specific metabolic drivers of gestational

dysglycemia to inform appropriate and effective prevention and management strategies across ethnic groups.

Mendelian Randomisation (MR) is an instrumental variable technique that can provide estimates of causal associations between an exposure (such as metabolites) and outcome (such as dysglycemia) [15-17] . However, no study has yet utilised MR to establish the presence of casual associations between metabolites and measures of glycaemia at or before the 28 week of pregnancy in an ethnic-specific manner. The present study aimed to address this using the multi-ethnic Born in Bradford (BiB) cohort to identify ethnic-specific metabolic drivers of GDM.

Methods

Born in Bradford Cohort

Exposure data

BiB is a prospective longitudinal birth cohort that aimed to recruit all mothers receiving maternity care in the Bradford Royal Infirmary between 2007-2010 [18]. Bradford, a large city in the north of England, has high levels of deprivation and a large SA population, predominantly of Pakistani ancestry. A total of 12,453 women (mean maternal age 27.8) were recruited, 45% of which were of SA ancestry [18, 19]. All participants provided written consent and ethnical approval was obtained from the Bradford Research Ethics Committee (ref07/H1302/112) [18].

Fasted plasma sample collection and high-throughput metabolite quantification by automated NMR (Nightingale Health©; Helsinki, Finland) has been previously described and validated to a high accuracy [2]. Briefly, samples were taken by trained phlebotomists from BiB participants (26-28 weeks' gestation) and were processed in the absence of freeze-thaw cycles within 2.5 hours before storage at -80°C. One hundred and forty-six absolute measures of metabolites were included in the analysis following the removal of metabolites expressed as a percentage or ratio to minimise redundancy. In total, 10 overarching classes of metabolites were included in the analysis: lipoproteins (n=97), amino acids (n=9), apolipoproteins (n=2), cholesterols (n=8), fatty acids (n=8), glycerides and phospholipids (n=8), glycolysis related metabolites (n=4), ketone bodies (n=2), measures of fluid balance and inflammation (n=3) and measures of lipoprotein particle diameter (n=3). A full list of included metabolites can be found in **Supplementary Table 1**.

Outcome data

Participants were prior to GDM assessment and the 28th week of pregnancy. In BiB, individuals were diagnosed with GDM if either their fasting glucose or if 2-hour post-load glucose concentration exceeded 6.1 mmol/L or 7.8 mmol/L following a 75g oral glucose tolerance test (OGTT) [20]. The OGTT was performed in the morning following an overnight fast. To maximise power, MR analyses were performed using continuous metabolite values and fasting glucose and 2-hour post glucose. Fasting glucose and 2-hour post glucose values were log normalised prior to analysis.

Metabolite data

Information on metabolite data preparation has been described in full elsewhere [9]. In brief, 11,480 blood samples were metabolically profiled from BiB, 54 of which were excluded due to failure of any one of five Nightingale© quality control measures leaving 11,426 samples prior to imputation. Missing data was imputed via multiple imputation using the missMDA package in R [21].

After combining with postprandial glucose data, 3,693 SA and 3,377 individuals whose samples were taken before the 28th week of pregnancy were retained before outlier removal. Outliers were removed (those outside of 1.5 x IQR) for each metabolite in each ethnicity separately and metabolite values were normalised by taking the log, square root or normal score transformation (NST) as appropriate following the visual inspection of histograms and QQ plots. Following the removal of outliers, the number of individuals available for GWAS analysis of each metabolite was not the same (**Supplementary Table 2**) but was relatively consistent: for SAs the average sample size for each metabolite was 3622 (range 3472-3688), while for WEs it was 3301 (range 3158-3354). Information on gestational age at sample collection and parity was obtained from obstetric records. Ethnicity was self-reported or obtained from primary care records if missing. Individuals of a SA descent other than Pakistani were excluded from the analysis due to the small sample sizes of these populations.

Genetic data

Imputed genetic data were obtained from BiB. All samples were genotyped using two chips: the Infinium Global Sequencing Array-24 v.1 (GSA) (~640K SNPs) and the Infinium CoreExome-24 v1.1 BeadChip (~550K SNPs) [22]. Genetic data from the Illumina Global Sequencing Array (GSA) and Illumina CoreExome SNPs were

166 combined. Where SNPs were missing in >5% of individuals, they were excluded [22].
167 When evaluating imputed data, the R^2 value can be a measure of quality control as it
168 reflects to the estimated proportion of genetic variation maintained in the imputed data.
169 As a result, SNPs with an R^2 <0.9 were excluded prior to analysis.

170 GWAS analysis

171 Conventionally a GWAS assumes individuals are unrelated and the inclusion
172 of related individuals can potentially lead to spurious associations [23, 24]. However,
173 the removal of individuals from the BiB sample with close ancestry would substantially
174 reduce the sample size. In addition, high rates of consanguinity in the SA stratum of
175 the cohort makes relatedness difficult to assess [25]. As such, a GWAS mixed linear
176 model association (MLMA) analysis was conducted in PLINK (version 1.9) that allowed
177 for the inclusion of related individuals [24, 26-28]. MLMA models include a fixed effect,
178 adjusted covariates, and an additional random effect comprised of a variance-
179 covariance matrix that models the correlation (here relatedness) between individuals
180 to be accounted for [24, 26]. GWAS MLMA models were implemented using the *GCTA*
181 (Genome-wide Complex Trait Analysis) command line tool for each metabolite in both
182 ethnicities [29]. To increase power, MLMA-loo (leave-one-out) analysis was utilised,
183 preventing a SNP from being included in both the fixed and random effects
184 concurrently, thereby avoiding double fitting [26]. MLMA models also included parity
185 and principal components (PC) 1 and 2 to account for population stratification
186 (**Supplementary Figure 1, Supplementary Table 3**). Gestational age, which showed
187 little variation, was not included in the modelling (median gestational age SA = 184
188 days, IQR= 182-186.7, median gestational age WE = 184 days, IQR= 182-187).
189 Genomic inflation factors (λ) were calculated for all models for a range of minor allele

frequency (MAF) cut-offs (MAF <0.001, 0.001≤ MAF < 0.005, 0.005 ≤ MAF < 0.01, 0.01 ≤MAF < 0.05, 0.05 ≤ MAF < 0.1, and MAF ≥ 0.1) in order to minimise data loss while also minimising false positives. λ ≥ 1.1 was considered indicative of genomic inflation. [30, 31] A MAF cut-off of <0.05 was the least stringent cut-off found to reduce λ to ~1 meaning this cut off was utilised in the analysis (**Supplementary Table 4**).

When a SNP was found to be associated with a metabolite value in only one ethnicity a fixed effect inverse-variance weighted meta-analysis was implemented to assess the heterogeneity (via the I^2 statistic) of identified associations between ethnicities and to see if the SNP retained significance in a larger sample. Meta-analyses were conducted within the command-line tool METAL [32].

One-sample MR

Genetic Instruments

One-Sample MR was conducted for all 146 metabolite values in both ethnic groups using SNPs identified as significant at a genome-wide suggestive level (p-value ≤ 1×10^{-5}) in the GWAS. Metabolites were grouped into their overall classes and SNPs in each class were thinned by linkage disequilibrium (LD) (R^2<0.2) via the NIH LDlink online tool (https://ldlink.nci.nih.gov) reducing the overlap of instruments in each class [33, 34]. For individuals of WE ancestry, all European (EUR) and South Asian (SAS) populations in LD link (software that utilises 1000 Genome data) were used to estimate LD due to the expected similarity in their LD structure allowing for an increase sample size and resultant improvement in the accuracy of LD estimates [35]. Similarity between 1000 Genome SA samples and BiB samples was assessed using principle components analysis (PCA) in PLINK (version 1.9) due to the fact that Pakistani

214 samples from BiB originate from a different region of Pakistan (the Mirpur Region) from
215 the 1000 Genome SA samples [27, 28]. This is of particular importance in SA
216 populations as even geographically close populations can have differing allele
217 frequencies due to differing Biraderi ('Brotherhood') membership between population
218 subgroups. Biraderi membership is assigned at birth, is an indicator of male lineage
219 as well as social-occupational status which largely governs partner choice and can
220 result in higher levels of consanguinity in the population [22]. PCA plots were created
221 using the *ggplot2* package in R studio (version 4.0.2) [36, 37]. No clear separation in
222 SA BiB samples and SA 1000 Genome (1000G) samples was identified indicating that
223 LD estimates obtained from 1000G was suitable for use in BiB (**Supplementary**
224 **Figures 2-4**).
225

226 ## Analysis

227 Genetic Risk Scores (GRS) were created in PLINK (version 1.9) for each
228 metabolite with each SNP receiving a weight based on its estimated effect size on the
229 metabolite obtained from the GWAS [38]. One-sample MR was then performed by
230 Two-Stage Least Squares regression (TSLS; *ivpack*, ivreg, and *AER* packages in R
231 version 4.0.2) to obtain a causal estimate for the effect of each metabolite value on
232 the log-normalised continuous measures of fasting glucose and 2-hour post glucose
233 following a 75g oral glucose tolerance test (OGTT) [36, 39, 40]. Here, the level of a
234 metabolite is regressed on its respective GRS and, subsequently, the outcome is
235 regressed onto these fitted GRS-metabolite values in the second stage. All MR results
236 have been reported according to STROBE-MR guidelines [41].

237 When significant associations were identified leave-one-out analysis was
238 performed. For this, SNPs were removed sequentially from the instrument and

239 changes to the effect estimate and F-statistic was assessed. If the exclusion of a SNP
240 was found to alter either the effect estimates or F-statistic (through the visualisation of
241 forest plots) it is possible that the SNP is influencing the outcome via an alternative
242 pathway to other SNPs, potentially highlighting a violation of the 2nd or 3rd MR
243 assumption. To further test for violations of these assumptions, included SNPs were
244 searched for in both the Phenoscanner and GWAS Catalog databases to identify
245 previously identified associations [42-44]. In both databases a p-value ≤ 1×10^{-5} was
246 interpreted as indicative of an additional association. Differences between MR and
247 linear regression results were also evaluated via the Wu-Hausman statistic to assess
248 deviation of the instrumental variable estimate from the ordinary least squares (OLS)
249 estimate [45]. Deviations in these two measures can indicate either confounding in the
250 OLS estimate (indicating a need for MR) or violations of the MR assumptions due to
251 pleiotropy.
252

253 **Post-hoc power analyses**

254 For metabolites that associated with a measure of postprandial glucose in only
255 one ethnic group, *post-hoc* power analyses were performed using the mRnd CNS
256 genomics tool (https://shiny.cnsgenomics.com/mRnd/) to assess whether the absence
257 of an association in the alternate ethnicity was due to limited power [46]. Observational
258 and 'true' associations required by the tool were obtained by performing linear
259 regression of the outcome on the metabolite and obtaining unadjusted and adjusted
260 estimates (adjusted for maternal age (years), BMI (continuous), smoking status,
261 multiple pregnancy, parity, and gestational age) respectively. Due to the *post-hoc*
262 nature of this analysis, additional power analyses could be conducted assuming the
263 MR estimate to be the true causal effect in the MR calculation. This analysis was

264 performed in the non-significant population for each metabolite associated in only one
265 ethnicity. If power was found to be adequate (80%) at the 5% level ($\alpha = 0.05$) power
266 was also assessed at the 1% level ($\alpha = 0.01$).
267
268

269 Results

270 GWAS of Metabolite Measures

271 A total of 6184 SNPs were associated with at least one metabolite in WEs at
272 the suggestive level (1×10^{-5}), with 2616 (42.3%) SNPs being associated with a single
273 metabolite measure. However, no SNPs were identified below the genome-wide
274 significant level (p-value $<5 \times 10^{-8}$) in WEs.
275 Fewer SNPs were identified at the suggestive level in SAs, with 3685 SNPs
276 SNP-metabolite associations in total, of which 1544 (41.9%) SNPs being associated
277 with only one metabolite measure. SNP associations were identified for 138/146
278 (94.5%) metabolite exposures in SAs (**Supplementary Table 5**). No SNP was
279 identified as being associated at the suggestive level in both ancestries, although
280 shared genomic regions were identified between ancestries (**Supplemental Excel**).
281 To evaluate the possibility of shared genomic predictors of metabolites, a
282 pooled meta-analysis of effect estimates in both ethnicities was performed. For 90
283 metabolite values, no associations were found to exceed the genome-wide suggestive
284 level ($p<10^{-5}$) following meta-analyses of both ethnicities. SNP associations were
285 identified at the suggestive level for four metabolite measures (concentration of XL-
286 HDL, total lipids in M-VLDL, mean density of VLDL and citrate) despite these differing
287 in direction of effects in SAs and WEs. In addition, 4 SNPs were associated with

288 alanine, despite these SNPs initially being associated with alanine only in the SA
289 population. These SNPs (rs12256633, rs17121228, rs7096521, rs12240368) are all
290 found on chromosome 10 in the receptor gene *SORCS1* and have not been
291 associated with alanine levels previously [47, 48].

292 MR results

293 After LD thinning, genetic instruments were available for all 146 metabolites in
294 WEs and for 136/146 (93.2%) metabolites in SAs. In WEs, 1040 SNPs were retained
295 following LD thinning including 423 (40.67%) that were unique to an individual
296 metabolite. Fewer SNPs were identified in SAs, where 383 SNPS remained after LD
297 thinning, 195 (50.9%) of which were unique to a single metabolite. (**Supplementary**
298 **Table 5**). 2.7% of included genetic instruments (4 metabolites) and 12.5% (9
299 metabolites) of included genetic instruments in WEs and SA respectively had an F-
300 statistic < 10, indicating that most instruments were at low risk of weak instrument bias
301 [49]. The average F-statistic for WEs instruments was 72.4, while in SAs it was
302 considerably lower at 26.7.
303

304 White Europeans

305 Two metabolite values, leucine and mean density of HDL lipoproteins (HDL_D),
306 were associated with both fasting glucose and 2-hour post glucose (**Table 1**).
307 Specifically, a 1mmol/L increase in blood leucine associated with lower fasting glucose
308 (-0.193 mmol/L, 95% CI -0.069, -0.319) and 2-hour post glucose (-0.443 mmol/L, 95%
309 CI -0.113, -0.774). Likewise, a 1nm increase in mean diameter of HDL associated with
310 lower fasting glucose (-0.082 mmol/L, 95% CI 0.026, 0.138) and 2-hour post glucose

311 (-0.191, 95% CI 0.043, 0.339 mmol/L) **(Supplementary Figures 4-5)**. No other
312 metabolites were associated with both measures of glucose in WEs.

313 For fasting glucose, an increase of 1mmol/L total cholesterol in M-HDL (M-HDL-
314 C) and cholesterol esters in M-HDL (M-HDL-CE) were associated with lower fasting
315 glucose measures (-0.189 mmol/L, 95% CI -0.021, -0.358, and -0.327 mmol/L, 95%
316 CI -0.069, -0.586 respectively). For 2-hr post-glucose, 8 metabolite values were
317 positively associated with this (HDLC, HDL2C, HDL3C, triglycerides in XS-VLDL,
318 cholesterol esters in XL-HDL, total concentration of L-HDL, total lipids in L-HDL and
319 cholesterol esters in S-HDL) and one (total concentration of S-LDL) was negatively
320 associated (**Table 1**). Cholesterol metabolites, measures of total cholesterols in
321 lipoproteins and total cholesterols in lipoproteins were the most common types of
322 metabolite class to be associated with postprandial glucose in WEs, with leucine being
323 the only amino acid identified. Wu-Hausman p-values < 0.05 indicate deviations of the
324 instrumental variable estimate from the OLS estimate **(Table 1)**.
325
326
327
328
329
330
331
332
333
334
335

	Class	Metabolite	Outcome	F statistic	β estimate (95% CI)	WuH
Lower Glucose	S-LDL	S-LDL-P	2-hour post	41.7	-1000 (-20, -1984)	0.017
	Amino Acids	Leucine	Fasting glucose	67.3	-0.193 (-0.069, -0.319)	0.005
			2-hour post		-0.443 (-0.113, -0.774)	0.008
	M-HDL	M-HDL-CE	Fasting glucose	62.4	-0.327 (-0.069, -0.586)	0.043
		M-HDL-C	Fasting glucose	117	-0.189 (-0.021, -0.358)	0.117
Increase Glucose	Lipoprotein Density	HDL_D	Fasting glucose	131	0.082 (0.026, 0.138)	0.004
			2-hour post		0.191 (0.043, 0.339)	0.024
	L-HDL	L-HDL-P	2-hour post	108	220 (41.3, 397)	0.02
		L-HDL_L	2-hour post	131	0.264 (0.062, 0,464)	0.014
		L-HDL-C	2-hour post	120	0.279 (0.012, 0.544)	0.048
	Cholesterol	HDL2C	2-hour post	103	0.288 (0.007, 0.583)	0.025
		HDLC	2-hour post	90.6	0.296 (0.007, 0.583)	0.047
		HDL3C	2-hour post	66.6	1.58 (0.002, 3.15)	0.074
	XL-HDL	XL-HDL-CE	2-hour post	109	0.541 (0.079, 1.00)	0.039
	XSVLDL	XS-VLDL-TG	2-hour post	87.8	0.841 (0.098, 1.58)	0.042
	S-HDL	S-HDL-CE	2-hour post	41.9	1.78 (0.448, 3.11)	0.007

Table 1: Significant MR results in white Europeans.

Glucose measures are expressed as mmol/L. HDL_D: mean diameter of HDLs (nm). HDLC: total cholesterol in HDL (mmol/L). HDL2C: total cholesterol in HDL2 (mmol/L). HDL3C: total cholesterol in HDL3 (mmol/L). L-HDL-C: total cholesterols in L-HDL (mmol/L). L-HDL_L: total lipids in L-HDL (mol/L). L-HDL-P: concentration of L-HDL (mol/L). M-HDL-C: total cholesterol in M-HDL (mmol/L). M-HDL-CE: cholesterol esters in M-HDL (mmol/L). S-HDL-CE: cholesterol esters in S-HDL (mmol/L). S-LDL-P: concentration of S-LDL (mol/L). XL-HDL-CE: cholesterol esters in XL-HDL (mmol/L). XS-VLDL-TG: triglycerides in XSVLDL (mmol/L). WuH: Wu-Hausman p-value.

354 ## White Europeans: Sensitivity analyses

355 For 6 of 13 metabolite values, leave-out one analyses maintained significance
356 ($P \leq 0.05$) indicating that no individual SNP was driving the identified associations in
357 WEs: leucine, mean diameter of HDL, total lipids in L-HDL, cholesterol esters in S-
358 HDL and cholesterol esters in M-HDL **(Supplementary Figure 7**). For the remaining
359 8 metabolites, β values were consistent across leave-one-out analyses although not
360 all associations remained significant. Additionally, for 12/13 metabolites (all but M-
361 HDL-CE), the F-statistic did not substantially differ through the exclusion of individual
362 SNPs from the instruments, which suggests they were not substantially driven by a
363 single SNP **(Supplementary Figure 8**). The exception to this was cholesterol esters
364 in M-HDL, where the exclusion of rs2138011 or rs739018 increased the F-statistic.

365 Three of the metabolites (leucine, L-HDL-L, L-HDL-C) that were associated with
366 postprandial glucose in WEs included a SNP that previous studies have associated (p
367 $\leq 1 \times 10^{-5}$) with at least one potential confounder (BMI, hypertension or waist
368 circumference) **(Supplementary Table 7**). The removal of these SNPs from the
369 instrument did not impact the significance of the associations identified for leucine or
370 L-HDL-L **(Table 2)**. However, for L-HDL-C instrument, the exclusion of two SNPs
371 (rs5576825 and rs6811162) previously associated with a potential confounder (waist
372 circumference and hypertension respectively) resulted in non-significant association
373 between L-HDL-C and 2-hour post glucose. Importantly, for both SNPs, it is
374 conceivable that the confounders could reside on their causal pathway (i.e., vertically
375 pleiotropic, where L-HDL-C effects 2-hr post-prandial glucose through its effect on
376 weight gain) rather than be in horizontal pleiotropy and may, therefore, violate the 2nd
377 MR assumption [49].

Ethnicity	Metabolite	SNP	Gene	Associated confounder	Initial		Confounder removal	
					β estimate (95% CI)	WuH	β estimate (95% CI)	WuH
WE	Leucine	rs2984433	ACTG1P9	BMI, Obesity class 1, weight	-0.193 (-0.319, -0.068) [FG]	0.004	-0.203 (-0.339, -0.068) [FG]	0.006
					-0.443 (-0.774, -0.113) [2H]	0.024	-0.547 (-0.909, -0.185) [2H]	0.002
	L-HDL_L	rs6811162	ENPEP	Self-reported hypertension, diagnosed high blood pressure	0.264 (0.062, 0.464)	0.014	0.301 (0.092, 0.510)	0.006
	L-HDL-C	rs5576825	LINC01621 ELOVL4	Waist circumference	0.279 (0.012, 0.544)	0.048	0.323 (0.0456, 0.602)	0.107
		rs6811162	ENPEP	Self-reported hypertension, diagnosed high blood pressure			0.241 (-0.037, 0.519)	0.026
		rs5576825 + rs6811162	-	-			0.287 (-0.004, 0.578)	0.063
	XL-HDL class	rs5576825	LINC01621 ELOVL4	Waist circumference	-0.285 (-0.552, -0.018)	0.015	-0.244 (-0.528, 0.040)	0.135
SA	LA	rs12720820	APOB	Self-reported high cholesterol, coronary artery disease, treatment with cholesterol lowering medication	0.477 (0.013, 0.939)	0.030	0.335 (-0.982, 0.763)	0.459
	FAw6	rs12720820	APOB	Self-reported high cholesterol, coronary artery disease, treatment with cholesterol lowering medication	0.445 (0.094, 0.794)	0.007	0.223 (-0.398, 0.843)	0.105
	L-HDL-PL	rs7486176	C12orf76	Systolic blood pressure, diagnosed high blood pressure, hypertension	0.692 (0.106, 1.28)	0.021	0.853 (0.170, 1.54)	0.013
	Fatty Acid class	rs12720820	APOB	Self-reported high cholesterol, coronary artery disease, treatment with cholesterol lowering medication	0.172 (0.018, 0.327)	0.018	0.142 (-0.049, 0.334)	0.119

Table 2: Removal of potentially pleiotropic SNPS.

2H: 2-hour post glucose. FAw6: Total n-6 fatty acids. FG: fasting glucose. LA: 18:2 linoleic acid (mmol/L). L-HDL-C: total cholesterols in L-HDL (mmol/L). L-HDL_L: total lipids in L-HDL (mmol/L). L-HDL-PL: phospholipids in L-HDL (mmol/L). SA: South Asian. WE: White European. WUH: Wu-Hausman p-value.

383 ## South Asians

384 No metabolite was associated with both fasting glucose and 2-hour post
385 glucose in SAs. Although, for fasting glucose, a 1 mmol/L increase in either total FAw3
386 or S-HDL-C was associated with an increase of fasting glucose by 0.432 mmol/L (95%
387 CI 0.063 – 0.798) and 1 mmol/L (95% CI 0.116 – 1.882) respectively. No metabolite
388 associated with a decrease in fasting glucose in SAs (**Supplementary Figures 4-5)**.

389 Nine metabolites associated with 2-hour post glucose levels in SAs. Of these,
390 4 metabolites, LA, FAw6, total lipids in M-VLDL (M-VLDL-L) and total phospholipids in
391 L-HDL (L-HDL-PL), associated with an increase in with 2-hour post glucose, with the
392 largest effect being identified for L-HDL-PL. Specifically, a 1mmol/L increase in L-HDL-
393 PL associated with a 0.692 mmol/L increase (95% CI 0.106 - 1.280) in 2-hour post
394 glucose. A further 5 additional metabolites were associated with a decrease in 2-hour
395 post glucose: concentration of L-LDL (L-LDL-P), total cholesterols in IDL (IDL-C),
396 cholesterol esters in IDL (IDL-CE) concentration, total cholesterols in IDL (IDL-C), total
397 lipids in small S-LDL (S-LDL-L), and total lipids in small S-HDL (S-HDL-L). The largest
398 decrease in 2-hour post glucose was observed for L-LDL-P where a 1mmol/L increase
399 in L-LDL-P associated with a 3.86 mmol/L decrease (95% 0.467 - 7.27) in 2-hour post
400 glucose levels **(Table 3)**.

401 Fatty acids were the class of metabolites most frequently associated with
402 postprandial glucose in SAs. All three fatty acids (LA, FAw3 and FAw6) associations
403 identified in SAs were of similar magnitude: a 1 mmol/l increase of FAw3 associated
404 with a +0.4 mmol/l increase in fasting glucose or and a 1 mmol/ increase od FAw6 and
405 LA associated with a +0.4 mmol/l increase of 2-hour post glucose.

406 No metabolite found to be associated with postprandial glucose measures in
407 WEs was found to be associated with postprandial glucose in SAs. However, in both

408 populations, members of the S-HDL and L-HDL class were found to be associated
409 with increased postprandial glucose.

410

411

412

413

414

	Class	Metabolite	Outcome	F statistic	β estimate (95% CI)	WuH
Lowers Glucose	L-LDL	LDL-P	2-hour post	11.7	-2238 (-4828 -193.7)	0.024
	S-LDL	SLDL-L	2-hour post	11.1	-3.86 (0.467, -7.27)	0.015
	IDL	IDL-C	2-hour post	10.9	-1.19 (-0.12, -2.27)	0.021
		IDL-CE	2-hour post	11.8	-1.34 (-0.144, -2.55)	0.023
	S-HDL	S-HDL-L	2-hour post	11.4	-1.23 (-0.137, -2.32)	0.012
Increase Glucose	Fatty Acids	LA	2-hour post	20.9	0.477 (0.013, 0.939)	0.030
		FAw3	Fasting glucose	10	0.432 (0.063, 0.798)	0.008
		FAw6	2-hour post	33.4	0.445 (0.094, 0.794)	0.007
	M-VLDL	M-VLDL-L	2-hour post	68.4	0.046 (0.009, 0.083)	0.008
	L-HDL	L-HDL-PL	2-hour post	43.3	0.692 (0.106, 1.28)	0.021
	S-HDL	S-HDL-C	Fasting glucose	22	1 (0.116,1.882)	0.012

415 416 417 418 419 420 421 422 423 424 425 426 427

428 **Table 3: Significant MR results in South Asians**

429 Glucose measures are expressed as mmol/L. FAw3: total n-3 fatty acids. FAw6:
430 total n-6 fatty acids. IDL-C: total cholesterols in LDL (mmol/L). IDL-CE:
431 cholesterol esters in LDL (mmol/L). LA: 18:2 Linoleic Acid (mmol/L). LDL_P:
432 concentration of LDL particles (mol/L). L-HDL-PL: phospholipids in L-HDL
433 (mmol/L). M-VLDL-L: total lipids in M-VLDL (mmol/L). S-HDL-C: total cholesterols
434 in S-HDL (mmol/L). S-HDL-L: total lipids in S-HDL (mmol/L). S-LDL-L: total lipids
435 in S-LDL (mmol/L). WuH: Wu-Hausman p-value.

436

437

438

South Asians: Sensitivity analyses

Six instruments in SAs were comprised of a single SNP meaning it was not possible to perform a leave-one-out analysis for these metabolites. For the remaining 5 metabolites, associations were consistent across each leave-one-out analyses **(Supplementary Figure 9**). Likewise, no large differences in F-statistics following the removal of individual SNPs were identified **(Supplementary Figure 10**).

Just as in WEs, 3 metabolites identified in SAs included SNPs associated with cholesterol or hypertension, which are potential confounders of the association between metabolites and dysglycemia **(Supplementary Table 6**). Significance was maintained following the removal of SNP rs7486176 (found within the *C12orf76* gene) from the total phospholipids in L-HDL instrument. For the LA and FAw6 exposures, the removal of SNP rs12720820 (found within the *APOB* gene) resulted in a non-significant association indicating that this SNP was the main driver of the identified association (**Table 2**). In leave-one-out analyses, the removal of SNP rs58865405 from the FAw6 instrument resulted in non-significance, although the biological role of this SNP remains unknown.

Post-hoc analysis: Analysis of metabolite classes

Numerous SNPs were found to be associated with more than one metabolite measure, particularly for metabolites in the same metabolite class (**Supplementary Figures 11-12**). This was anticipated since many metabolomic pathways are biologically intertwined. To minimise the risk of violation of the third MR assumption (that the genetic instrument must only influence the outcome via the exposure and not via an alternative biological pathway) [15], the collective effect of an entire class of metabolites on postprandial glucose measures was examined. A composite score for

464 each metabolite class was created by placing all metabolites in a single class (e.g. all
465 LDL metabolites), conducting a PCA and extracting PC1. This was only possible for
466 20 classes in WEs and 21 classes in SAs that had > 2 metabolites and ≥70% of the
467 class variation was explained by PC1 (**Supplementary Table 7**). To assess the impact
468 of outliers on PCA, outliers were defined and removed based on two cut-offs: standard
469 (1.5 X IQR from the median) and stringent (3 x IQR from the median). For all classes,
470 PC1 and PC2 scores were comparable after removal of both types of outliers so only
471 3xIQR outliers were removed prior to analyses (**Supplementary Table 8**).

472 138 SNPS remained after LD thinning in WEs, 87 (63.04%) of which were
473 unique to a single metabolite class exposure. 19/20 (95%) of the metabolite classes
474 examined in WEs had an F-statistic ≥10 (the only exception being the MHDL class).
475 54 SNPS remained after LD thinning in SAs, 42 (77.78%) of which were unique to a
476 single metabolite class. Despite the lower number of SNPs identified in SAs,17/20
477 (85%) of the metabolite classes examined had a F-statistic ≥10 indicating that
478 instrument strength was still sufficient, for all classes except for the non-branched
479 amino acids, LLDL class, and all VLDL classes. On average, the mean F-statistic of
480 metabolite class instruments was 19.83%. On average genetic instruments of
481 composite measures of the metabolite classes were weaker than the instruments for
482 individual metabolites in both WEs and SAs (**Supplementary Figure 13**).

483 In WEs, 4 metabolite classes were associated with a glucose measure: S-LDLs
484 associated with fasting and 2-hour post glucose; M-LDL and all LDLS (i.e. the
485 collective grouping of HDLs, MDLs and SDLs) were associated with fasting glucose,
486 and XL-HDLs were associated with 2-hour post glucose (**Table 4**). In SAs, the fatty
487 acid metabolite class were associated with 2-hour post glucose levels. No other
488 associations were identified in SAs.

As with the sensitivity analyses of individual metabolites, the removal of individual SNPs was not found to greatly impact the F-statistic of most instruments (**Supplementary Figures 14-15**). However, for the fatty acid metabolites class, the exclusion of rs12720820 or rs7159441, and for 'XL-HDL', the removal of rs55768285, resulted in non-significant associations, suggesting that these SNPSs are key drivers of the association. **(Table 4, Supplementary Table 9**).

Ethnicity	Metabolite class	Outcome	F statistic	β estimate (95% CI)	WUH
White European	XL-HDL	2-hour post glucose	96.2	-0.285 (-0.018, -0.552)	0.015
	M-LDL	Fasting glucose	95.5	-0.048 (-0.004, - 0.091)	0.024
	S-LDL	Fasting glucose	98	0.084 (0.007, 0.162)	0.024
		2-hour post glucose	98	-0.249 (-0.016, -0.482)	0.066
	All LDL	Fasting glucose	95	0.038 (-0.005, -0.068)	0.016
South Asian	Fatty Acids	2-hour post glucose	32.8	0.172 (0.018, 0.327)	0.0184

Table 4: Significant MR results from the analysis of metabolite classes

Glucose measures are expressed as mmol/L. WuH: Wu-Hausman p-value.

Power analysis

R^2 values were consistently lower in the ethnic group where an effect was not detected but all genetic instruments for the metabolite values had an F-statistic ≥ 10, indicating that weak instrument bias was not responsible for the absence of significant effects. Where an association was identified in one ethnic group but not another, to determine whether the absence of an association was potentially due a lack of power in the other ethnicity rather than an ethnic-specific effect, *post-hoc* power analyses were performed.

517 When using MR estimates as an estimate for the true causal effect both the
518 analyses of FAw3 and the overall fatty acid class in WEs were adequately powered to
519 detect the observed MR effect in both populations. Therefore, the absence of an effect
520 of FAw3 in WEs is unlikely due to inadequate power (**Supplementary Table 10**). The
521 analysis of HDL2C and HDL3C in SAs was also sufficiently powered to detect the
522 observed MR effect in WEs.

523 Discussion

524 This study has identified ethnically distinct associations between a range of
525 metabolites and postprandial glucose measures taken during pregnancy in SAs and
526 WEs, with notably no shared associations were identified. Fourteen metabolites were
527 found to be associated with postprandial glucose measures in WEs. Whereas, a
528 distinct set of 11 metabolites were associated in SAs. In WEs, cholesterols and
529 lipoproteins were the metabolite classes associated with postprandial glucose
530 measures, while in SAs fatty acids were the most commonly associated.

531 Furthermore, through an extensive GWAS of metabolites, this study identified
532 novel genome-wide significant associations in relation to acetate (1 SNP, rs10945476)
533 and tyrosine (15 SNPs, all on chromosome 17) in SAs. No previous associations have
534 been identified for SNP rs10945476, found within the non-coding transcript gene
535 *PRDM15* in relation to acetate or any other exposure.

536 Interestingly, 3 of the 15 SNPs associated with tyrosine are found in a
537 transmembrane transporter gene, *SLC13A2*. Moreover, an additional 10 of the newly
538 identified 15 SNPs associated with tyrosine were found in the *FOXN* gene, a
539 transcription factor that has previously been identified to be associated with ceramide
540 levels (a lipid metabolite) in a GWAS from a Chinese cohort [50]. Moreover, ceramide

541 has been shown to induce tyrosine phosphorylation in membrane proteins meaning it
542 is plausible that a gene associated with ceramide is also associated with tyrosine
543 levels in an Asian population [51]. Interestingly, ceramide has been proposed as a
544 mediator of the interaction between saturated fat and insulin resistance and has been
545 associated with T2D and cardiovascular disease [52]. To the best of our knowledge
546 tyrosine levels have not previously been associated with either *FOXN SLC13A2*. The
547 remaining 2 of the 15 SNPs identified as being associated with tyrosine in SAs, are
548 currently not in any known genes. All 15 SNPs identified as being associated with
549 tyrosine in SAs are in LD with each other in 1000G SA populations (all $R^2 \geq 0.38$).

550 Identified associations in WEs

551 Leucine

552 Branched chain amino acids (BCAAs), including leucine, are predominantly
553 metabolised in skeletal muscles where they regulate protein synthesis and
554 mitochondrial functions [53]. In addition, BCAAs are hormonal signalling regulators
555 and are expected to module insulin resistance (IR) through increasing insulin secretion
556 in human pancreatic β-cells [54, 55]. Our study found leucine to be negatively
557 associated with both fasting glucose and 2-hour post glucose levels during pregnancy
558 in WEs; with 1 mmol/L of leucine associated with a decrease of 0.193 mmol/L in fasting
559 glucose and 0.327 mmol/L 2-hour post glucose respectively. Although few studies
560 have investigated the role of leucine in glucose regulation during pregnancy,
561 interestingly the ratio of leucine/isoleucine was similarly found associated with
562 reductions in fasting glucose in the HAPO study, a multi-ethnic cohort of pregnant
563 women of Afro-Caribbean, Mexican American, Northern European, and Thai ancestry
564 [56]. Common dietary sources of leucine include meat products and cheese, with

565 smaller amounts also being present in other dairy products (such as dairy and
566 yoghurt), fish and in certain legumes and nuts, such as dried raw broad beans and
567 pine nuts [53]. Hence, dietary interventions aimed at increasing leucine levels during
568 pregnancy, possibly through a dietary intervention promoting the consumption of lean
569 animal protein, low-fat dairy and nuts, may help improve pregnancy hyperglycaemia
570 in WEs.
571

572 **Cholesterols**

573 HDL cholesterol is colloquially described as 'good cholesterol' due to its role in
574 the removal of cholesterols from atherosclerotic plaque, thereby reducing an
575 individual's risk of CVD [57]. Furthermore, low HDL levels have commonly been
576 associated with diabetes in humans, with HDL shown to increase insulin secretion and
577 β-cell survival [58, 59]. We identified four associations between HDL cholesterol and
578 postprandial glucose measures in WEs. Herein, 1 mmol/L increase in S-HDL-CE
579 confers a 1.78 mmol/L increase (95% CI 0.49 – 3.11) in 2-hour post glucose. This is
580 consistent with previous evidence from a Finnish sample of overweight and obese
581 women where cholesterol esters in S-HDL were higher in the serum samples of GDM
582 cases at ~14 weeks gestation [60]. Discrepancies in the direct effect of HDL
583 cholesterol on dysglycemia have also been identified in the genetic literature [59], with
584 a recent review highlighting that while a genetic study utilising linear relation analysis
585 did find HDLs to have a protective effects against T2D (n cases = 2,447) [61], the same
586 effect was not been replicated in an MR setting (n cases= 47,627) [62]. When
587 considering LDL cholesterols, only S-LDLs was found to be significantly associated
588 with a postprandial glucose measure (fasting glucose) in WEs. Additionally, in our
589 composite analysis of metabolite classes, S-LDLs were associated with fasting

590 glucose and 2-hour post glucose in WEs, whereas the M-LDL and all LDL (a combined
591 measure of S-LDLs, M-LDLs and HDLs) classes were associated with fasting glucose.
592 Unfortunately, because composite scores were comprised of PC1 coordinates the
593 direction of effect of these associations could not be evaluated. To our knowledge, no
594 previous study has conducted an MR of metabolites on dysglycemic predictors of
595 GDM.
596

597 **Triglycerides**

598 Triglycerides are an abundant class of lipid particles found in the blood,
599 originating from either from the consumption of dietary fats or as a result of hepatic
600 metabolism [63, 64]. Once in the blood, triglycerides can be incorporated into HDL
601 and LDL cholesterol particles. In addition to dietary triglyceride consumption, dietary
602 fatty acids can be converted into triglycerides before they enter circulation, highlighting
603 the complex relationship between triglyceride, cholesterol, and fatty acid levels [63].
604 Our results suggest triglycerides in XSVLDL (XS-VLDL-TG) associate with
605 increased 2-hour post glucose (0.841 mmol/L) in WEs. In agreement with these
606 findings, increased triglycerides in XSVLDL levels have also previously been
607 associated with increased likelihood of GDM in a Finnish population [60]. No other
608 triglyceride was found to be associated with in WEs. One explanation no additional
609 associations were detected could be due to the average BMI of the WEs in BiB. For
610 example, an analysis of a prospective Irish cohort (~94% WE) found that triglyceride
611 levels were only associated with GDM in obese individuals, a higher average BMI than
612 that observed in the BiB cohort [65]. Further confirmation of these findings of increased
613 triglycerides in XS-VLDLs would suggest that this association is, at least in part,

614 responsible for the identified associations between diets high in fats and increased
615 prevalence of GDM in WEs [11].
616

617 Identified associations in SAs

618 Fatty acids

619 Polyunsaturated acids (PUFAs) are consumed in the diet and can be converted
620 to long-chain PUFAs (LC-PUFAs) through a process of desaturation and elongation
621 reactions that predominately occur in the liver [66]. Changes in dietary patterns can
622 have a large impact on fatty acid composition in the body and with-it disease risk. For
623 example, a western dietary pattern, which has high levels of n-6 fatty acids, has been
624 associated with GDM risk [11, 67]. In a cohort of Chinese adults, total n-6 fatty acids
625 and 18:2 n-6 levels at baseline in venous blood samples were both found to associate
626 with an increased risk of T2D after ~8 years of follow up, while increased n-3 fatty acid
627 levels were protective [68]. However, a recent two-sample MR suggested only a
628 negligible effect of n-6 PUFA synthesis on T2D in a predominantly WE cohort [69].
629 Moreover, the relationship between n-3, n-6, n-9 fatty acids and GDM remains
630 inconclusive, and in a recent (2021) systematic review none of the identified studies
631 (n=15) was conducted in a SA population [67], highlighting the need for more studies
632 exploring the role of fatty acids in GDM development in Asians [70].
633 This study provides evidence of an association between LA and total FAw6
634 levels and an increase in 2-hour post glucose levels during pregnancy in SAs. In
635 addition, the fatty acid class associated with 2-hour post glucose in SAs. Through a
636 leave-out-one sensitivity analyses for the FAw6 and LA instruments, the removal of
637 the SNP rs12720820 (found within the *APOB* gene) resulted in non-significant

638 associations for both exposures, indicating that this rs12720820 was the largest
639 contributor to the identified associations and has previously been associated with
640 cholesterol levels and the use of cholesterol lowering drugs [71]. Interestingly, FAw3
641 also associated with increased 2-hour post glucose in SAs; however, with only one
642 SNP potential pleiotropy could not be explored.

643 It is well established that fatty acid profiles can impact blood cholesterol levels
644 [72-74]. In addition, increased dietary cholesterol has previously been associated with
645 an increased risk of GDM in a systematic review of observational studies [75]. Taken
646 together, our data confirm that fatty acids and cholesterol metabolites are in vertically
647 pleiotropy and are likely impacting gestational dysglycemia via the same causal
648 pathway. Unlike horizontal pleiotropy, vertical pleiotropy does not result in a violation
649 of the 2nd MR assumption as cholesterol is not acting as a confounder, meaning MR
650 estimates are still valid (**Figure 1**). Furthermore, it is also possible that this interaction
651 between fatty acids and cholesterols may be ethnic-specific due to the absence of
652 associations identified between fatty acids and postprandial glucose measures in
653 WEs. In addition to possible variations in cholesterol metabolism, it is plausible that
654 variations in fatty acid synthesis are also partially responsible for the increased GDM
655 risk experienced by SAs. For example, variants within the *FADS* genes impact LC-
656 PUFA conversion [76, 77]. Current evidence suggests that SAs are likely to synthesise
657 LC-PUFA more quickly than WEs, which could contribute to elevated risk of prolonged
658 exposure to elevated LC-PUFA levels (namely, w6) and risk of dysglycemia [76, 77].
659 If these ethnic differences in fatty acid metabolism are confirmed to be linked to
660 disease risk, it would aid in the development of tailored GDM prevention strategies
661 that focus on modifying fatty acid profiles in an ethnic-specific manner.

The analyses found no association between triglycerides and dysglycemia in SAs. This agrees with a recent meta-analysis that concluded, although triglyceride levels associated with likelihood of GDM ($I^2 \geq 84\%$) [78], after stratification by culture/geographical location, they found no association between triglyceride levels and likelihood of GDM. The reasons for this are unclear but it has also been shown that SAs have a higher prevalence of hypertriglyceridemia than WEs and at lower BMI levels, meaning it is possible that the difference in triglyceride levels and in SA GDM cases and controls is less pronounced than in WEs [79].

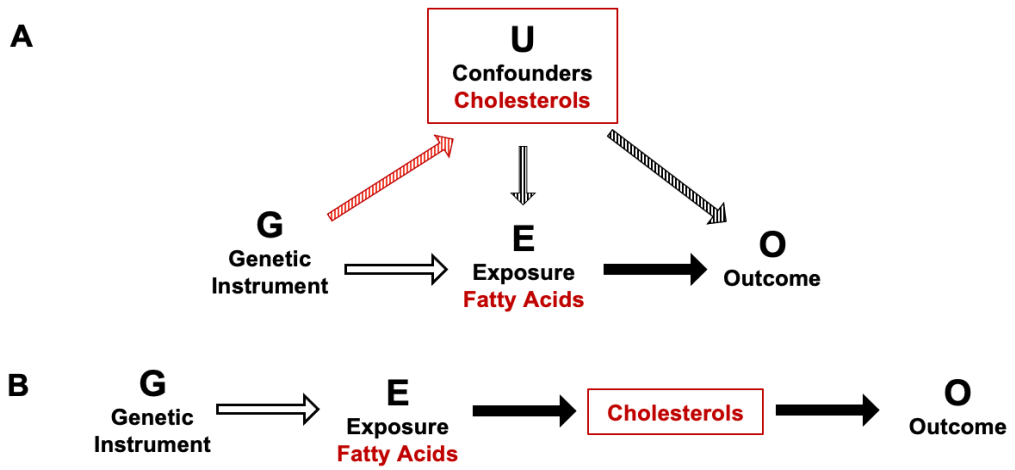


Figure 1: Schematic of potential horizontal and vertical pleiotropy in relation to fatty acid and cholesterol metabolites and postprandial glucose measures.

A: Illustration of horizontal pleiotropy. **B:** Illustration of vertical pleiotropy. Vertical pleiotropy does not result in a violation of the 2nd MR assumption.

Strengths and limitations

This analysis has several strengths. Firstly, this study involved a large and comprehensive panel of metabolites allowing for the relationships between

682 metabolites and postprandial glucose to be thoroughly investigated. Secondly, this is
683 the first MR study to investigate dysglycemia during pregnancy while also being one
684 of the few MR studies to be conducted in a SA population. Finally, through leave-one-
685 out analyses and the searching of both Phenoscanner and GWAS Catalog databases,
686 violations of the 2nd and 3rd MR assumptions were thoroughly investigated meaning
687 that it was possible to conclude that identified associations may not be subjected to
688 horizontal pleiotropy and that identified causal associations are valid due to the
689 absence of detectable violations of the MR assumptions.

690 Nonetheless, this study has some limitations. Metabolites are highly correlated
691 meaning it is not possible to confidently interpret that an individual metabolite is
692 independently associated with a postprandial outcome measure. To account for this
693 limitation MR analyses were performed on composite measures of each metabolite
694 class (when PC1 explained ≥70% of the variation in the metabolite class) in order to
695 assess the overall impact of each metabolite class on pregnancy dysglycemia.
696 Furthermore, MR also assumes the level of genetically conferred exposure from
697 conception to the time of measurement is constant, which is unknown when studying
698 metabolites. The limited sample size of this study also means that further adjustment
699 could not be made at the GWAS stage since missing data in certain variables would
700 further reduce sample size (eg, age) and some associations may have been
701 underpowered to detect an effect. However, a *post hoc* power analyses found that for
702 some metabolite values significant effects only identified in one ethnicity were possible
703 to detect in the alternate ethnicity. In addition, some genetic instruments included only
704 one SNP meaning it was not possible to evaluate the impact of pleiotropy for any
705 identified associations involving these instruments. Finally, due to limitations in data

availability in SAs a two-sample MR could not be conducted meaning it was not possible to assess the generalisability of these findings.

Conclusions

The presence of causal relationships between a comprehensive set of metabolites and postprandial glucose measures (fasting glucose and 2-hour post glucose) in mid-pregnancy has been established in a UK SA and WE population. This study has found a range of metabolite values to be associated with postprandial glucose measures in WEs and high-risk SA women, although more associations were identified in WEs despite these individuals being at lower risk of GDM. In high-risk SA women, total n-6 fatty acids and the n-6 fatty acid, LA appear to increase postprandial glucose levels suggesting that fatty acids may be responsible for a large proportion of metabolically driven risk for GDM experienced by this population. Future work in a larger sample (potentially using a two-sample MR) and a larger panel of metabolites is needed to investigate our findings and hypotheses more closely, ideally over the course of a pregnancy in order to aid in GDM prevention in this high-risk population.

Acknowledgements

Born in Bradford is only possible because of the enthusiasm and commitment of the children and parents in BiB. We are grateful to all the participants, health professionals, schools and researchers who have made Born in Bradford happen. This work was undertaken on ARC4, part of the High Performance Computing facilities at the University of Leeds, UK.

References

1. Plows JF, Stanley JL, Baker PN, Reynolds CM, Vickers MH. The Pathophysiology of Gestational Diabetes Mellitus. Int J Mol Sci. 2018;19(11). Epub 2018/10/31. doi: 10.3390/ijms19113342. PubMed PMID: 30373146; PubMed Central PMCID: PMCPMC6274679.
2. Taylor K, Ferreira DLS, West J, Yang T, Caputo M, Lawlor DA. Differences in Pregnancy Metabolic Profiles and Their Determinants between White European and South Asian Women: Findings from the Born in Bradford Cohort. Metabolites. 2019;9(9). Epub 2019/09/22. doi: 10.3390/metabo9090190. PubMed PMID: 31540515; PubMed Central PMCID: PMCPMC6780545.
3. Mills HL, Patel N, White SL, Pasupathy D, Briley AL, Santos Ferreira DL, et al. The effect of a lifestyle intervention in obese pregnant women on gestational metabolic profiles: findings from the UK Pregnancies Better Eating and Activity Trial (UPBEAT) randomised controlled trial. BMC Med. 2019;17(1):15. Epub 2019/01/22. doi: 10.1186/s12916-018-1248-7. PubMed PMID: 30661507; PubMed Central PMCID: PMCPMC6340185.
4. Kampmann U, Knorr S, Fuglsang J, Ovesen P. Determinants of Maternal Insulin Resistance during Pregnancy: An Updated Overview. Journal of Diabetes Research. 2019;2019:5320156. doi: 10.1155/2019/5320156.
5. McIntyre HD, Catalano P, Zhang C, Desoye G, Mathiesen ER, Damm P. Gestational diabetes mellitus. Nat Rev Dis Primers. 2019;5(1):47. Epub 2019/07/13. doi: 10.1038/s41572-019-0098-8. PubMed PMID: 31296866.
6. Rodrigo N, Glastras SJ. The Emerging Role of Biomarkers in the Diagnosis of Gestational Diabetes Mellitus. J Clin Med. 2018;7(6). Epub 2018/06/09. doi: 10.3390/jcm7060120. PubMed PMID: 29882903; PubMed Central PMCID: PMCPMC6024961.
7. Menard V, Sotunde OF, Weiler HA. Ethnicity and Immigration Status as Risk Factors for Gestational Diabetes Mellitus, Anemia and Pregnancy Outcomes Among Food Insecure Women Attending the Montreal Diet Dispensary Program. Can J Diabetes. 2020;44(2):139-45 e1. Epub 2019/08/21. doi: 10.1016/j.jcjd.2019.05.004. PubMed PMID: 31427254.
8. Gadve SS, Chavanda S, Mukherjee AD, Aziz S, Joshi A, Patwardhan M. Risk of Developing Type 2 Diabetes Mellitus in South Asian Women with History of Gestational Diabetes Mellitus: A Systematic Review and Meta-Analysis. Indian J Endocrinol Metab. 2021;25(3):176-81. Epub 2021/11/12. doi: 10.4103/ijem.IJEM_57_21. PubMed PMID: 34760669; PubMed Central PMCID: PMCPMC8547406.
9. Fuller H, Iles M, Moore JB, Zulyniak MA. Unique Metabolic Profiles Associate with Gestational Diabetes and Ethnicity in Low and High-Risk Women Living in the UK. J Nutr. 2022. Epub 20220726. doi: 10.1093/jn/nxac163. PubMed PMID: 35883228.
10. Taylor K, Ferreira DLS, West J, Yang T, Caputo M, Lawlor DA. Differences in Pregnancy Metabolic Profiles and Their Determinants between White European and South Asian Women: Findings from the Born in Bradford Cohort. Metabolites. 2019;9(9):190. doi: 10.3390/metabo9090190. PubMed PMID: 31540515.
11. Fuller H, Moore JB, Iles MM, Zulyniak MA. Ethnic-specific associations between dietary consumption and gestational diabetes mellitus incidence: A meta-analysis. PLOS Global Public Health. 2022;2(5):e0000250. doi: 10.1371/journal.pgph.0000250. PubMed PMID: pub.1147817197.

12. Griffith RJ, Alsweiler J, Moore AE, Brown S, Middleton P, Shepherd E, et al. Interventions to prevent women from developing gestational diabetes mellitus: an overview of Cochrane Reviews. Cochrane Database Syst Rev. 2020;6:CD012394. Epub 2020/06/12. doi: 10.1002/14651858.CD012394.pub3. PubMed PMID: 32526091; PubMed Central PMCID: PMCPMC7388385.
13. de Souza RJ, Shanmuganathan M, Lamri A, Atkinson SA, Becker A, Desai D, et al. Maternal Diet and the Serum Metabolome in Pregnancy: Robust Dietary Biomarkers Generalizable to a Multiethnic Birth Cohort. Curr Dev Nutr. 2020;4(10):nzaa144-nzaa. doi: 10.1093/cdn/nzaa144. PubMed PMID: 33073162.
14. Wang QY, You LH, Xiang LL, Zhu YT, Zeng Y. Current progress in metabolomics of gestational diabetes mellitus. World J Diabetes. 2021;12(8):1164-86. Epub 2021/09/14. doi: 10.4239/wjd.v12.i8.1164. PubMed PMID: 34512885; PubMed Central PMCID: PMCPMC8394228.
15. Sheehan NA, Meng S, Didelez V. Mendelian Randomisation: A Tool for Assessing Causality in Observational Epidemiology. In: Teare MD, editor. Genetic Epidemiology. Totowa, NJ: Humana Press; 2011. p. 153-66.
16. Liu J, van Klinken JB, Semiz S, van Dijk KW, Verhoeven A, Hankemeier T, et al. A Mendelian Randomization Study of Metabolite Profiles, Fasting Glucose, and Type 2 Diabetes. Diabetes. 2017;66(11):2915-26. Epub 2017/08/30. doi: 10.2337/db17-0199. PubMed PMID: 28847883.
17. Carreras-Torres R, Johansson M, Gaborieau V, Haycock PC, Wade KH, Relton CL, et al. The Role of Obesity, Type 2 Diabetes, and Metabolic Factors in Pancreatic Cancer: A Mendelian Randomization Study. J Natl Cancer Inst. 2017;109(9). Epub 2017/09/28. doi: 10.1093/jnci/djx012. PubMed PMID: 28954281; PubMed Central PMCID: PMCPMC5721813.
18. Wright J, Small N, Raynor P, Tuffnell D, Bhopal R, Cameron N, et al. Cohort Profile: the Born in Bradford multi-ethnic family cohort study. Int J Epidemiol. 2013;42(4):978-91. Epub 20121012. doi: 10.1093/ije/dys112. PubMed PMID: 23064411.
19. Bird PK, McEachan RRC, Mon-Williams M, Small N, West J, Whincup P, et al. Growing up in Bradford: protocol for the age 7-11 follow up of the Born in Bradford birth cohort. BMC Public Health. 2019;19(1):939. Epub 20190712. doi: 10.1186/s12889-019-7222-2. PubMed PMID: 31300003; PubMed Central PMCID: PMCPMC6626420.
20. Lawlor DA, West J, Fairley L, Nelson SM, Bhopal RS, Tuffnell D, et al. Pregnancy glycaemia and cord-blood levels of insulin and leptin in Pakistani and white British mother-offspring pairs: findings from a prospective pregnancy cohort. Diabetologia. 2014;57(12):2492-500. Epub 2014/10/03. doi: 10.1007/s00125-014-3386-6. PubMed PMID: 25273345; PubMed Central PMCID: PMCPMC4218974.
21. Josse J, Husson F. missMDA: A Package for Handling Missing Values in Multivariate Data Analysis. Journal of Statistical Software. 2016;70(1). doi: 10.18637/jss.v070.i01.
22. Arciero E, Dogra SA, Malawsky DS, Mezzavilla M, Tsismentzoglou T, Huang QQ, et al. Fine-scale population structure and demographic history of British Pakistanis. Nat Commun. 2021;12(1):7189. Epub 2021/12/12. doi: 10.1038/s41467-021-27394-2. PubMed PMID: 34893604; PubMed Central PMCID: PMCPMC8664933.
23. Marees AT, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, et al. A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. Int J Methods Psychiatr Res. 2018;27(2):e1608. Epub 2018/02/28. doi: 10.1002/mpr.1608. PubMed PMID: 29484742; PubMed Central PMCID: PMCPMC6001694.

24. Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, et al. Genome-wide association studies. Nature Reviews Methods Primers. 2021;1(1). doi: 10.1038/s43586-021-00056-9.
25. Sheridan E, Wright J, Small N, Corry PC, Oddie S, Whibley C, et al. Risk factors for congenital anomaly in a multiethnic birth cohort: an analysis of the Born in Bradford study. The Lancet. 2013;382(9901):1350-9. doi: https://doi.org/10.1016/S0140-6736(13)61132-0.
26. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. Advantages and pitfalls in the application of mixed-model association methods. Nat Genet. 2014;46(2):100-6. Epub 2014/01/30. doi: 10.1038/ng.2876. PubMed PMID: 24473328; PubMed Central PMCID: PMCPMC3989144.
27. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559-75. Epub 20070725. doi: 10.1086/519795. PubMed PMID: 17701901; PubMed Central PMCID: PMCPMC1950838.
28. Purcell S. PLINK (Version 1.9).
29. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88(1):76-82. Epub 2010/12/21. doi: 10.1016/j.ajhg.2010.11.011. PubMed PMID: 21167468; PubMed Central PMCID: PMCPMC3014363.
30. van den Berg S, Vandenplas J, van Eeuwijk FA, Lopes MS, Veerkamp RF. Significance testing and genomic inflation factor using high-density genotypes or whole-genome sequence data. J Anim Breed Genet. 2019;136(6):418-29. Epub 2019/06/20. doi: 10.1111/jbg.12419. PubMed PMID: 31215703; PubMed Central PMCID: PMCPMC6900143.
31. Rivadeneira F, Uitterlinden AG. Chapter 18 - Genetics of osteoporosis. In: Dempster DW, Cauley JA, Bouxsein ML, Cosman F, editors. Marcus and Feldman's Osteoporosis (Fifth Edition): Academic Press; 2021. p. 405-51.
32. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26(17):2190-1. Epub 2010/07/10. doi: 10.1093/bioinformatics/btq340. PubMed PMID: 20616382; PubMed Central PMCID: PMCPMC2922887.
33. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics. 2015;31(21):3555-7. Epub 2015/07/04. doi: 10.1093/bioinformatics/btv402. PubMed PMID: 26139635; PubMed Central PMCID: PMCPMC4626747.
34. Myers TA, Chanock SJ, Machiela MJ. LDlinkR: An R Package for Rapidly Calculating Linkage Disequilibrium Statistics in Diverse Populations. Front Genet. 2020;11:157. Epub 2020/03/18. doi: 10.3389/fgene.2020.00157. PubMed PMID: 32180801; PubMed Central PMCID: PMCPMC7059597.
35. Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015;526(7571):68-74. Epub 2015/10/04. doi: 10.1038/nature15393. PubMed PMID: 26432245; PubMed Central PMCID: PMCPMC4750478.
36. R Development Core Team. R: A language and environment for statistical computing Vienna, Austria: R Foundation for Statistical Computing; 2010.
37. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2016.

865 38. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK:
866 rising to the challenge of larger and richer datasets. GigaScience. 2015;4(1). doi:
867 10.1186/s13742-015-0047-8.
868 39. Budu-Aggrey A, Brumpton B, Tyrrell J, Watkins S, Modalsli EH, Celis-Morales C, et al.
869 Evidence of a causal relationship between body mass index and psoriasis: A mendelian
870 randomization study. PLoS Med. 2019;16(1):e1002739. Epub 2019/02/01. doi:
871 10.1371/journal.pmed.1002739. PubMed PMID: 30703100.
872 40. Fox J, Kleiber, C., Zeileis, A. _ivreg: Instrumental-Variables Regression by '2SLS', '2SM',
873 or '2SMM', with Diagnostics_. . R package version 0.6-1 ed2021.
874 41. Skrivankova VW, Richmond RC, Woolf BAR, Davies NM, Swanson SA, VanderWeele TJ,
875 et al. Strengthening the reporting of observational studies in epidemiology using mendelian
876 randomisation (STROBE-MR): explanation and elaboration. BMJ. 2021;375:n2233. Epub
877 2021/10/28. doi: 10.1136/bmj.n2233. PubMed PMID: 34702754.
878 42. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The
879 NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and
880 summary statistics 2019. Nucleic Acids Research. 2019;47.
881 43. Staley JR, Blackshaw J, Kamat MA, Ellis S, Surendran P, Sun BB, et al. PhenoScanner: a
882 database of human genotype-phenotype associations. Bioinformatics. 2016;32(20):3207-9.
883 Epub 2016/06/19. doi: 10.1093/bioinformatics/btw373. PubMed PMID: 27318201; PubMed
884 Central PMCID: PMCPMC5048068.
885 44. Kamat MA, Blackshaw JA, Young R, Surendran P, Burgess S, Danesh J, et al.
886 PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations.
887 Bioinformatics. 2019;35(22):4851-3. Epub 2019/06/25. doi: 10.1093/bioinformatics/btz469.
888 PubMed PMID: 31233103; PubMed Central PMCID: PMCPMC6853652.
889 45. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a
890 guide, glossary, and checklist for clinicians. BMJ. 2018;362:k601. Epub 2018/07/12. doi:
891 10.1136/bmj.k601. PubMed PMID: 30002074; PubMed Central PMCID: PMCPMC6041728.
892 46. Brion M-JA, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian
893 randomization studies. International journal of epidemiology. 2012;42(5):1497-501.
894 47. Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, et al. Ensembl
895 2021. Nucleic Acids Research. 2020;49(D1):D884-D91. doi: 10.1093/nar/gkaa942.
896 48. Ensembl. Gene: SORCS1 e! Ensembl: Ensembl release 105; 2021 [cited 2022
897 01/04/2022]. Available from:
898 https://www.ensembl.org/Homo_sapiens/Gene/Summary?g=ENSG00000108018;r=10:1065
899 73663-107164706.
900 49. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian
901 randomization: using genes as instruments for making causal inferences in epidemiology. Stat
902 Med. 2008;27(8):1133-63. doi: 10.1002/sim.3034. PubMed PMID: 17886233.
903 50. Chai JF, Raichur S, Khor IW, Torta F, Chew WS, Herr DR, et al. Associations with
904 metabolites in Chinese suggest new metabolic roles in Alzheimer's and Parkinson's diseases.
905 Hum Mol Genet. 2020;29(2):189-201. Epub 2019/10/20. doi: 10.1093/hmg/ddz246. PubMed
906 PMID: 31628463.
907 51. Gulbins E, Szabo I, Baltzer K, Lang F. Ceramide-induced inhibition of T lymphocyte
908 voltage-gated potassium channel is mediated by tyrosine kinases. Proc Natl Acad Sci U S A.
909 1997;94(14):7661-6. Epub 1997/07/08. doi: 10.1073/pnas.94.14.7661. PubMed PMID:
910 9207149; PubMed Central PMCID: PMCPMC23879.

52. Yki-Jarvinen H, Luukkonen PK, Hodson L, Moore JB. Dietary carbohydrates and fats in nonalcoholic fatty liver disease. Nat Rev Gastroenterol Hepatol. 2021;18(11):770-86. Epub 20210713. doi: 10.1038/s41575-021-00472-y. PubMed PMID: 34257427.
53. Rondanelli M, Nichetti M, Peroni G, Faliva MA, Naso M, Gasparri C, et al. Where to Find Leucine in Food and How to Feed Elderly With Sarcopenia in Order to Counteract Loss of Muscle Mass: Practical Advice. Front Nutr. 2021;7:622391-. doi: 10.3389/fnut.2020.622391. PubMed PMID: 33585538.
54. Nie C, He T, Zhang W, Zhang G, Ma X. Branched Chain Amino Acids: Beyond Nutrition Metabolism. Int J Mol Sci. 2018;19(4). Epub 2018/03/24. doi: 10.3390/ijms19040954. PubMed PMID: 29570613; PubMed Central PMCID: PMCPMC5979320.
55. Leenders M, van Loon LJ. Leucine as a pharmaconutrient to prevent and treat sarcopenia and type 2 diabetes. Nutrition Reviews. 2011;69(11):675-89. doi: 10.1111/j.1753-4887.2011.00443.x.
56. Liu Y, Kuang A, Talbot O, Bain JR, Muehlbauer MJ, Hayes MG, et al. Metabolomic and genetic associations with insulin resistance in pregnancy. Diabetologia. 2020;63(9):1783-95. Epub 2020/06/20. doi: 10.1007/s00125-020-05198-1. PubMed PMID: 32556615; PubMed Central PMCID: PMCPMC7416451.
57. CDC. LDL and HDL Cholesterol: "Bad" and "Good" Cholesterol: US.gov; 2020 [cited 2022 01/04/2022]. Available from: https://www.cdc.gov/cholesterol/ldl_hdl.htm.
58. Ryckman KK, Spracklen CN, Smith CJ, Robinson JG, Saftlas AF. Maternal lipid levels during pregnancy and gestational diabetes: a systematic review and meta-analysis. BJOG. 2015;122(5):643-51. Epub 2015/01/23. doi: 10.1111/1471-0528.13261. PubMed PMID: 25612005.
59. Wong NKP, Nicholls SJ, Tan JTM, Bursill CA. The Role of High-Density Lipoproteins in Diabetes and Its Vascular Complications. International journal of molecular sciences. 2018;19(6):1680. doi: 10.3390/ijms19061680. PubMed PMID: 29874886.
60. Mokkala K, Vahlberg T, Pellonperä O, Houttu N, Koivuniemi E, Laitinen K. Distinct Metabolic Profile in Early Pregnancy of Overweight and Obese Women Developing Gestational Diabetes. J Nutr. 2020;150(1):31-7. Epub 2019/09/19. doi: 10.1093/jn/nxz220. PubMed PMID: 31529056.
61. Qi Q, Liang L, Doria A, Hu FB, Qi L. Genetic predisposition to dyslipidemia and type 2 diabetes risk in two prospective cohorts. Diabetes. 2012;61(3):745-52. Epub 20120207. doi: 10.2337/db11-1254. PubMed PMID: 22315312; PubMed Central PMCID: PMCPMC3282815.
62. Haase CL, Tybjærg-Hansen A, Nordestgaard BG, Frikke-Schmidt R. HDL Cholesterol and Risk of Type 2 Diabetes: A Mendelian Randomization Study. Diabetes. 2015;64(9):3328-33. Epub 20150513. doi: 10.2337/db14-1603. PubMed PMID: 25972569.
63. Alexopoulos A-S, Qamar A, Hutchins K, Crowley MJ, Batch BC, Guyton JR. Triglycerides: Emerging Targets in Diabetes Care? Review of Moderate Hypertriglyceridemia in Diabetes. Curr Diab Rep. 2019;19(4):13-. doi: 10.1007/s11892-019-1136-3. PubMed PMID: 30806837.
64. Laufs U, Parhofer KG, Ginsberg HN, Hegele RA. Clinical review on triglycerides. Eur Heart J. 2020;41(1):99-109c. doi: 10.1093/eurheartj/ehz785. PubMed PMID: 31764986.
65. O'Malley EG, Reynolds CME, Killalea A, O'Kelly R, Sheehan SR, Turner MJ. Maternal obesity and dyslipidemia associated with gestational diabetes mellitus (GDM). Eur J Obstet Gynecol Reprod Biol. 2020;246:67-71. Epub 2020/01/22. doi: 10.1016/j.ejogrb.2020.01.007. PubMed PMID: 31962258.
66. Gonzalez-Soto M, Mutch DM. Diet Regulation of Long-Chain PUFA Synthesis: Role of Macronutrients, Micronutrients, and Polyphenols on Δ-5/Δ-6 Desaturases and Elongases 2/5.

958 Adv Nutr. 2021;12(3):980-94. doi: 10.1093/advances/nmaa142. PubMed PMID: 33186986;
959 PubMed Central PMCID: PMCPMC8166571.
960 67. Hosseinkhani S, Dehghanbanadaki H, Aazami H, Pasalar P, Asadi M, Razi F. Association
961 of circulating omega 3, 6 and 9 fatty acids with gestational diabetes mellitus: a systematic
962 review. BMC Endocrine Disorders. 2021;21(1):120. doi: 10.1186/s12902-021-00783-w.
963 68. Bragg F, Kartsonaki C, Guo Y, Holmes M, Du H, Yu C, et al. Circulating Metabolites and
964 the Development of Type 2 Diabetes in Chinese Adults. Diabetes Care. 2021;45(2):477-80.
965 doi: 10.2337/dc21-1415.
966 69. Zulyniak MA, Fuller H, Iles MM. Investigation of the Causal Association between Long-
967 Chain n-6 Polyunsaturated Fatty Acid Synthesis and the Risk of Type 2 Diabetes: A Mendelian
968 Randomization Analysis. Lifestyle Genomics. 2020;13(5):146-53. doi: 10.1159/000509663.
969 70. Pan X-F, Huang Y, Li X, Wang Y, Ye Y, Chen H, et al. Circulating fatty acids and risk of
970 gestational diabetes mellitus: prospective analyses in China. European Journal of
971 Endocrinology. 2021;185(1):87-97. doi: 10.1530/eje-21-0118.
972 71. MedlinePlus. APOB gene National Library of Medicine (US): NIH; 2021 [cited 2022
973 01/04/2022]. Available from: https://medlineplus.gov/genetics/gene/apob/.
974 72. Mensink RP, Zock PL, Kester AD, Katan MB. Effects of dietary fatty acids and
975 carbohydrates on the ratio of serum total to HDL cholesterol and on serum lipids and
976 apolipoproteins: a meta-analysis of 60 controlled trials. The American Journal of Clinical
977 Nutrition. 2003;77(5):1146-55. doi: 10.1093/ajcn/77.5.1146.
978 73. Jacobson TA, Glickstein SB, Rowe JD, Soni PN. Effects of eicosapentaenoic acid and
979 docosahexaenoic acid on low-density lipoprotein cholesterol and other lipids: a review. J Clin
980 Lipidol. 2012;6(1):5-18. Epub 2012/01/24. doi: 10.1016/j.jacl.2011.10.018. PubMed PMID:
981 22264569.
982 74. Ooi EM, Ng TW, Watts GF, Barrett PH. Dietary fatty acids and lipoprotein metabolism:
983 new insights and updates. Curr Opin Lipidol. 2013;24(3):192-7. Epub 2013/04/27. doi:
984 10.1097/MOL.0b013e3283613ba2. PubMed PMID: 23619368.
985 75. Schoenaker DA, Mishra GD, Callaway LK, Soedamah-Muthu SS. The Role of Energy,
986 Nutrients, Foods, and Dietary Patterns in the Development of Gestational Diabetes Mellitus:
987 A Systematic Review of Observational Studies. Diabetes Care. 2016;39(1):16-23. Epub
988 2015/12/24. doi: 10.2337/dc15-0540. PubMed PMID: 26696657.
989 76. Koletzko B, Reischl E, Tanjung C, Gonzalez-Casanova I, Ramakrishnan U, Meldrum S, et
990 al. FADS1 and FADS2 Polymorphisms Modulate Fatty Acid Metabolism and Dietary Impact on
991 Health. Annu Rev Nutr. 2019;39:21-44. Epub 2019/08/23. doi: 10.1146/annurev-nutr-
992 082018-124250. PubMed PMID: 31433740.
993 77. Ameur A, Enroth S, Johansson A, Zaboli G, Igl W, Johansson AC, et al. Genetic
994 adaptation of fatty-acid metabolism: a human-specific haplotype increasing the biosynthesis
995 of long-chain omega-3 and omega-6 fatty acids. Am J Hum Genet. 2012;90(5):809-20. Epub
996 2012/04/17. doi: 10.1016/j.ajhg.2012.03.014. PubMed PMID: 22503634; PubMed Central
997 PMCID: PMCPMC3376635.
998 78. Hu J, Gillies CL, Lin S, Stewart ZA, Melford SE, Abrams KR, et al. Association of maternal
999 lipid profile and gestational diabetes mellitus: A systematic review and meta-analysis of 292
1000 studies and 97,880 women. EClinicalMedicine. 2021;34:100830. Epub 2021/05/18. doi:
1001 10.1016/j.eclinm.2021.100830. PubMed PMID: 33997732; PubMed Central PMCID:
1002 PMCPMC8102708.

79. Makshood M, Post WS, Kanaya AM. Lipids in South Asians: Epidemiology and Management. Curr Cardiovasc Risk Rep. 2019;13(8):24. Epub 2019/07/11. doi: 10.1007/s12170-019-0618-9. PubMed PMID: 33833849.

Supplementary Figure 1: PCA to account for population stratification.

A: Schematic of data processing steps taken prior to PCA of genetic data. R^2 cut-off of 0.3 and 50 variant windows were utilised for LD pruning. MAF: minor allele frequency. B: PCA plot of BiB genotype data calculated to account for population stratification. Blue: South Asians. Red: white Europeans.

Supplementary Figure 2: PCA comparing BiB data to all SA data from 1000G.

A: PCA plot of SA BiB and SA data from 1000G. BiB: Born in Bradford, BEB: Bengali in Bangladesh, GIH: Gujarati Indian from Houston, Texas. ITU: Indian Telugu in the UK. PJL: Punjabi in Lahore, Pakistan STU: Sri Lankan Tamil in the UK. B: PCA plot of BiB data (WE and SA) and South Asian and European data from 1000G.

Supplementary Figure 3: PCA comparing BiB data to all data from 1000G.

A: PCA plot of PC1 vs PC2. B: PCA plot of PC2 vs PC3. C: PCA plot of PC3 vs PC4. PK_BIB: Pakistani BiB sample; WE_BIB: white European BiB Sample.

Supplementary Figure 4: Modified Supplementary Figure 3 including the Pakistani BiB and South Asian 1000G populations.

Supplementary Figure 5: Forest plots of ethnic specific MR estimates for the association between metabolites and fasting glucose.

Forest plots showing β values and 95% CIs of the association between metabolites and fasting glucose. Significant associations (p-value ≤ 0.05) are shown in shaded circles.

Supplementary Figure 6: Forest plots of ethnic specific MR estimates for the association between metabolites and fasting glucose.

Forest plots showing β values and 95% CIs of the association between metabolites and 2-hour post glucose. Significant associations (p-value ≤ 0.05) are shown in shaded circles.

Supplementary Figure 7: Forest plots of β values following leave-one-out analyses for identified associations in white Europeans.

Forest plots showing β values and 95% CIs following the leave-one-out analyses of each SNP in each instrument. Dashed line represents no effect. Associations shown in purple indicating associations with 2-hour post glucose while associations shown in blue indicate associations with fasting glucose. A: Leucine. B: HDL_D. C: HDLC. D: HDL2C. E: HDL3C. F: XS-VLDL-TG. G: XL-HDL-CE. H: L-HDL-P. I: L-HDL_L. J: L-HDL-C. K: S-HDL-CE. L: M-HDL-C. M: M-HDL-CE.

Supplementary Figure 8: F-statistics following leave-one-out analyses for identified associations in white Europeans.

Dashed line indicates an F-statistic of 10, below which an instrument is classified as weak. Associations shown in purple indicate associations with 2-hour post glucose while associations shown in blue indicate associations with fasting glucose. Green bars indicate metabolite measures associated with both fasting glucose and 2-hour post glucose. A: Leucine. B: HDL_D. C: HDLC. D: HDL2C. E: HDL3C. F: XS-VLDL-TG. G: XL-HDL-CE. H: L-HDL-P. I: L-HDL_L. J: L-HDL-C. K: S-HDL-CE. L: M-HDL-C. M: M-HDL-CE

Supplementary Figure 9: Forest plots of β values following leave-one-out analyses for identified associations in South Asians.

Supplementary Figure 10: F-statistics following leave-one- out analyses for identified associations in South Asians.

Dashed line indicates an F-statistic of 10, below which an instrument is classified as weak Associations shown in purple indicate associations with 2-hour post glucose while associations shown in blue indicate associations with fasting glucose. A: FAw6. B: LA. C: M-VLDL-L. D: L-HDL-PL. E: S-HDL-C.

Supplementary Figure 11: Overlap of SNPs identified by class in white Europeans.

Venn diagram showing the overlap between suggestive (1×10^{-5}) SNPs in WEs. SNPs in the apolipoprotein, lipoprotein density, ketone bodies and fluid balance/ inflammation, aromatic amino acids and non-branched amino acids were not found to overlap in WEs. Classes with ≥5 metabolites have been split for clarity. A: Fatty Acids. B: Cholesterols. C: Glycerides and Phospholipids. D: Glycolysis Related Metabolites. E: Branched Amino Acids. Plots were created using the online tool (https://bioinformatics.psb.ugent.be/webtools/Venn/) by the Bioinformatics and Genetics Group at the Department of Plant Systems Biology.

Supplementary Figure 12: Overlap of SNPs identified by class in South Asians.

Venn diagram highlighting the overlap between suggestive (1×10^{-5}) SNPs in SAs. SNPs in the apolipoprotein, lipoprotein density, ketone Bodies and fluid Balance/ inflammation, and amino acids were not found to overlap in WEs. Classes with ≥5 metabolites have been split for clarity. A: Fatty Acids. B: Cholesterols. C: Glycerides and Phospholipids. D: Glycolysis Related Metabolites. Plots were created using the online tool (https://bioinformatics.psb.ugent.be/webtools/Venn/) by the Bioinformatics and Genetics Group at the Department of Plant Systems Biology.

Supplementary Figure 13: Histograms of the number of SNPs and strength of each instrument in the analysis of each metabolite class.

A: Histogram of number of SNPs identified for each metabolite class at the suggestive level ($p \leq 1 \times 10^{-5}$). B: Histogram of the number of SNPs remaining after thinning by LD (R^2>0.2).C: Histogram of the F-statistics for each instrument. Dashed line shows an F-statistic of 10. An F-statistic < 10 is an indicator of weak instrument bias. Blue: South Asians. Red: white Europeans.

Supplementary Figure 14: Forest plots of β values following leave-one-out analyses for identified associations in the analysis of metabolite class.

Forest plots showing β values and 95% CIs following the leave-one-out analyses of each SNP in each instrument. Dashed line represents no effect. Associations shown in purple indicate associations with 2-hour post glucose while associations shown in blue indicate associations with fasting glucose. A: Fatty acid class, SAs. B: XL-HDL class, WEs. C: M-LDL class, WEs. D: All LDL class, WEs. E: S-LDL class, WEs.

Supplementary Figure 15: F-statistics following leave-one-out analyses for identified associations in the analysis of metabolite classes.

Dashed line indicates an F-statistic of 10, below which an instrument is classified as weak. Associations shown in purple indicating associations with 2-hour post glucose while associations shown in blue indicate associations with fasting glucose. Green bars indicate metabolite measures associated with both fasting glucose and 2-hour post glucose. A: Fatty acid class, SAs. B: XL-HDL class, WEs. C: M-LDL class, WEs. D: All LDL class, WEs. E: S-LDL class, WEs.

Supplementary Table 1: List of metabolite values included within this study.

Supplementary Table 2: Proportion of outliers identified in each ethnicity.

Supplementary Table 3: Regions of high linkage disequilibrium (LD) excluded from PCA of genetic data.

BP: Base pair

Supplementary Table 4: Absolute deviations in λ from 1 in each ethnicity.

Average differences of λ values from 1 in each ethnicity averaged across 146 metabolite values. MAF: Minor Allele Frequency.

A: Map of the Indian subcontinent with the location of each South Asian 1000G population illustrated. Brackets represent the country in which the sample was taken from. Colours of labels illustrate data points in PCA plots B: PC1 vs PC2. C: PC2 vs PC3. Base map for panel A was obtained from the *rworld* map package in R studio.

1138

1139 **Supplementary Table 5: Number of SNPs identified in each ethnicity.**

1140 SNPs were classified as being in LD if their R^2 value exceeded 0.2. Suggestive level:
1141 p-value≤ 1 x 10^{-5}. Numbers in brackets represent the number of SNPs associated at
1142 the genome wide significant level (p-value≤ 5 x 10^{-8}).

1143

1144 **Supplementary Table 6: Investigation of pleiotropy in the GWAS Catalog and**
1145 **Phenoscanner databases.**

1146 Bolded SNPs indicate potential confounders. A: Individual metabolites in white
1147 Europeans. B: Individual metabolites in South Asians.

1148

1149 **Supplementary Table 7: Percentage of variation explained by PC1 in each**
1150 **metabolite class.**

1151 Percentage of variation explained by PC1 in the PCA analysis of each metabolite class
1152 following the removal of extreme outliers (3 x IQR). GWAS were performed on all
1153 classes with ≥ 2 metabolites where PC1 explained ≥70% of the variation.

1154

1155 **Supplementary Table 8: Correlation between PC1 and PC2 following outlier**
1156 **removal.**

1157 Correlation between PC1 and PC2 following the removal of outliers (1.5 x IQR) and
1158 extreme outliers (3 x 1QR) in metabolite classes where PC1 explained ≥ 70 the
1159 variation in the metabolite class in at least one ethnicity.

1160

1161 **Supplementary Table 9: Investigation of pleiotropy in the GWAS Catalog and**
1162 **Phenoscanner databases for identified metabolite classes.**

1163

1164 **Supplementary Table 10: Post-hoc power analysis.**

1165 Tables showing *post-hoc* MR calculation from mRND
1166 (https://shiny.cnsgenomics.com/mRnd/). Observational associations were obtained
1167 from linear regression models adjusted for maternal age (years), BMI (continuous),
1168 smoking status, multiple pregnancy, parity, and gestational age. Initial true β estimates
1169 were obtained from unadjusted linear regression models. Additional power analyses
1170 were performed in the non-significant ethnicity to determine the power to predict the β
1171 estimate obtained from the MR analyses and the power to detect the β estimate from

the significant model in the alternative ethnicity. If the power from either analysis exceeded 80% then power was also calculated for $\alpha = 0.01$. Obs: β from adjusted observational studies A: Individual metabolite analysis in South Asians. B: Analysis of metabolite classes in South Asians. C: Analysis of Individual metabolites in White Europeans. D: Analysis of metabolite classes in White Europeans.

Supplementary: Metabolic drivers of dysglycemia in pregnancy: A GWAS of metabolites and one-sample Mendelian Randomisation study in a UK multi-ethnic birth cohort.

Authors: Harriett Fuller, Mark M Iles, J. Bernadette Moore, Michael A. Zulyniak.

*Corresponding Author: Michael Zulyniak. Email. m.a.zulyniak@leeds.ac.uk. Phone +44 113

Table of Contents

Supplementary Figure 1: PCA to account for population stratification. 3

Supplementary Figure 2: PCA comparing BiB data to all SA data from 1000G. . . 4

Supplementary Figure 3: PCA comparing BiB data to all data from 1000G. 5

Supplementary Figure 4: Modified Supplementary Figure 3 including the Pakistani BiB and South Asian 1000G populations. 6

Supplementary Figure 5: Forest plots of ethnic-specific MR estimates for the association between metabolites and fasting glucose. 7

..... 7

Supplementary Figure 5 continued: Forest plots of ethnic-specific MR estimates for the association between metabolites and fasting glucose. 8

..... 8

Supplementary Figure 6: Forest plots of ethnic-specific MR estimates for the association between metabolites and 2-hour post glucose. 9

Supplementary Figure 6 continued: Forest plots of ethnic-specific MR estimates for the association between metabolites and 2-hour post glucose. . . 10

Supplementary Figure 7: Forest plots of β values following leave-one-out analyses for identified associations in white Europeans. 11

Supplementary Figure 7 continued: Forest plots of β values following leave-one-out analyses for identified associations in white Europeans. 12

Supplementary Figure 8: F-statistics following leave-one-out analyses for identified associations in white Europeans. 13

..... 13

Supplementary Figure 8 continued: F-statistics following leave-one-out analyses for identified associations in white Europeans. 14

Supplementary Figure 9: Forest plots of β values following leave-one-out analyses for identified associations in South Asians. 15

..... 15

Supplementary Figure 10: F-statistics following leave-one- out analyses for identified associations in South Asians. 16

Supplementary Figure 11: Overlap of SNPs identified by class in white Europeans. 17

Supplementary Figure 12: Overlap of SNPs identified by class in South Asians. 18

Supplementary Table 1: List of metabolite values included within this study. 19

Supplementary Table 2: Proportion of outliers identified in each ethnicity. 23

Supplementary Table 3: Regions of high linkage disequilibrium (LD) excluded from PCA of genetic data. 26

Supplementary Table 4: Absolute deviations in λ from 1 in each ethnicity. 27

Supplementary Table 5: Number of SNPs identified in each ethnicity 28

Supplementary Table 6: Investigation of pleiotropy in the GWAS Catalog and Phenoscanner databases. 31

Supplementary Table 7: Percentage of variation explained by PC1 in each metabolite class. 37

Supplementary Table 8: Correlation between PC1 and PC2 following outlier removal. 38

Supplementary Figure 13: Histograms of the number of SNPs and strength of each instrument in the analysis of each metabolite class. 39

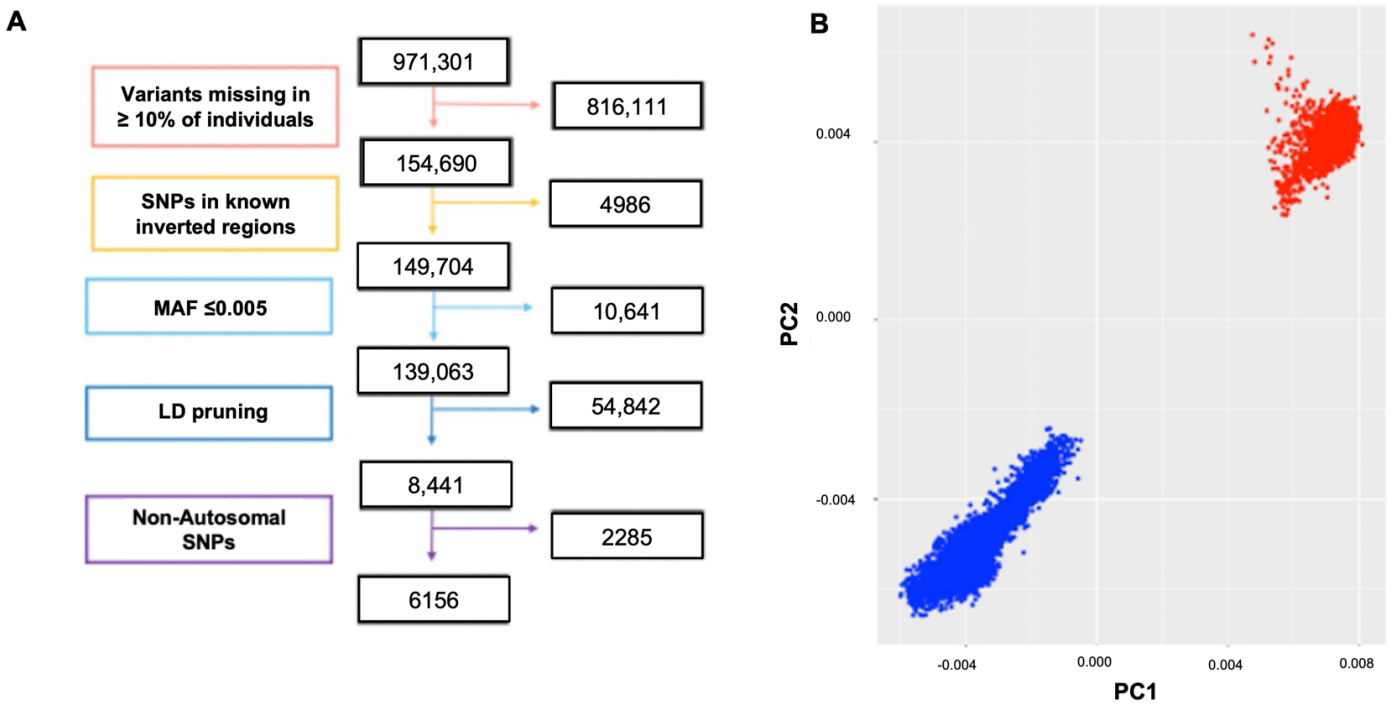
Supplementary Figure 14: Forest plots of β values following leave-one-out analyses for identified associations in the analysis of metabolite class. 40

Supplementary Figure 15: F-statistics following leave-one-out analyses for identified associations in the analysis of metabolite classes. 41

Supplementary Table 9: Investigation of pleiotropy in the GWAS Catalog and Phenoscanner databases for identified metabolite classes. 42

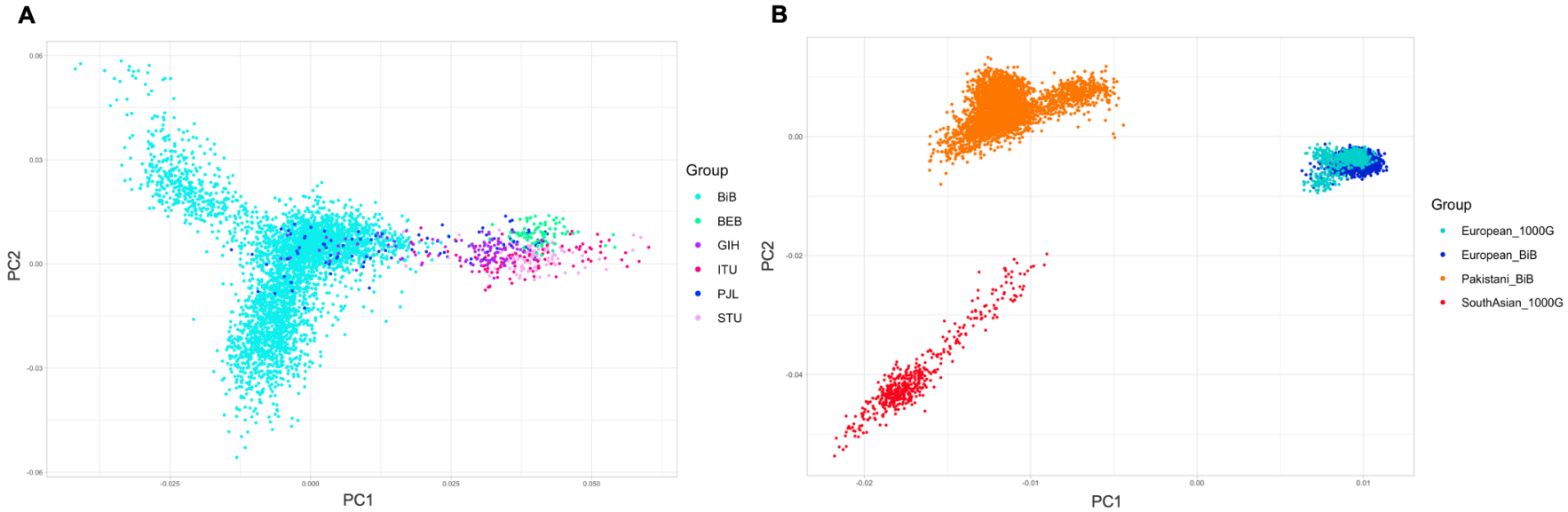
Supplementary Table 10: Post-hoc power analysis. 44

Supplementary Figure 1: PCA to account for population stratification.



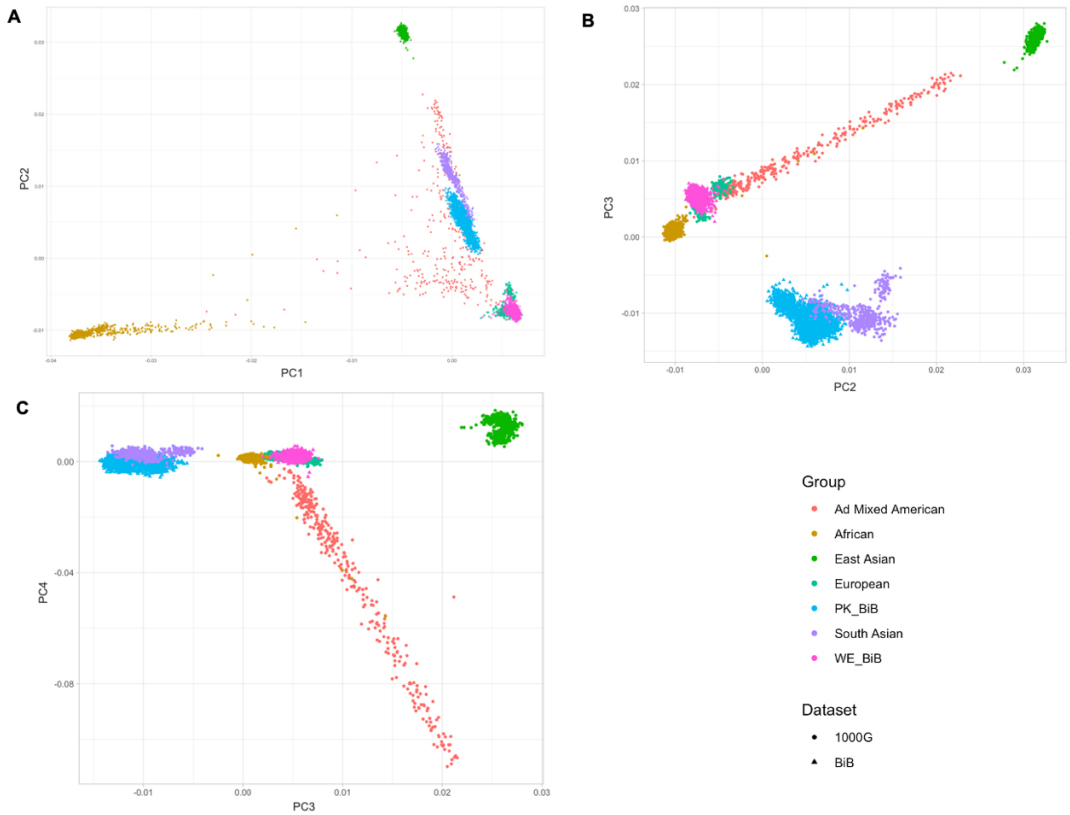
A: Schematic of data processing steps taken prior to PCA of genetic data. A R^2 cut-off of 0.3 and 50 variant windows were utilised for LD pruning. MAF: minor allele frequency. **B**: PCA plot of BiB genotype data calculated to account for population stratification. Blue: South Asians. Red: white Europeans.

Supplementary Figure 2: PCA comparing BiB data to all SA data from 1000G.



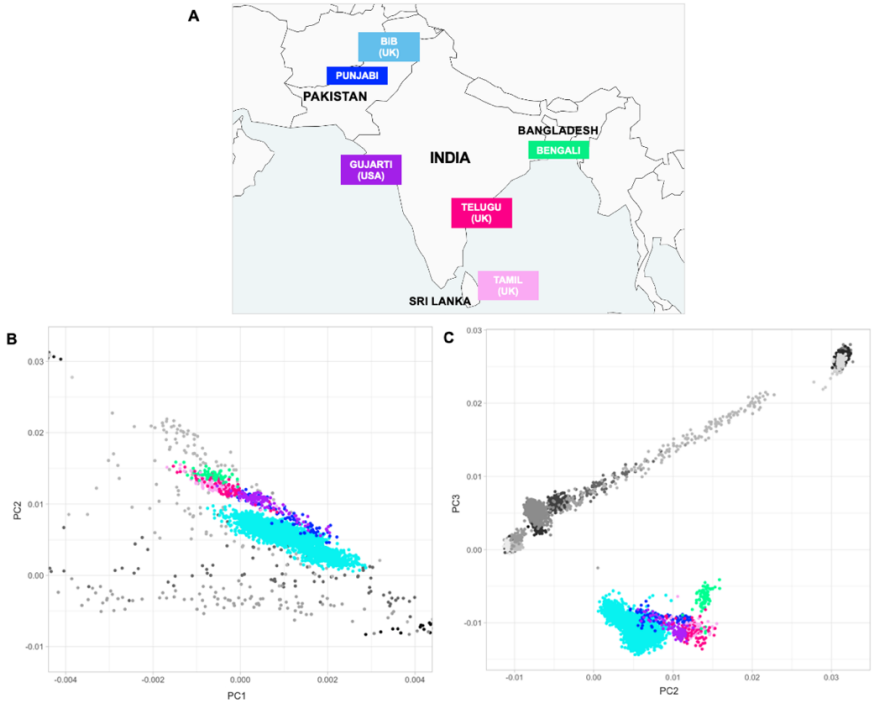
A: PCA plot of SA BiB and SA data from 1000G. BiB: Born in Bradford, BEB: Bengali in Bangladesh, GIH: Gujarati Indian from Houston, Texas. ITU: Indian Telugu in the UK. PJL: Punjabi in Lahore, Pakistan STU: Sri Lankan Tamil in the UK. **B**: PCA plot of BiB data (WE and SA) and South Asian and European data from 1000G.

Supplementary Figure 3: PCA comparing BiB data to all data from 1000G.



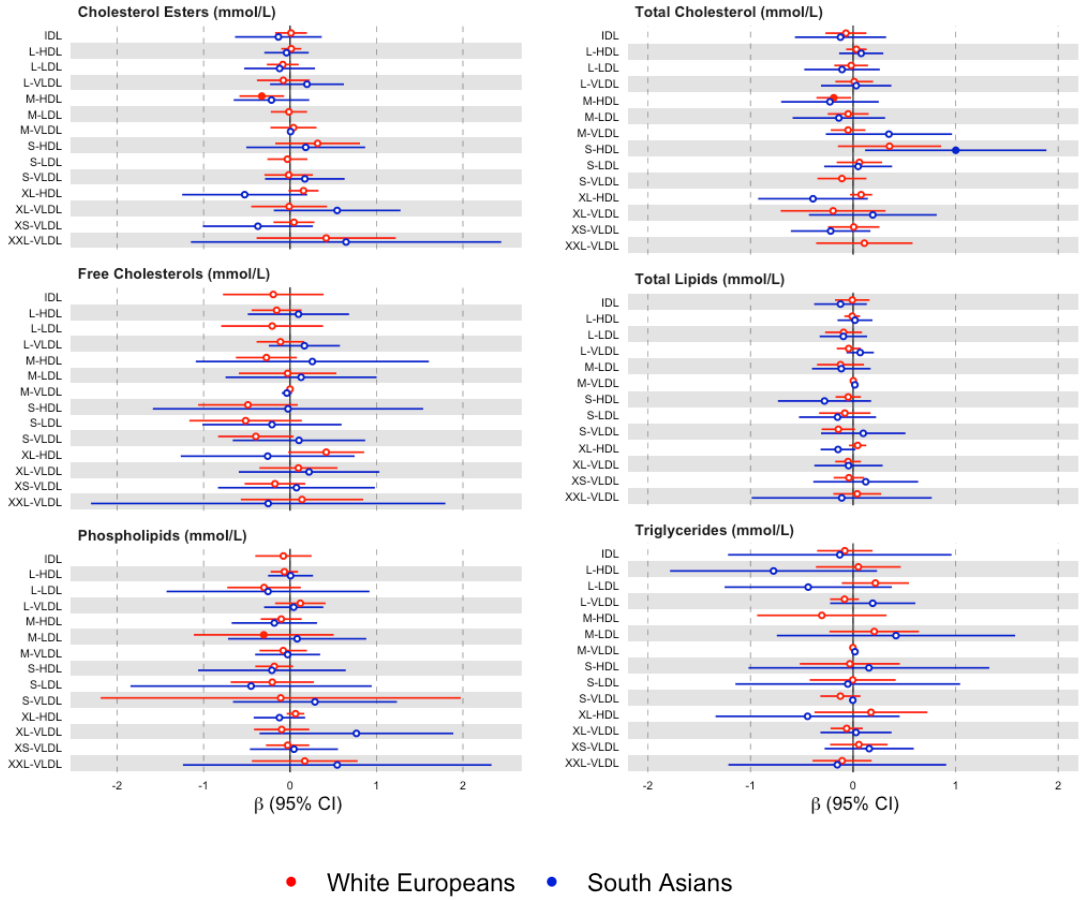
A: PCA plot of PC1 vs PC2. **B**: PCA plot of PC2 vs PC3. **C**: PCA plot of PC3 vs PC4. PK_BIB: Pakistani BiB sample; WE_BIB: white European BiB Sample.

Supplementary Figure 4: Modified Supplementary Figure 3 including the Pakistani BiB and South Asian 1000G populations.



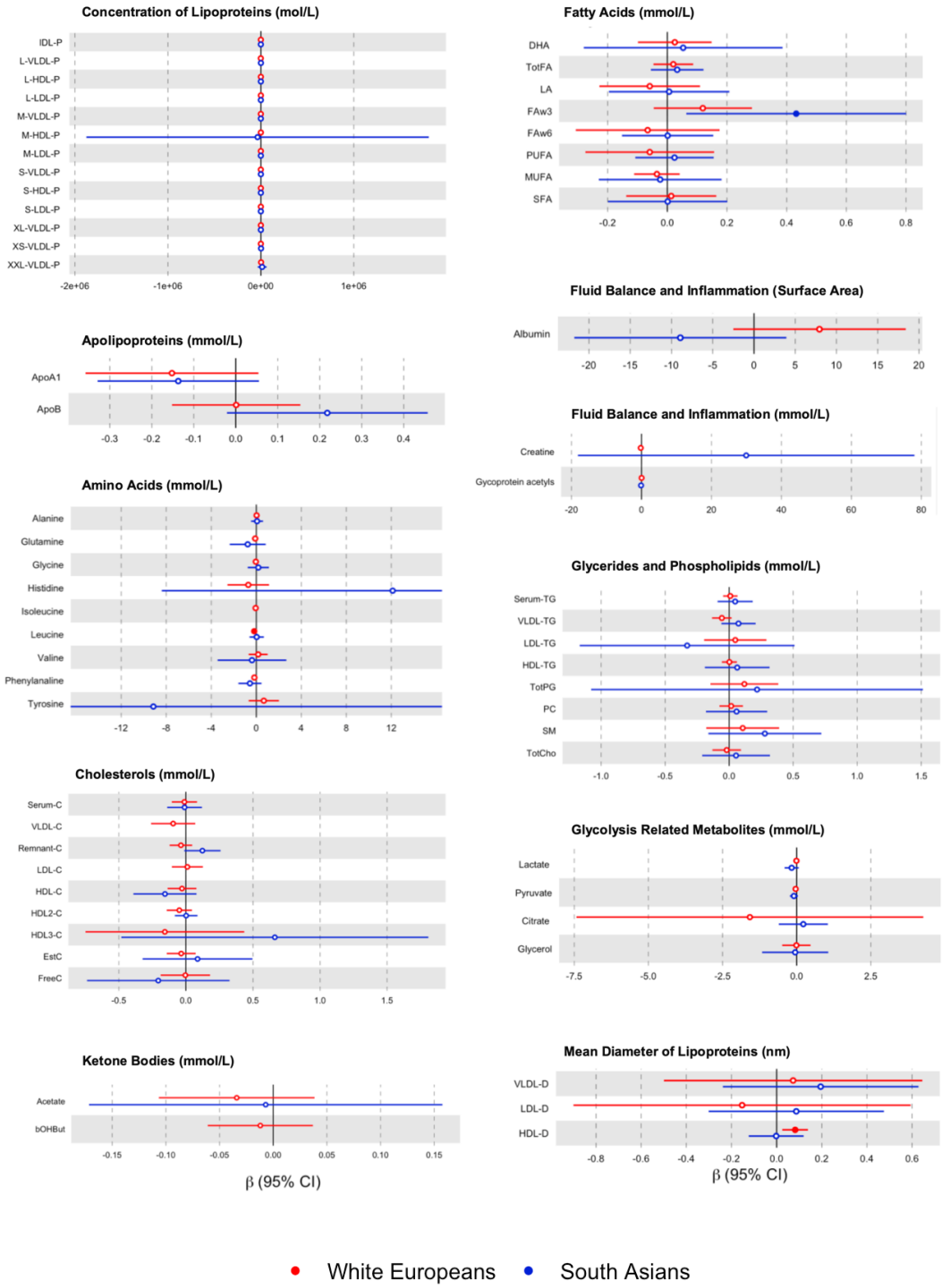
A: Map of the Indian subcontinent with the location of each South Asian 1000G population illustrated. Brackets represent the country in which the sample was taken from. Colours of labels illustrate data points in PCA plots **B:** PC1 vs PC2. **C:** PC2 vs PC3. Base map for panel A was obtained from the *rworld* map package in R studio.

Supplementary Figure 5: Forest plots of ethnic-specific MR estimates for the association between metabolites and fasting glucose.



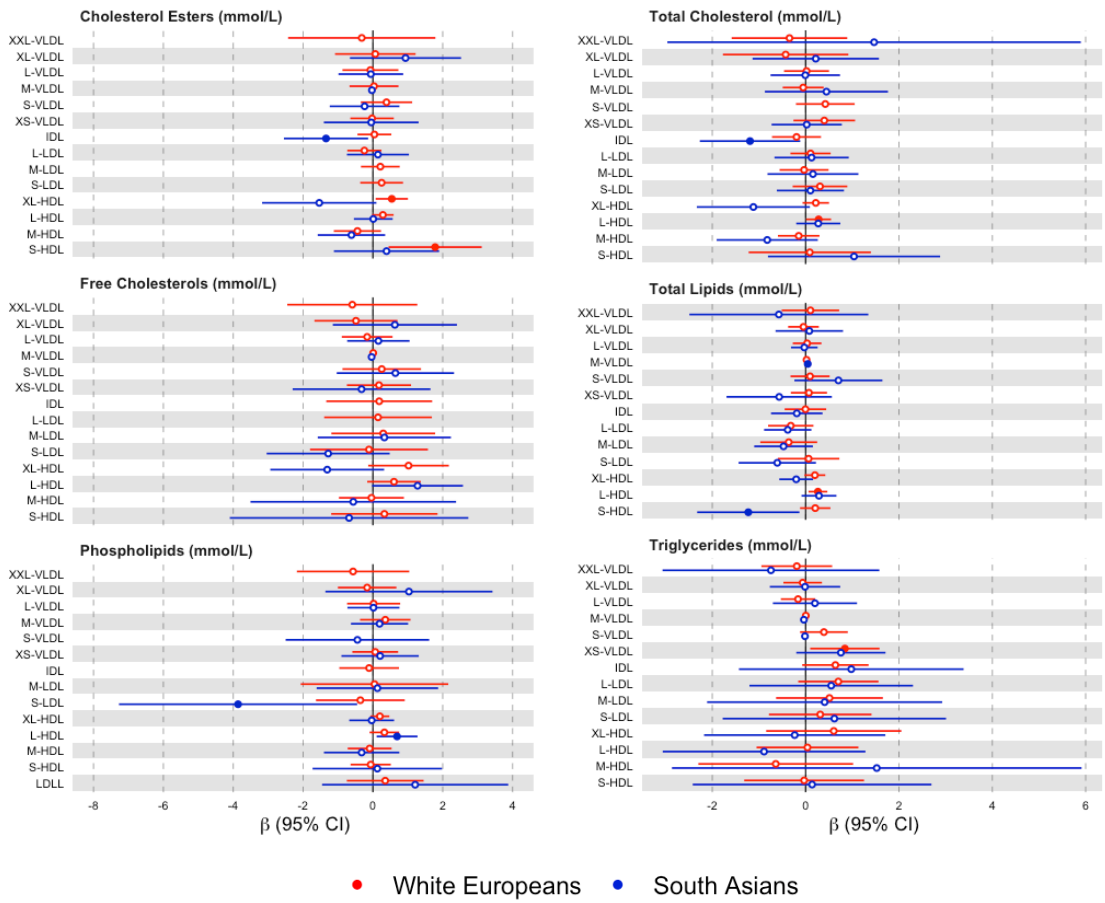
Forest plots showing β values and 95% CIs of the association between metabolites and fasting glucose. Significant associations (P value ≤ 0.05) are shown in shaded circles.

Supplementary Figure 5 continued: Forest plots of ethnic-specific MR estimates for the association between metabolites and fasting glucose.



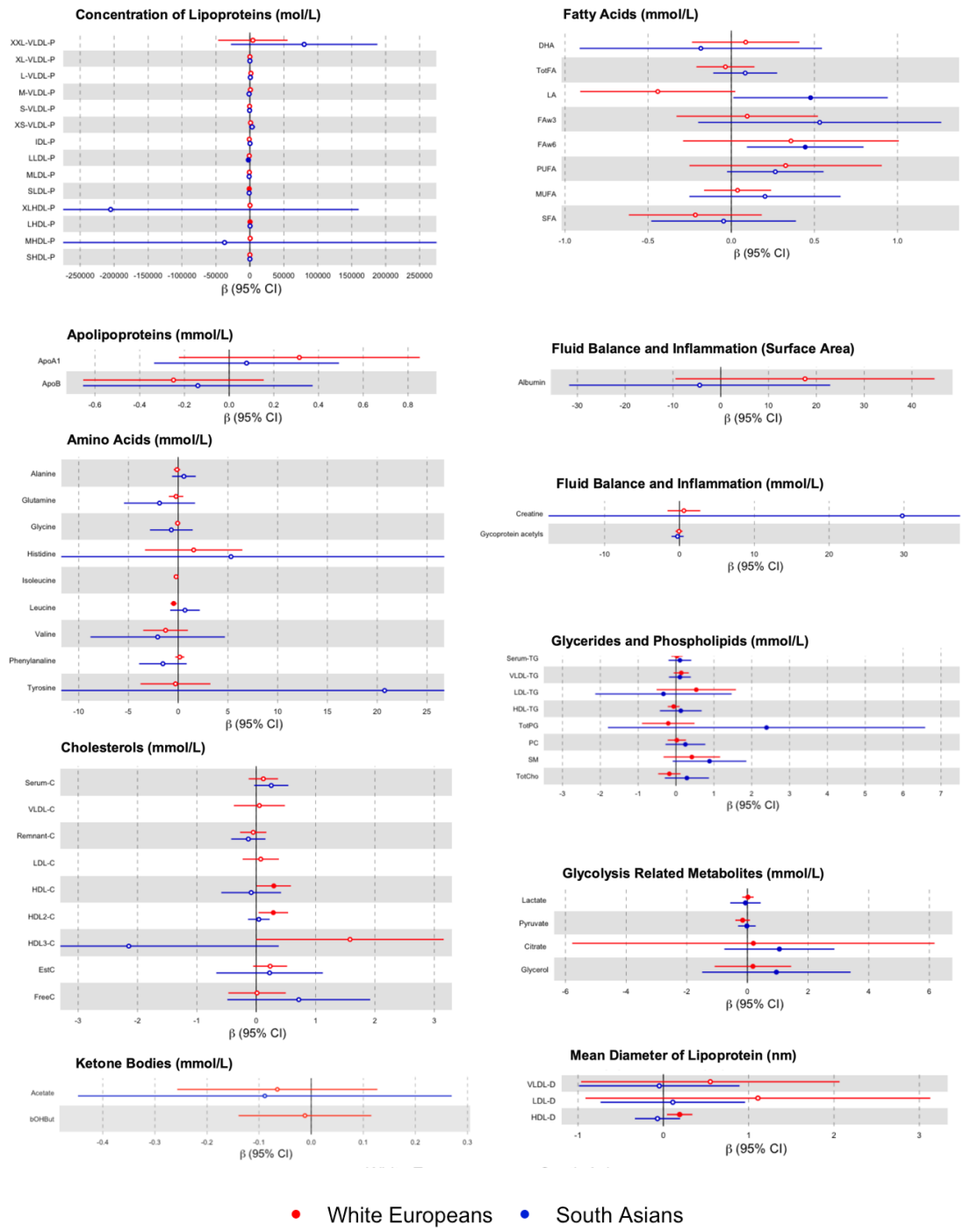
Forest plots showing β values and 95% CIs of the association between metabolites and fasting glucose. Significant associations (P value ≤ 0.05) are shown in shaded circles.

Supplementary Figure 6: Forest plots of ethnic-specific MR estimates for the association between metabolites and 2-hour post glucose.



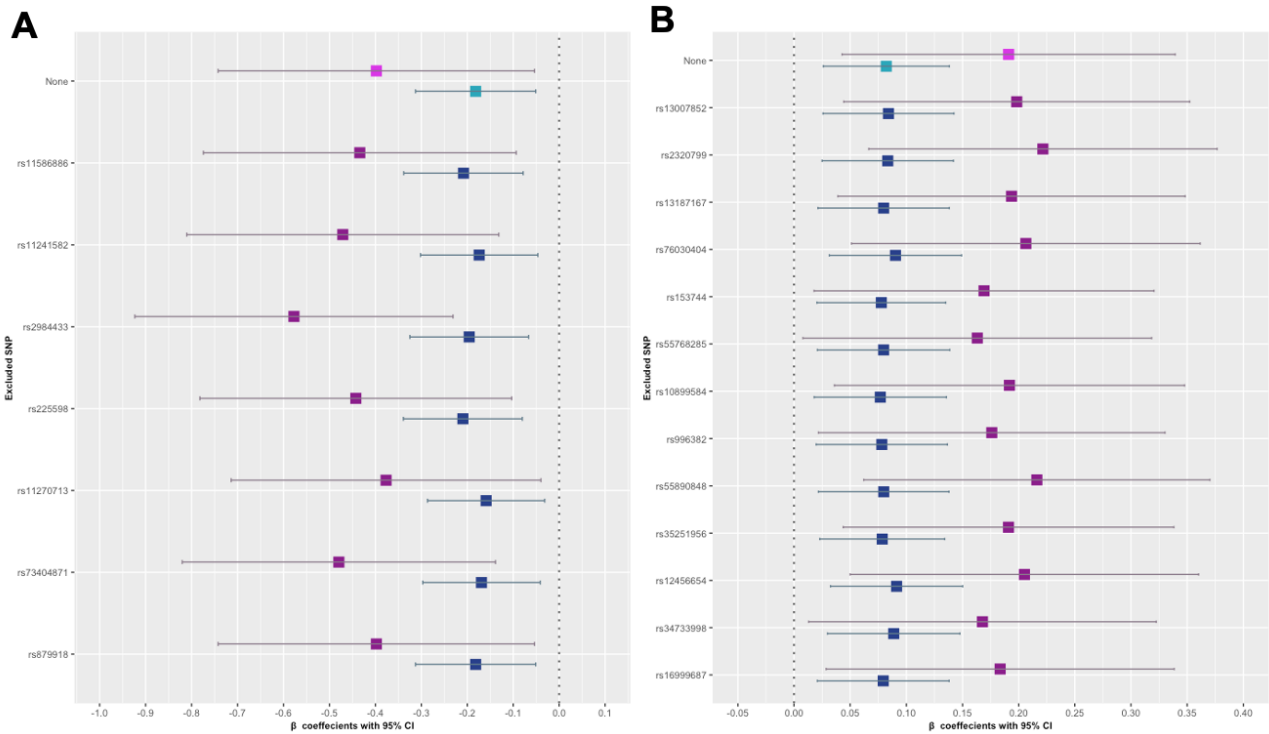
Forest plots showing β values and 95% CIs of the association between metabolites and 2-hour post glucose. Significant associations (P value ≤ 0.05) are shown in shaded circles.

Supplementary Figure 6 continued: Forest plots of ethnic-specific MR estimates for the association between metabolites and 2-hour post glucose.



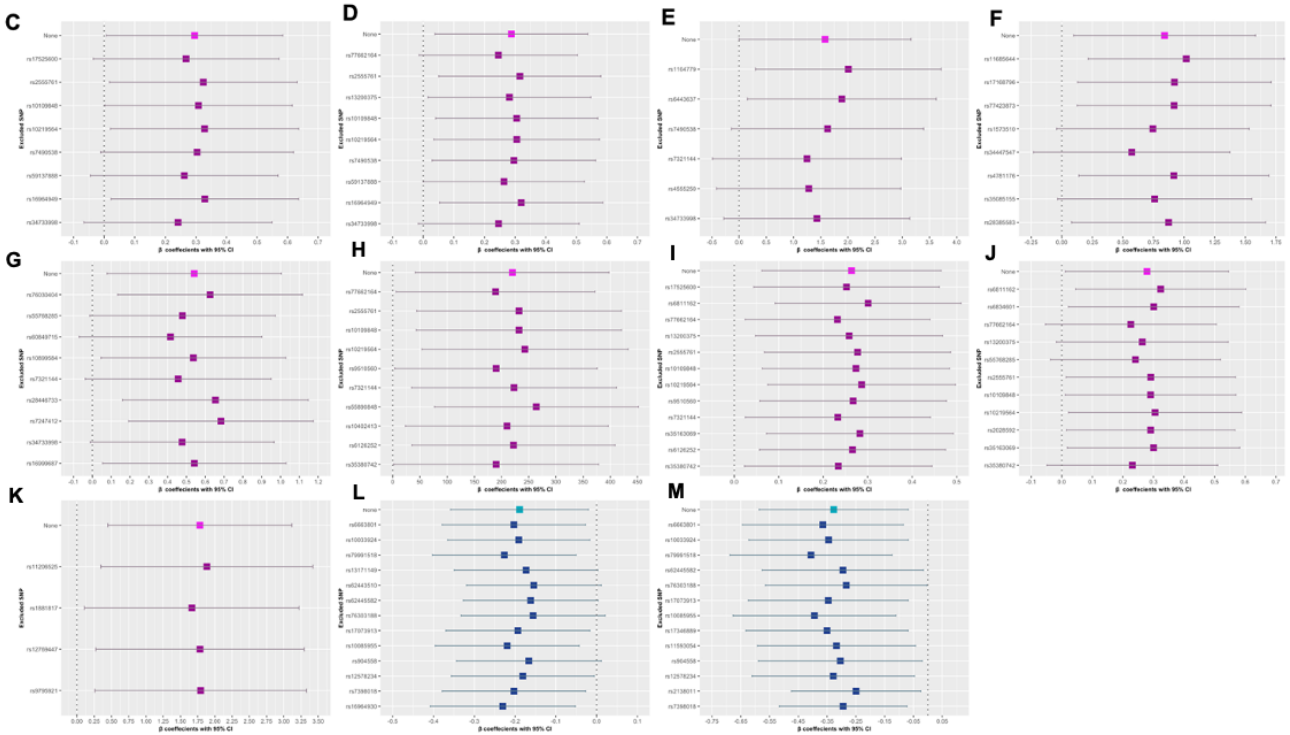
Forest plots showing β values and 95% CIs of the association between metabolites and 2-hour post glucose. Significant associations (P value ≤ 0.05) are shown in shaded circles.

Supplementary Figure 7: Forest plots of β values following leave-one-out analyses for identified associations in white Europeans.



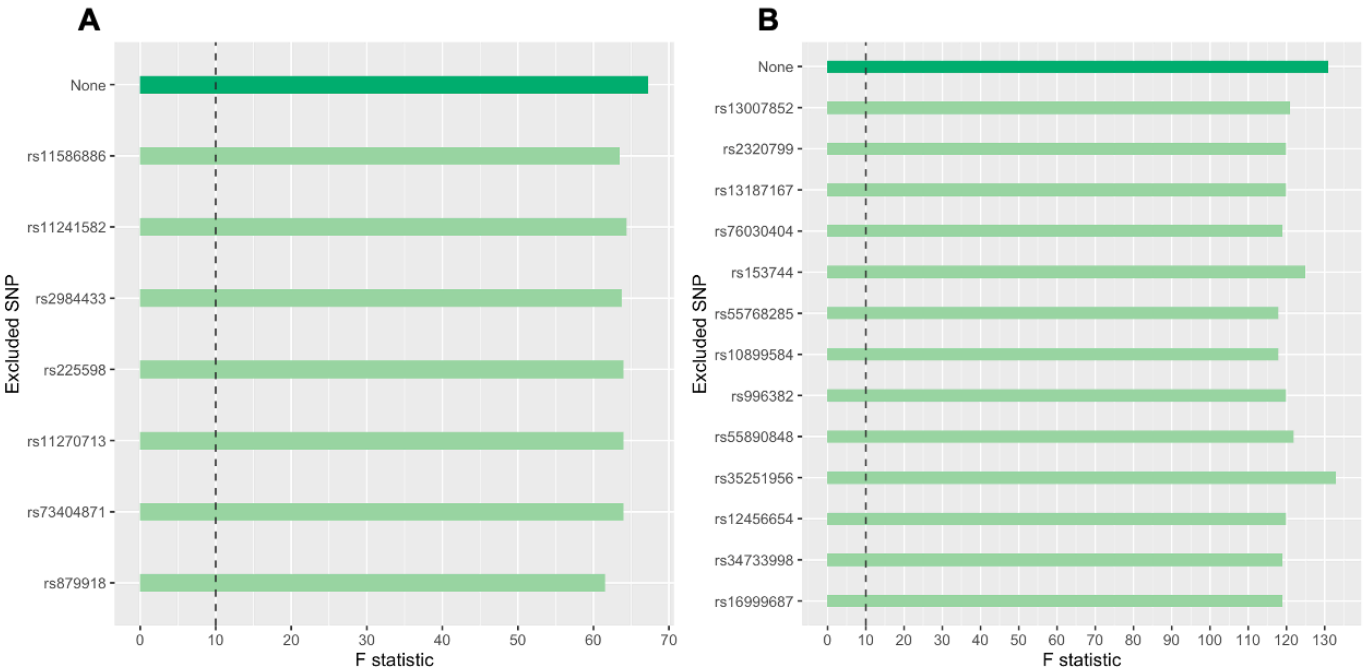
Forest plots showing β values and 95% CIs following the leave-one-out analyses of each SNP in each instrument. Dashed line represents no effect. Associations shown in purple indicating associations with 2-hour post glucose while associations shown in blue indicate associations with fasting glucose. **A**: Leucine. **B**: HDL_D.

Supplementary Figure 7 continued: Forest plots of β values following leave-one-out analyses for identified associations in white Europeans.



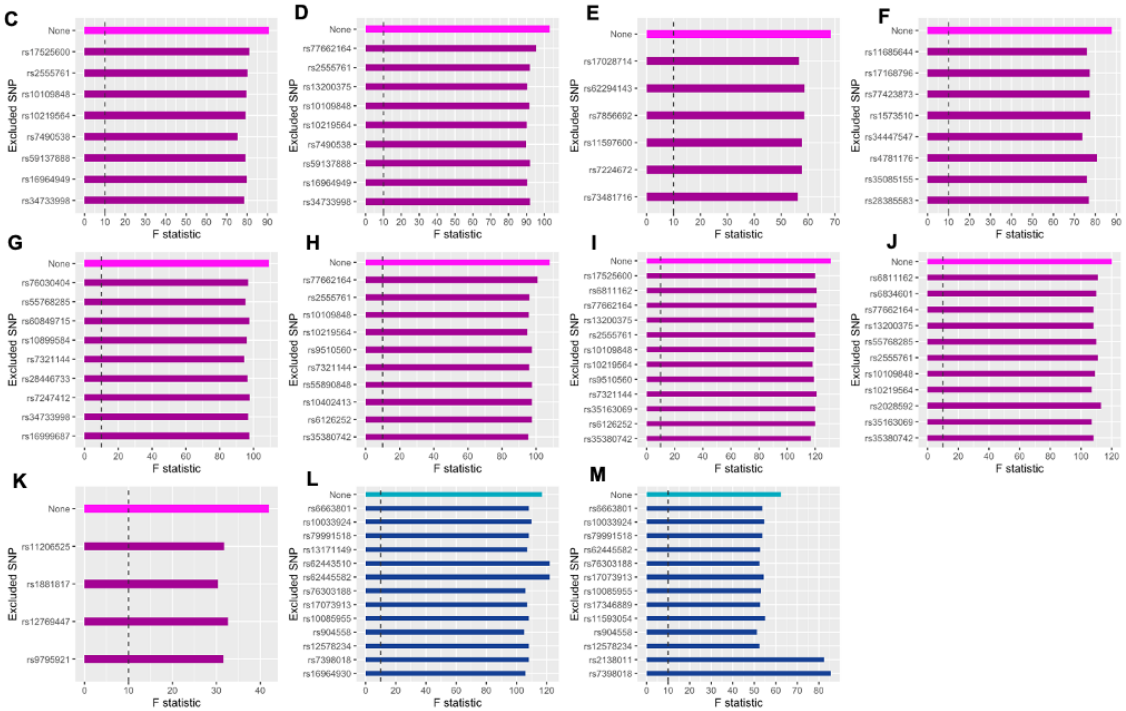
Forest plots showing β values and 95% CIs following the leave-one-out analyses of each SNP in each instrument. Dashed line represents no effect. Associations shown in purple indicating associations with 2-hour post glucose while associations shown in blue indicate associations with fasting glucose. **C**: HDLC. **D**: HDL2C. **E**: HDL3C. **F**: XS-VLDL-TG. **G**: XL-HDL-CE. **H**: L-HDL-P. **I**: L-HDL_L. **J**: L-HDL-C. **K**: S-HDL-CE. **L**: M-HDL-C. **M**: M-HDL-CE.

Supplementary Figure 8: F-statistics following leave-one-out analyses for identified associations in white Europeans.



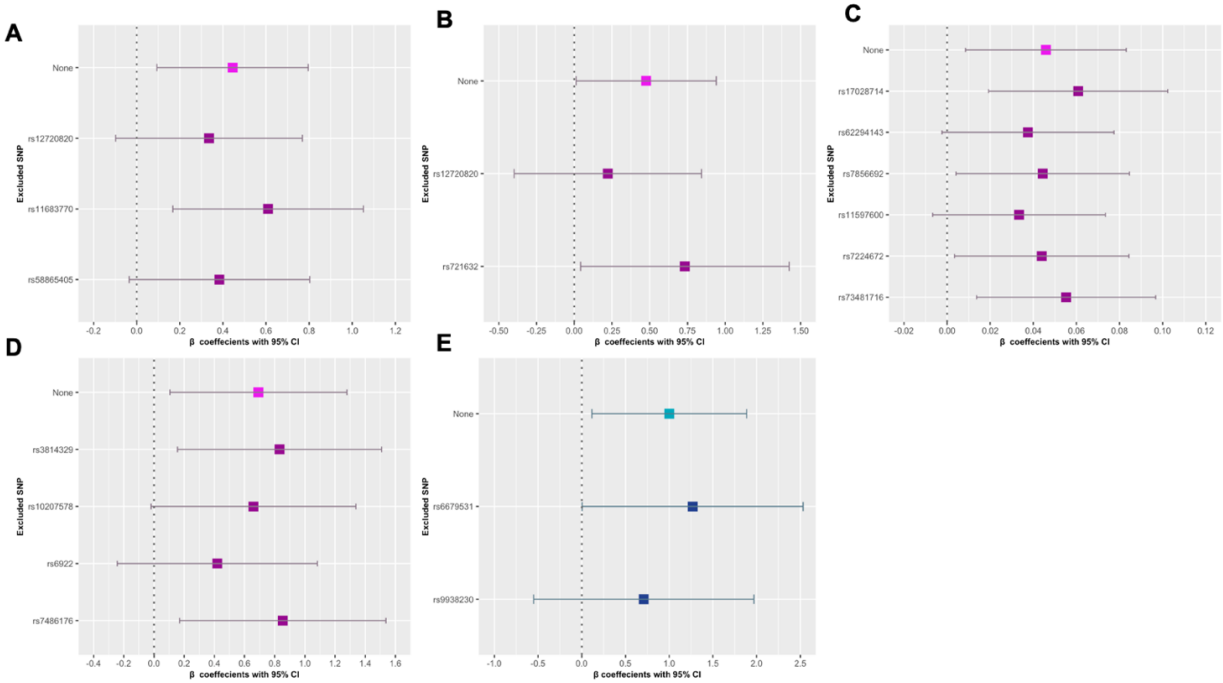
Dashed line indicates a F statistic of 10, below which an instrument is classified as weak. Associations shown in purple indicate associations with 2-hour post glucose while associations shown in blue indicate associations with fasting glucose. Green bars indicate metabolite measures associated with both fasting glucose and 2-hour post glucose. **A:** Leucine. **B**: HDL_D.

Supplementary Figure 8 continued: F-statistics following leave-one-out analyses for identified associations in white Europeans.



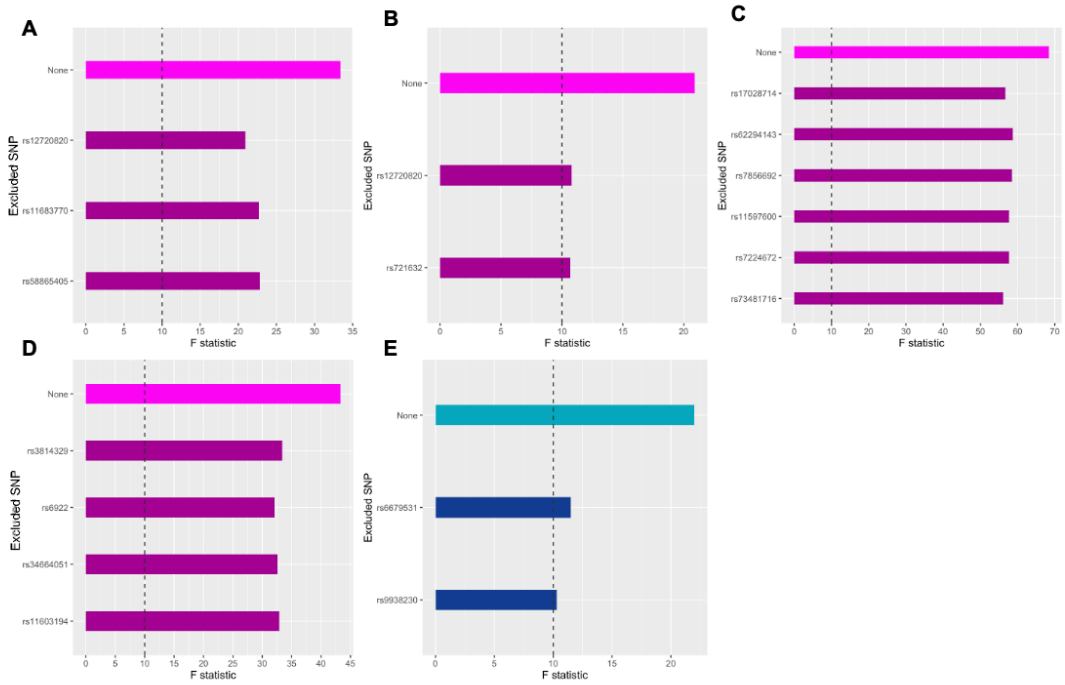
Dashed line indicates an F statistic of 10, below which an instrument is classified as weak. Associations shown in purple indicate associations with 2-hour post glucose while associations shown in blue indicate associations with fasting glucose. Green bars indicate metabolite measures associated with both fasting glucose and 2-hour post glucose. **C:** HDLC. **D:** HDL2C. **E:** HDL3C. **F:** XS-VLDL-TG. **G:** XL-HDL-CE. **H:** L-HDL-P. **I:** L-HDL_L. **J:** L-HDL-C. **K:** S-HDL-CE. **L:** M-HDL-C. **M:** M-HDL-CE

Supplementary Figure 9: Forest plots of β values following leave-one-out analyses for identified associations in South Asians.



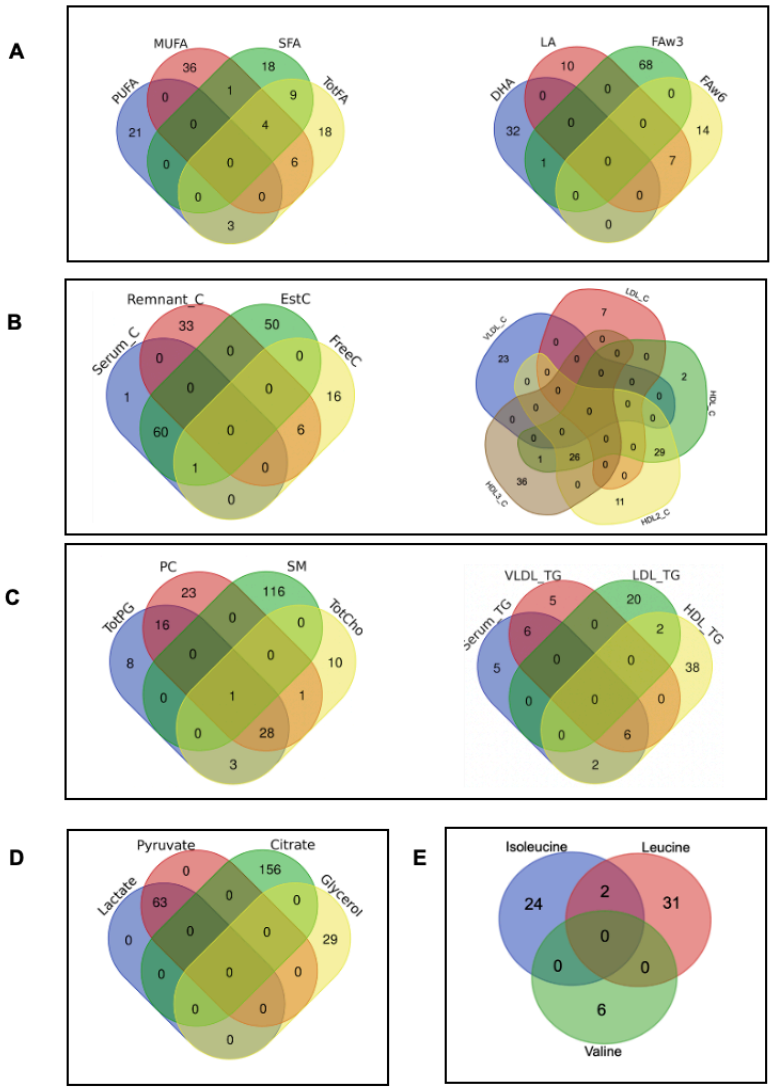
Forest plots showing β values and 95% CIs following the leave-one-out analyses of each SNP in each instrument. Dashed line represents no effect. Associations shown in purple indicating associations with 2-hour post glucose while associations shown in blue indicate associations with fasting glucose. **A:** FAw6. **B:** LA. **C:** M-VLDL-L. **D:** L-HDL-PL. **E:** S-HDL-C.

Supplementary Figure 10: F-statistics following leave-one- out analyses for identified associations in South Asians.



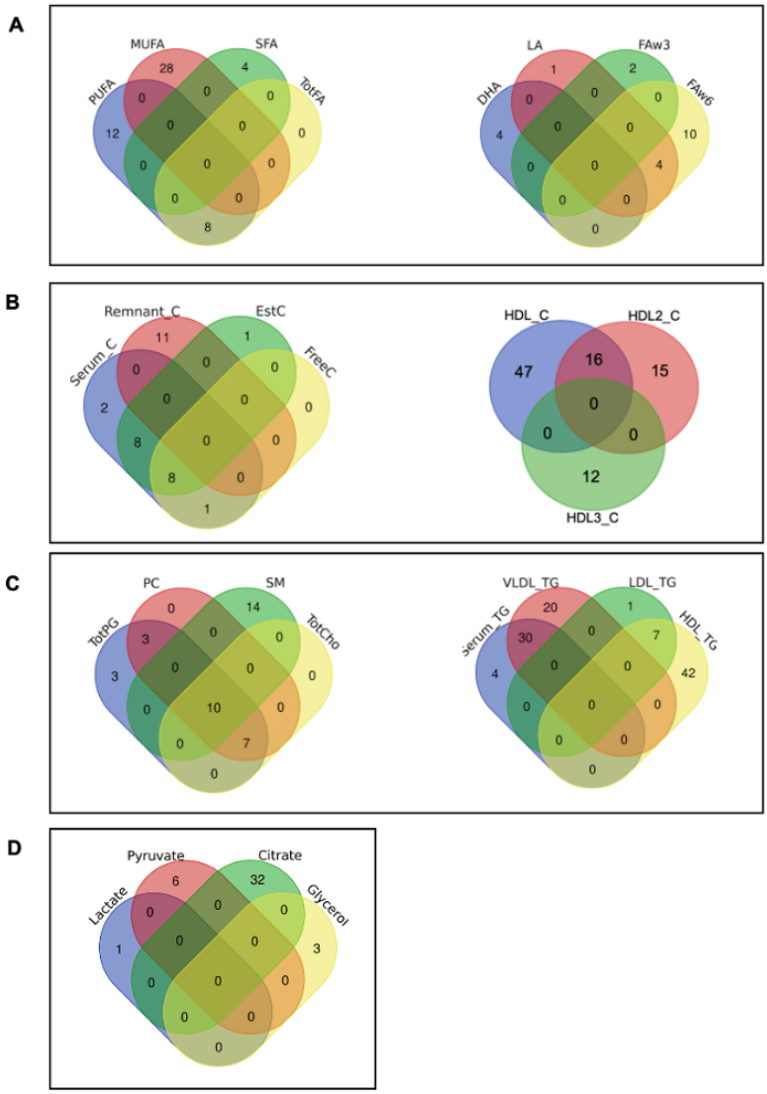
Dashed line indicates an F statistic of 10, below which an instrument is classified as weak Associations shown in purple indicate associations with 2-hour post glucose while associations shown in blue indicate associations with fasting glucose. **A:** FAw6. **B:** LA. **C:** M-VLDL-L. **D:** L-HDL-PL. **E:** S-HDL-C.

Supplementary Figure 11: Overlap of SNPs identified by class in white Europeans.



Venn diagram showing the overlap between suggestive (1 x 10^{-5}) SNPs in WEs. SNPs in the apolipoprotein, lipoprotein density, ketone bodies and fluid balance/ inflammation, aromatic amino acids and non-branched amino acids were not found to overlap in WEs. Classes with ≥5 metabolites have been split for clarity. **A**: Fatty Acids. **B**: Cholesterols. **C**: Glycerides and Phospholipids. **D**: Glycolysis Related Metabolites. **E**: Branched Amino Acids. Plots were created using the online tool (https://bioinformatics.psb.ugent.be/webtools/Venn/) by the Bioinformatics and Genetics Group at the Department of Plant Systems Biology

Supplementary Figure 12: Overlap of SNPs identified by class in South Asians.



Venn diagram highlighting the overlap between suggestive (1 x 10^{-5}) SNPs in SAs. SNPs in the apolipoprotein, lipoprotein density, ketone Bodies and fluid Balance/ inflammation, and amino acids were not found to overlap in WEs. Classes with ≥5 metabolites have been split for clarity. **A**: Fatty Acids. **B**: Cholesterols. **C**: Glycerides and Phospholipids. **D**: Glycolysis Related Metabolites. Plots were created using the online tool (https://bioinformatics.psb.ugent.be/webtools/Venn/) by the Bioinformatics and Genetics Group at the Department of Plant Systems Biology.

Supplementary Table 1: List of metabolite values included within this study.

Metabolite Class	Metabolite Code	Full Name (Unit)
XXL-VLDL	XXL-VLDL-P	Concentration of XXL-VLDL (mol/l)
	XXL-VLDL-L	Total lipids in XXL-VLDL (mmol/l)
	XXL-VLDL-PL	Phospholipids in XXL-VLDL (mmol/l)
	XXL-VLDL-C	Total Cholesterol in XXL-VLDL (mmol/l)
	XXL-VLDL-CE	Cholesterol esters in XXL-VLDL (mmol/l)
	XXL-VLDL-FC	Free cholesterol in XXL-VLDL (mmol/l)
	XXL-VLDL-TG	Triglycerides in XXL-VLDL (mmol/l)
XL-VLDL	XL-VLDL-P	Concentration of XL-VLDL (mol/l)
	XL-VLDL-L	Total lipids in XL-VLDL (mmol/l)
	XL-VLDL-PL	Phospholipids in XL-VLDL (mmol/l)
	XL-VLDL-C	Total Cholesterol in XL-VLDL (mmol/l)
	XL-VLDL-CE	Cholesterol esters in XL-VLDL (mmol/l)
	XL-VLDL-FC	Free cholesterol in XL-VLDL (mmol/l)
	XL-VLDL-TG	Triglycerides in XL-VLDL (mmol/l)
L-VLDL	L-VLDL-P	Concentration of L-VLDL (mol/l)
	L-VLDL-L	Total lipids in L-VLDL (mmol/l)
	L-VLDL-PL	Phospholipids in L-VLDL (mmol/l)
	L-VLDL-C	Total Cholesterol in L-VLDL (mmol/l)
	L-VLDL-CE	Cholesterol esters in L-VLDL (mmol/l)
	L-VLDL-FC	Free cholesterol in L-VLDL (mmol/l)
	L-VLDL-TG	Triglycerides in L-VLDL (mmol/l)
M-VLDL	M-VLDL-P	Concentration of M-VLDL (mol/l)
	M-VLDL-L	Total lipids in M-VLDL (mmol/l)
	M-VLDL-PL	Phospholipids in M-VLDL (mmol/l)
	M-VLDL-C	Total Cholesterol in M-VLDL (mmol/l)
	M-VLDL-CE	Cholesterol esters in M-VLDL (mmol/l)
	M-VLDL-FC	Free cholesterol in M-VLDL (mmol/l)
	M-VLDL-TG	Triglycerides in M-VLDL (mmol/l)
S-VLDL	S-VLDL-P	Concentration of S-VLDL (mol/l)
	S-VLDL-L	Total lipids in S-VLDL (mmol/l)
	S-VLDL-PL	Phospholipids in S-VLDL (mmol/l)
	S-VLDL-C	Total Cholesterol in S-VLDL (mmol/l)
	S-VLDL-CE	Cholesterol esters in S-VLDL (mmol/l)
	S-VLDL-FC	Free cholesterol in S-VLDL (mmol/l)
	S-VLDL-TG	Triglycerides in S-VLDL (mmol/l)
XS-VLDL	XS-VLDL-P	Concentration of XS-VLDL (mol/l)
	XS-VLDL-L	Total lipids in XS-VLDL (mmol/l)
	XS-VLDL-PL	Phospholipids in XS-VLDL (mmol/l)
	XS-VLDL-C	Total Cholesterol in XS-VLDL (mmol/l)
	XS-VLDL-CE	Cholesterol esters in XS-VLDL (mmol/l)
	XS-VLDL-FC	Free cholesterol in XS-VLDL (mmol/l)

	XS-VLDL-TG	Triglycerides in XS-VLDL (mmol/l)
IDL	IDL-P	Concentration of IDL-P (mol/l)
	IDL-L	Total lipids in IDL-L (mmol/l)
	IDL-PL	Phospholipids in IDL-PL (mmol/l)
	IDL-C	Total Cholesterol in IDL-C (mmol/l)
	IDL-CE	Cholesterol esters in IDL-CE (mmol/l)
	IDL-FC	Free cholesterol in IDL-FC (mmol/l)
	IDL-TG	Triglycerides in IDL-TG (mmol/l)
L-LDL	L-LDL-P	Concentration of L-LDL-P (mol/l)
	L-LDL-L	Total lipids in L-LDL-L (mmol/l)
	L-LDL-PL	Phospholipids in L-LDL-PL (mmol/l)
	L-LDL-C	Total Cholesterol in L-LDL-C (mmol/l)
	L-LDL-CE	Cholesterol esters in L-LDL-CE (mmol/l)
	L-LDL-FC	Free cholesterol in L-LDL-FC (mmol/l)
	L-LDL-TG	Triglycerides in L-LDL-TG (mmol/l)
M-LDL	M-LDL-P	Concentration of M-LDL-P (mol/l)
	M-LDL-L	Total lipids in M-LDL-L (mmol/l)
	M-LDL-PL	Phospholipids in M-LDL-PL (mmol/l)
	M-LDL-C	Total Cholesterol in M-LDL-C (mmol/l)
	M-LDL-CE	Cholesterol esters in M-LDL-CE (mmol/l)
	M-LDL-FC	Free cholesterol in M-LDL-FC (mmol/l)
	M-LDL-TG	Triglycerides in M-LDL-TG (mmol/l)
S-LDL	S-LDL-P	Concentration of S-LDL-P (mol/l)
	S-LDL-L	Total lipids in S-LDL-L (mmol/l)
	S-LDL-PL	Phospholipids in S-LDL-PL (mmol/l)
	S-LDL-C	Total Cholesterol in S-LDL-C (mmol/l)
	S-LDL-CE	Cholesterol esters in S-LDL-CE (mmol/l)
	S-LDL-FC	Free cholesterol in S-LDL-FC (mmol/l)
	S-LDL-TG	Triglycerides in S-LDL-TG (mmol/l)
XL-HDL	XL-HDL-P	Concentration of XL-HDL-P (mol/l)
	XL-HDL-L	Total lipids in XL-HDL-L (mmol/l)
	XL-HDL-PL	Phospholipids in XL-HDL-PL (mmol/l)
	XL-HDL-C	Total Cholesterol in XL-HDL-C (mmol/l)
	XL-HDL-CE	Cholesterol esters in XL-HDL-CE (mmol/l)
	XL-HDL-FC	Free cholesterol in XL-HDL-FC (mmol/l)
	XL-HDL-TG	Triglycerides in XL-HDL-TG (mmol/l)
L-HDL	L-HDL-P	Concentration of L-HDL-P (mol/l)
	L-HDL-L	Total lipids in L-HDL-L (mmol/l)
	L-HDL-PL	Phospholipids in L-HDL-PL (mmol/l)
	L-HDL-C	Total Cholesterol in L-HDL-C (mmol/l)
	L-HDL-CE	Cholesterol esters in L-HDL-CE (mmol/l)
	L-HDL-FC	Free cholesterol in L-HDL-FC (mmol/l)
	L-HDL-TG	Triglycerides in L-HDL-TG (mmol/l)
M-HDL	M-HDL-P	Concentration of M-HDL-P (mol/l)
	M-HDL-L	Total lipids in M-HDL-L (mmol/l)
	M-HDL-PL	Phospholipids in M-HDL-PL (mmol/l)
	M-HDL-C	Total Cholesterol in M-HDL-C (mmol/l)
	M-HDL-CE	Cholesterol esters in M-HDL-CE (mmol/l)
	M-HDL-FC	Free cholesterol in M-HDL-FC (mmol/l)

	M-HDL-TG	Triglycerides in M-HDL-TG (mmol/l)
S-HDL	S-HDL-P	Concentration of S-HDL-P (mol/l)
	S-HDL-L	Total lipids in S-HDL-L (mmol/l)
	S-HDL-PL	Phospholipids in S-HDL-PL (mmol/l)
	S-HDL-C	Total Cholesterol in S-HDL-C (mmol/l)
	S-HDL-CE	Cholesterol esters in S-HDL-CE (mmol/l)
	S-HDL-FC	Free cholesterol in S-HDL-FC (mmol/l)
	S-HDL-TG	Triglycerides in S-HDL-TG (mmol/l)
Lipoprotien Density	VLDL_D	Mean diameter of VLDL (nm)
	LDL_D	Mean diameter of LDL (nm)
	HDL_D	Mean diameter of HDL (nm)
Cholesterols	Serum_C	Total serum cholesterol (mmol/l)
	VLDL_C	VLDL cholesterol (mmol/l)
	Remnant_C	Remnant cholesterol* (mmol/l)
	LDL_C	LDL cholesterol (mmol/l)
	HDL_C	HDL cholesterol (mmol/l)
	HDL2_C	HDL2 cholesterol (mmol/l)
	HDL3_C	HDL3 cholesterol (mmol/l)
	EstC	Total esterified cholesterol (mmol/l)
	FreeC	Free Cholesterol (mmol/l)
Glycerides and Phospholipids	Serum_TG	Total serum triglycerides (mmol/l)
	VLDL-TG	VLDL triglycerides (mmol/l)
	LDL-TG	LDL triglycerides (mmol/l)
	HDL-TG	HDL triglycerides (mmol/l)
	TotPG	Total phosphoglycerides (mmol/l)
	PC	Phosphatidylcholine* (mmol/l)
	SM	Sphingomyelins (mmol/l)
	TotCho	Total cholines (mmol/l)
Apolipoproteins	ApoA1	Apolipoprotein A1 (g/l)
	ApoB	Apolipoprotein B (g/l)
Fatty Acids	TotFA	Total fatty acids (mmol/l)
	DHA	22:6, docosahexaenoic acid (mmol/l)
	LA	18:2, linoleic acid (mmol/l)
	FAw3	Omega-3 fatty acids (mmol/l)
	FAw6	Omega-6 fatty acids (mmol/l)
	PUFA	Polyunsaturated fatty acids
	MUFA	Monounsaturated fatty acids ** (mmol/l)
	SFA	Saturated fatty acids (mmol/l)
Glycolysis Related Metabolites	Lac	Lactate (mmol/l)
	Pyr	Pyruvate (mmol/l)
	Cit	Citrate (mmol/l)
	Glol	Glycerol (mmol/l)
Amino Acids	Ala	Alanine (mmol/l)
	Gln	Glycine (mmol/l)
	Gly	Glycerol (mmol/l)
	His	Histidine (mmol/l)
	Ile	Isoleucine (mmol/l)
	Leu	Leucine (mmol/l)
	Val	Valine (mmol/l)
	Phe	Phenylalanine (mmol/l)

	Tyr	Tyrosine (mmol/l)
Ketone Bodies	Ace	Acetate (mmol/l)
	bOHBut	3-hydroxybutyrate (mmol/l)
Fluid Balance and Inflammation	Crea	Creatine (mmol/l)
	Alb	Albumin (signal area)
	Gp	Glycoprotein acetyls (mmol/l)

Supplementary Table 2: Proportion of outliers identified in each ethnicity.

Metabolite	South Asian		White European	
	Final Sample Size	% Removed Outliers	Final Sample Size	% Removed Outliers
XXL-VLDL-P	3586	2.90	3247	3.79
XXL-VLDL-L	3589	2.82	3250	3.70
XXL-VLDL-PL	3592	2.73	3253	3.61
XXL-VLDL-C	3605	2.38	3270	3.11
XXL-VLDL-CE	3615	2.11	3276	2.93
XXL-VLDL-FC	3591	2.76	3264	3.29
XXL-VLDL-TG	3584	2.95	3241	3.97
XL-VLDL-P	3585	2.92	3261	3.38
XL-VLDL-L	3587	2.87	3261	3.38
XL-VLDL-PL	3587	2.87	3255	3.56
XL-VLDL-C	3599	2.55	3259	3.44
XL-VLDL-CE	3600	2.52	3256	3.53
XL-VLDL-FC	3595	2.65	3258	3.47
XL-VLDL-TG	3587	2.87	3262	3.35
L-VLDL-P	3588	2.84	3264	3.29
L-VLDL-L	3592	2.73	3265	3.26
L-VLDL-PL	3596	2.63	3266	3.23
L-VLDL-C	3590	2.79	3257	3.50
L-VLDL-CE	3607	2.33	3256	3.53
L-VLDL-FC	3593	2.71	3265	3.26
L-VLDL-TG	3588	2.84	3262	3.35
M-VLDL-P	3601	2.49	3266	3.23
M-VLDL-L	3603	2.44	3261	3.38
M-VLDL-PL	3610	2.25	3268	3.17
M-VLDL-C	3614	2.14	3271	3.08
M-VLDL-CE	3616	2.09	3289	2.55
M-VLDL-FC	3608	2.30	3266	3.23
M-VLDL-TG	3589	2.82	3268	3.17
S-VLDL-P	3617	2.06	3295	2.37
S-VLDL-L	3615	2.11	3291	2.49
S-VLDL-PL	3627	1.79	3301	2.19
S-VLDL-C	3627	1.79	3303	2.13
S-VLDL-CE	3627	1.79	3309	1.96
S-VLDL-FC	3617	2.06	3300	2.22
S-VLDL-TG	3606	2.36	3271	3.08
XS-VLDL-P	3623	1.90	3316	1.75
XS-VLDL-L	3627	1.79	3315	1.78
XS-VLDL-PL	3627	1.79	3311	1.90
XS-VLDL-C	3632	1.65	3305	2.07
XS-VLDL-CE	3634	1.60	3308	1.99
XS-VLDL-FC	3626	1.81	3308	1.99
XS-VLDL-TG	3614	2.14	3295	2.37
IDL-P	3628	1.76	3312	1.87
IDL-L	3629	1.73	3310	1.93
IDL-PL	3628	1.76	3307	2.01
IDL-C	3628	1.76	3311	1.90
IDL-CE	3630	1.71	3310	1.93
IDL-FC	3624	1.87	3318	1.69
IDL-TG	3619	2.00	3303	2.13
L-LDL-P	3630	1.71	3307	2.01
L-LDL-L	3630	1.71	3308	1.99

L-LDL-PL	3632	1.65	3304	2.10
L-LDL-C	3627	1.79	3307	2.01
L-LDL-CE	3628	1.76	3306	2.04
L-LDL-FC	3622	1.92	3315	1.78
L-LDL-TG	3621	1.95	3310	1.93
M-LDL-P	3629	1.73	3307	2.01
M-LDL-L	3628	1.76	3310	1.93
M-LDL-PL	3634	1.60	3307	2.01
M-LDL-C	3628	1.76	3308	1.99
M-LDL-CE	3628	1.76	3304	2.10
M-LDL-FC	3634	1.60	3307	2.01
M-LDL-TG	3621	1.95	3310	1.93
S-LDL-P	3631	1.68	3313	1.84
S-LDL-L	3636	1.54	3306	2.04
S-LDL-PL	3635	1.57	3309	1.96
S-LDL-C	3631	1.68	3310	1.93
S-LDL-CE	3623	1.90	3308	1.99
S-LDL-FC	3630	1.71	3311	1.90
S-LDL-TG	3628	1.76	3303	2.13
XL-HDL-P	3646	1.27	3340	1.04
XL-HDL-L	3645	1.30	3343	0.95
XL-HDL-PL	3649	1.19	3345	0.89
XL-HDL-C	3640	1.44	3340	1.04
XL-HDL-CE	3646	1.27	3344	0.92
XL-HDL-FC	3633	1.62	3338	1.10
XL-HDL-TG	3642	1.38	3323	1.54
L-HDL-P	3646	1.27	3340	1.04
L-HDL-L	3645	1.30	3340	1.04
L-HDL-PL	3648	1.22	3342	0.98
L-HDL-C	3644	1.33	3342	0.98
L-HDL-CE	3647	1.25	3342	0.98
L-HDL-FC	3647	1.25	3339	1.07
L-HDL-TG	3638	1.49	3325	1.48
M-HDL-P	3645	1.30	3320	1.63
M-HDL-L	3642	1.38	3321	1.60
M-HDL-PL	3645	1.30	3312	1.87
M-HDL-C	3640	1.44	3332	1.27
M-HDL-CE	3638	1.49	3330	1.33
M-HDL-FC	3643	1.35	3329	1.36
M-HDL-TG	3634	1.60	3315	1.78
S-HDL-P	3635	1.57	3324	1.51
S-HDL-L	3633	1.62	3324	1.51
S-HDL-PL	3635	1.57	3307	2.01
S-HDL-C	3625	1.84	3307	2.01
S-HDL-CE	3618	2.03	3317	1.72
S-HDL-FC	3636	1.54	3310	1.93
S-HDL-TG	3618	2.03	3311	1.90
VLDL_D	3644	1.33	3326	1.45
LDL_D	3625	1.84	3234	4.18
HDL_D	3668	0.68	3354	0.62
Serum_C	3638	1.49	3309	1.96
VLDL_C	3626	1.81	3292	2.46
Remnant_C	3626	1.81	3314	1.81
LDL_C	3625	1.84	3304	2.10
HDL_C	3643	1.35	3340	1.04
HDL2_C	3650	1.16	3338	1.10
HDL3_C	3631	1.68	3323	1.54
EstC	3640	1.44	3311	1.90
FreeC	3640	1.44	3306	2.04
Serum_TG	3611	2.22	3278	2.87

VLDL-TG	3601	2.49	3266	3.23
LDL-TG	3621	1.95	3308	1.99
HDL-TG	3618	2.03	3319	1.66
TotPG	3642	1.38	3326	1.45
PC	3648	1.22	3311	1.90
SM	3641	1.41	3312	1.87
TotCho	3637	1.52	3320	1.63
ApoA1	3642	1.38	3328	1.39
ApoB	3626	1.81	3312	1.87
TotFA	3630	1.71	3311	1.90
DHA	3602	2.46	3303	2.13
LA	3638	1.49	3323	1.54
FAw3	3616	2.09	3305	2.07
FAw6	3639	1.46	3314	1.81
PUFA	3640	1.44	3308	1.99
MUFA	3611	2.22	3298	2.28
SFA	3622	1.92	3302	2.16
Lac	3516	4.79	3244	3.88
Pyr	3578	3.11	3288	2.58
Cit	3641	1.41	3302	2.16
Glol	3624	1.87	3293	2.43
Ala	3640	1.44	3338	1.10
Gln	3642	1.38	3320	1.63
Gly	3629	1.73	3328	1.39
His	3634	1.60	3324	1.51
Ile	3609	2.27	3284	2.70
Leu	3635	1.57	3309	1.96
Val	3621	1.95	3319	1.66
Phe	3640	1.44	3328	1.39
Tyr	3649	1.19	3320	1.63
Ace	3594	2.68	3280	2.81
bOHBut	3519	4.71	3184	5.66
Crea	3654	1.06	3329	1.36
Alb	3472	5.98	3158	6.43
Gp	3630	1.71	3296	2.34

Supplementary Table 3: Regions of high linkage disequilibrium (LD) excluded from PCA of genetic data.

Chromosome	Start (bp)	End (bp)
1	48000000	52000000
2	86000000	100500000
	134500000	138000000
	183000000	190000000
3	47500000	50000000
	83500000	87000000
	89000000	97500000
5	44500000	50500000
	98000000	100500000
	129000000	132000000
	135500000	138500000
	25500000	33500000
6	57000000	64000000
	140000000	142500000
7	55000000	66000000
8	8000000	12000000
	43000000	50000000
	112000000	115000000
10	37000000	43000000
11	87500000	90500000
	46000000	57000000
12	33000000	40000000
	109500000	112000000
20	32000000	34500000

BP: Base pair

Supplementary Table 4: Absolute deviations in λ from 1 in each ethnicity.

MAF cut off	White European	South Asian
MAF <0.001	0.184	0.047
0.001≤ MAF <0.005	0.071	0.028
0.005≤ MAF <0.01	0.070	0.030
0.01≤ MAF <0.05	0.029	0.017
0.05≤ MAF <0.10	0.020	0.026
0.10≤ MAF <0.15	0.023	0.018
0.15≤ MAF <0.20	0.011	0.010
0.20≤ MAF <0.25	0.011	0.010
0.25≤ MAF <0.30	0.015	0.012
0.30≤ MAF <0.35	0.016	0.014
0.35≤ MAF <0.40	0.022	0.014
0.40≤ MAF <0.45	0.026	0.020
0.45≤ MAF <0.50	0.022	0.023

Average differences of λ values from 1 in each ethnicity averaged across 146 metabolite values. MAF: Minor Allele Frequency.

Supplementary Table 5: Number of SNPs identified in each ethnicity

Metabolite class	Metabolite	White European		South Asian	
		Suggestive (significant)	Thinned	Suggestive (significant)	Thinned
XXL-VLDL	XXL-VLDL-P	25	4	1	1
	XXL-VLDL-L	55	9	29	1
	XXL-VLDL-PL	56	10	52	2
	XXL-VLDL-C	60	11	3	1
	XXL-VLDL-CE	64	6	6	2
	XXL-VLDL-FC	34	11	24	2
	XXL-VLDL-TG	65	11	24	1
XL-VLDL	XL-VLDL-P	44	7	53	5
	XL-VLDL-L	42	10	41	2
	XL-VLDL-PL	31	9	54	3
	XL-VLDL-C	31	6	13	3
	XL-VLDL-CE	26	7	17	4
	XL-VLDL-FC	38	8	24	4
	XL-VLDL-TG	42	10	50	3
L-VLDL	L-VLDL-P	11	3	58	4
	L-VLDL-L	15	4	69	5
	L-VLDL-PL	17	5	68	4
	L-VLDL-C	21	7	79	3
	L-VLDL-CE	17	5	27	4
	L-VLDL-FC	17	6	39	4
	L-VLDL-TG	24	5	43	5
M-VLDL	M-VLDL-P	26	5	53	1
	M-VLDL-L	222	47	102	6
	M-VLDL-PL	24	4	42	3
	M-VLDL-C	35	9	26	1
	M-VLDL-CE	16	7	89	8
	M-VLDL-FC	226	35	85	3
	M-VLDL-TG	254	33	91	5
S-VLDL	S-VLDL-P	52	5	24	3
	S-VLDL-L	47	4	25	1
	S-VLDL-PL	48	4	32	1
	S-VLDL-C	35	5	2	0
	S-VLDL-CE	11	5	7	3
	S-VLDL-FC	49	4	33	2
	S-VLDL-TG	12	6	193	24
XS-VLDL	XS-VLDL-P	10	8	6	1
	XS-VLDL-L	16	8	9	1
	XS-VLDL-PL	21	8	7	3
	XS-VLDL-C	14	5	2	2
	XS-VLDL-CE	23	9	1	1
	XS-VLDL-FC	18	9	8	2
	XS-VLDL-TG	61	8	59	5
IDL	IDL-P	7	4	13	3
	IDL-L	6	3	11	2
	IDL-PL	8	3	0	0
	IDL-C	5	3	1	1
	IDL-CE	10	6	7	1
	IDL-FC	5	1	0	0
	IDL-TG	28	10	2	1
L-LDL	L-LDL-P	7	2	6	1

	L-LDL-L	8	2	9	2
	L-LDL-PL	12	2	8	2
	L-LDL-C	10	3	8	1
	L-LDL-CE	13	3	8	1
	L-LDL-FC	4	1	0	0
	L-LDL-TG	30	8	8	2
M-LDL	M-LDL-P	5	3	14	2
	M-LDL-L	6	2	14	2
	M-LDL-PL	6	1	7	1
	M-LDL-C	14	3	8	1
	M-LDL-CE	14	3	0	0
	M-LDL-FC	8	3	9	2
	M-LDL-TG	37	8	9	3
S-LDL	S-LDL-P	9	4	14	2
	S-LDL-L	11	3	9	2
	S-LDL-PL	33	5	1	1
	S-LDL-C	13	4	9	2
	S-LDL-CE	19	4	0	0
	S-LDL-FC	21	4	13	4
	S-LDL-TG	40	14	11	3
XL-HDL	XL-HDL-P	76	10	8	2
	XL-HDL-L	84	11	8	2
	XL-HDL-PL	48	10	10	2
	XL-HDL-C	52	13	5	1
	XL-HDL-CE	54	9	6	1
	XL-HDL-FC	41	9	12	3
	XL-HDL-TG	16	7	37	5
L-HDL	L-HDL-P	57	10	85	4
	L-HDL-L	57	12	83	4
	L-HDL-PL	63	6	61	4
	L-HDL-C	53	11	75	4
	L-HDL-CE	52	11	8	4
	L-HDL-FC	61	10	8	4
	L-HDL-TG	32	10	7	3
M HDL	M-HDL-P	32	11	7	3
	M-HDL-L	53	7	10	2
	M-HDL-PL	64	9	17	3
	M-HDL-C	49	13	2	1
	M-HDL-CE	36	13	6	2
	M-HDL-FC	74	12	13	3
	M-HDL-TG	20	6	1	1
S-HDL	S-HDL-P	56	8	1	1
	S-HDL-L	81	9	1	1
	S-HDL-PL	140	10	19	1
	S-HDL-C	19	4	3	2
	S-HDL-CE	12	4	9	3
	S-HDL-FC	76	9	3	2
	S-HDL-TG	57	7	18	2
Lipoprotein Density	VLDL_D	39	7	20	3
	LDL_D	20	6	46	5
	HDL_D	74	12	31	4
Cholesterol	Serum_C	62	4	19	3
	VLDL_C	23	3	0	0
	Remnant_C	39	8	11	5
	LDL_C	7	3	0	0
	HDL_C	58	8	63	4
	HDL2_C	67	9	31	6
	HDL3_C	63	6	12	3
	EstC	111	4	17	3
	FreeC	23	3	9	2

Triglycerides	Serum_TG	19	4	34	3
	VLDL-TG	17	7	50	3
	LDL-TG	22	8	8	1
	HDL-TG	49	11	49	2
	TotPG	56	7	23	4
	PC	69	10	20	2
	SM	117	6	24	4
	TotCho	43	7	17	2
Apolipoproteins	ApoA1	58	10	32	6
	ApoB	33	6	13	4
Fatty Acids	TotFA	40	3	8	3
	DHA	33	5	4	1
	LA	17	2	5	2
	FAw3	69	2	2	1
	FAw6	21	1	15	3
	PUFA	26	1	20	4
	MUFA	47	5	28	1
	SFA	32	2	4	2
Glycolysis Related Metabolites	Lac	63	5	1	1
	Pyr	63	4	6	3
	Cit	156	12	32	6
	Glol	29	7	3	2
Unbranched Amino Acids	Ala	156	14	53	2
	Gln	5	2	2	2
	Gly	15	8	9	2
	His	127	7	9	5
Branched Chain Amino Acids	Ile	26	6	0	0
	Leu	32	7	1	1
	Val	6	4	8	3
Aromatic Amino Acids	Phe	5	3	59	4
	Tyr	11	5	23 (15)	2
Ketone Bodies	Ace	27	5	49 (1)	4
	bOHBut	23	6	0	0
Fluid Balance and Inflammation	Crea	108	13	52	6
	Alb	15	1	1	1
	Gp	22	5	2	2

SNPs were classified as being in LD if their R^2 value exceeded 0.2. Suggestive level: p value≤ 1 x 10^{-5}. Numbers in brackets represent the number of SNPs associated at the genome wide significant level (p value≤ 5 x 10^{-8}).

Supplementary Table 6: Investigation of pleiotropy in the GWAS Catalog and Phenoscanner databases.

A

Metabolite	SNP	GWAS Catalog	Phenoscanner	Gene
Leucine	rs11586886	-	-	-
	rs11241582	-	-	-
	rs2984433	-	BMI, Obesity class 1, weight,	ACTG1P9
	rs225598	-	-	-
	rs11270713	-	-	-
	rs73404871	-	Mean corpuscular haemoglobin, mean corpuscular volume	TRIM56
	rs879918	-	Self-reported Hepatitis A	-
HDL_D	rs13007852	-	-	-
	rs2320799	-	-	-
	rs13187167	-	-	-
	rs76030404	-	-	-
	rs153744	-	-	-
	rs55768285	-	Height, sitting height, comparative height size at age 10, trunk fat-free mass, whole body fat free mass, arm predicted mass (left and right), whole water body mass, arm fat free mass (left)	RP11-207F8.1
	rs10899584	-	-	-
	rs996382	-	-	-
	rs55890848	-	-	-
	rs35251956	-	Other malignant neoplasms of skin, mean corpuscular volume, self- reported malignant melanoma, red blood cell count, mean platelet volume, self-reported basal cell carcinoma	CPNE7
	rs12456654	-	-	-

	rs34733998	-	-	-
	rs16999687	-	-	-
HDLC	rs17525600	-	Stomatitis and related lesions	AC007682.1
	rs2555761	-	-	-
	rs10109848	-	-	-
	rs10219564	-	-	-
	rs7490538	-	-	-
	rs59137888	-	-	-
	rs16964949	-	-	-
	rs34733998	-	-	-
HDL2C	rs77662164	-	-	-
	rs2555761	-	-	-
	rs13200375	-	-	-
	rs10109848	-	-	-
	rs10219564	-	-	-
	rs7490538	-	-	-
	rs59137888	-	-	-
	rs16964949	-	-	-
	rs34733998	-	-	-
HDL3C	rs1104779	-	-	-
	rs6443637	-	-	-
	rs7490538	-	-	-
	rs7321144	-	-	-
	rs4555250	-	-	-
	rs34733998	-	-	-
XS-VLDL-TG	rs11685644	-	-	-
	rs17168796	-	-	-
	rs77423873	-	-	-
	rs1573510	-	Height	OR5AL2P
	rs34447547	Total PHF- tau (SNP x SNP interaction)	-	-
	rs4781176	-	Qualifications: college or university degree	

	rs35085155	-	Hand grip strength left	AP000472.3
	rs28385583	-	-	-
S-LDL-P	rs10915339	-	-	-
	rs77266229	-	-	-
	rs10908948	Age at menarche	Age at menarche	UNQ6494
	rs35710612	-	-	-
XL-HDL-CE	rs7766216	-	-	-
	rs2555761	-	-	-
	rs10109848	-	-	-
	rs10219564	-	-	-
	rs9510560	-	-	-
	rs7321144	-	-	-
	rs55890848	-	-	-
	rs10402413	-	-	-
	rs6126252	-	-	-
	rs35380742	-	-	-
L-HDL-P	rs77662164	-	-	-
	rs2555761	-	-	-
	rs10109848	-	-	-
	rs10219564	-	-	-
	rs9510560	-	-	-
	rs7321144	-	-	-
	rs55890848	-	-	-
	rs10402413	-	-	-
	rs6126252	-	-	-
	rs35380742	-	-	-
L-HDL_L	rs17525600	-	Stomatitis and related lesions	AC007682.1
	rs6811162	-	Self-reported hypertension. Vascular or heart problems diagnosed by doctor: high blood pressure	ENPEP
	rs77662164	-	-	-
	rs13200375	-	-	-
	rs2555761	-	-	-

	rs10109848	-	-	-
	rs10219564	-	-	-
	rs9510560	-	-	-
	rs7321144	-	-	-
	rs35163069	-	-	-
	rs6126252	-	-	-
	rs35380742	-	-	-
L-HDL-C	**rs6811162**	-	Self-reported hypertension. Vascular or heart problems diagnosed by doctor: high blood pressure	ENPEP
	rs6834601	-	-	-
	rs77662164	-	-	-
	rs13200375	-	-	-
	rs55768285	Waist circumference adjusted for BMI	Height, sitting height, comparative height size at age 10, Trunk fat-free mass, whole body fat free mass, arm predicted mass (left and right), whole water body mass, arm fat free mass (left)	LINC01621, ELOVL4
	rs2555761	-	-	-
	rs10109848	-	-	-
	rs10219564	-	-	-
	rs2028592	-	-	-
	rs35163069	-	-	-
	rs35380742	-	-	-
M-HDL-C	rs6663801	-	Comparative height size at age 10	RP5-855F14.2
	rs10033924	-	-	-
	rs79991518	-	-	-
	rs13171149	-	-	-
	rs62443510	-	-	-
	rs62445582	-	-	-
	rs76303188	-	-	-
	rs17073913	-	-	-
	rs10085955	-	-	-

	rs904558	-	-	-
	rs12578234	-	-	-
	rs7398018	-	-	-
	rs16964930	-	-	-
M-HDL-CE	rs6663801	-	Comparative height size at age 10	RP5-855F14.2
	rs10033924	-	-	-
	rs79991518	-	-	-
	rs62445582	-	-	-
	rs76303188	-	-	-
	rs17073913	-	-	-
	rs10085955	-	-	-
	rs17346889	-	-	-
	rs11593054	-	-	-
	rs904558	-	-	-
	rs12578234	-	-	-
	rs2138011	-	-	-
	rs7398018	-	-	-
S-HDL-CE	rs11206525	-	-	-
	rs1881817	-	-	-
	rs12769447	-	Lymphocyte count	C10orf128
	rs9795921	-	-	-

B

Metabolite	SNP	GWAS Catalog	Phenoscanner	Gene
LA	**rs12720820**	-	Self-reported high cholesterol, coronary artery disease, treatment with cholesterol lowering medication	APOB
	rs721632	-	-	-
FAw3	rs7040631	-	-	
FAw6	**rs12720820**	-	Self-reported high cholesterol, coronary artery disease, treatment with cholesterol lowering medication	
	rs11683770	-	mDC:%32+; mDC subset (CD32+)	
	rs58865405	-	-	
M-VLDL-L	rs17028714	-	-	
	rs62294143	-	-	
	rs7856692	-	-	
	rs11597600	-	-	
	rs7224672	-	-	
	rs73481716	-	-	
IDL-C	rs41286967	-	-	-
IDL-CE	rs41286967	-	-	-
L-LDL-P	rs41286967	-	-	-
S-LDL-PL	rs41286968	-	-	-
L-HDL-PL	rs3814329	-	Sitting height	RGL1
	rs10207578	-	-	-
	rs6922	-	-	-
	rs7486176	-	Systolic blood pressure, vascular or heart problems diagnosed by doctor: high blood pressure, self-reported hypertension	C12orf76
S-HDL-L	rs6490057	-	-	-
S-HDL-C	rs6679531	-	-	-
	rs9938230	-	-	-

Bolded SNPs indicate potential confounders. **A**: Individual metabolites in white Europeans. **B**: Individual metabolites in South Asians.

Supplementary Table 7: Percentage of variation explained by PC1 in each metabolite class.

Metabolite class	White Europeans	South Asians
XS-VLDL	88	91.5
S-VLDL	96.8	91.8
M-VLDL	98.9	95.6
L-VLDL	98.5	99.0
XL-VLDL	99.1	96.7
XXL-VLDL	96.7	95.1
All VLDL classes	91.9	86
IDL	85.8	92.1
S-LDL	98.0	97.5
M-LDL	95.5	95.3
L-LDL	92.5	91.7
All LDL classes	95.0	94.5
S-HDL	57.3	56
M-HDL	92.1	97.3
L-HDL	97.0	91.9
XL-HDL	95.2	97.8
All HDL classes	49.2	97.8
Glycolysis Related Metabolites	63.9	51.7
Fatty Acids	82.7	87.4
Lipoprotein Density	80.9	72.2
Cholesterol	69.7	59.7
Glycerides and Phospholipids	79.1	81.5
Branched Amino Acids	78.8	79.3
Unbranched Amino Acids	48.3	84.8
All Amino Acids	45.8	66

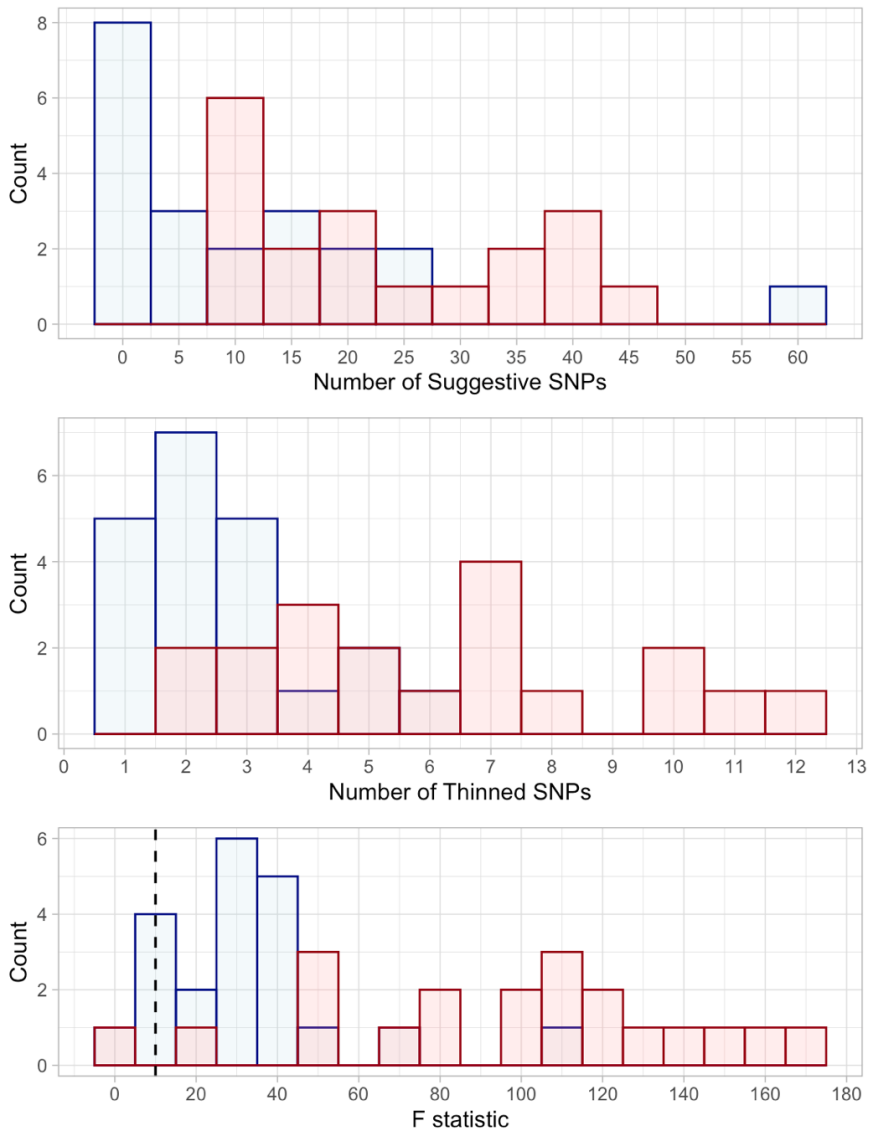
Percentage of variation explained by PC1 in the PCA analysis of each metabolite class following the removal of extreme outliers (3 x IQR). GWAS were performed on all classes with ≥ 2 metabolites where PC1 explained ≥70% of the variation.

Supplementary Table 8: Correlation between PC1 and PC2 following outlier removal.

	PC1		PC2	
Metabolite Class	**White Europeans**	**South Asians**	**White Europeans**	**South Asians**
XXL-VLDL	1	1	0.99	0.97
XL-VLDL	1	1	0.99	0.59
L-VLDL	1	1	-0.39	0.89
M-VLDL	1	1	0.99	0.98
S-VLDL	-0.99	-0.99	0.99	0.97
XS-VLDL	1	1	1	1
All VLDL	1	1	1	1
IDL	-0.83	1	-1	-0.63
L-LDL	-0.79	1	0.99	0.53
M-LDL	1	1	1	-1
S-LDL	1	1	-0.37	0.98
XL-HDL	1	1	1	1
L-HDL	1	-0.79	1	1
M-HDL	1	1	1	0.99
S-HDL	1	1	1	1
All HDL	0.94	0.94	1	0.41
Lipoprotein Density	1	1	1	1
Cholesterols	1	-0.95	1	0.95
Glycerides and Phospholipids	1	0.97	1	0.84
Fatty Acids	1	1	0.99	1
Branched Amino Acids	1	1	1	1
Unbranched Amino Acids	1	1	1	1

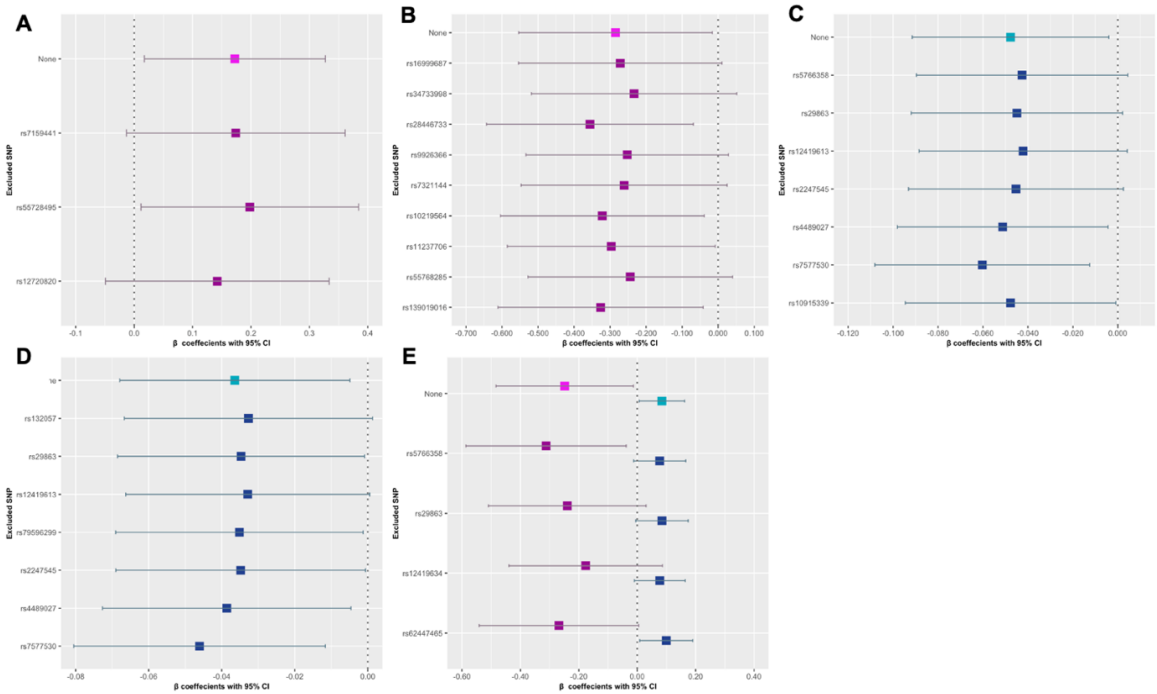
Correlation between PC1 and PC2 following the removal of outliers (1.5 x IQR) and extreme outliers (3 x 1QR) in metabolite classes where PC1 explained ≥ 70 the variation in the metabolite class in at least one ethnicity.

Supplementary Figure 13: Histograms of the number of SNPs and strength of each instrument in the analysis of each metabolite class.



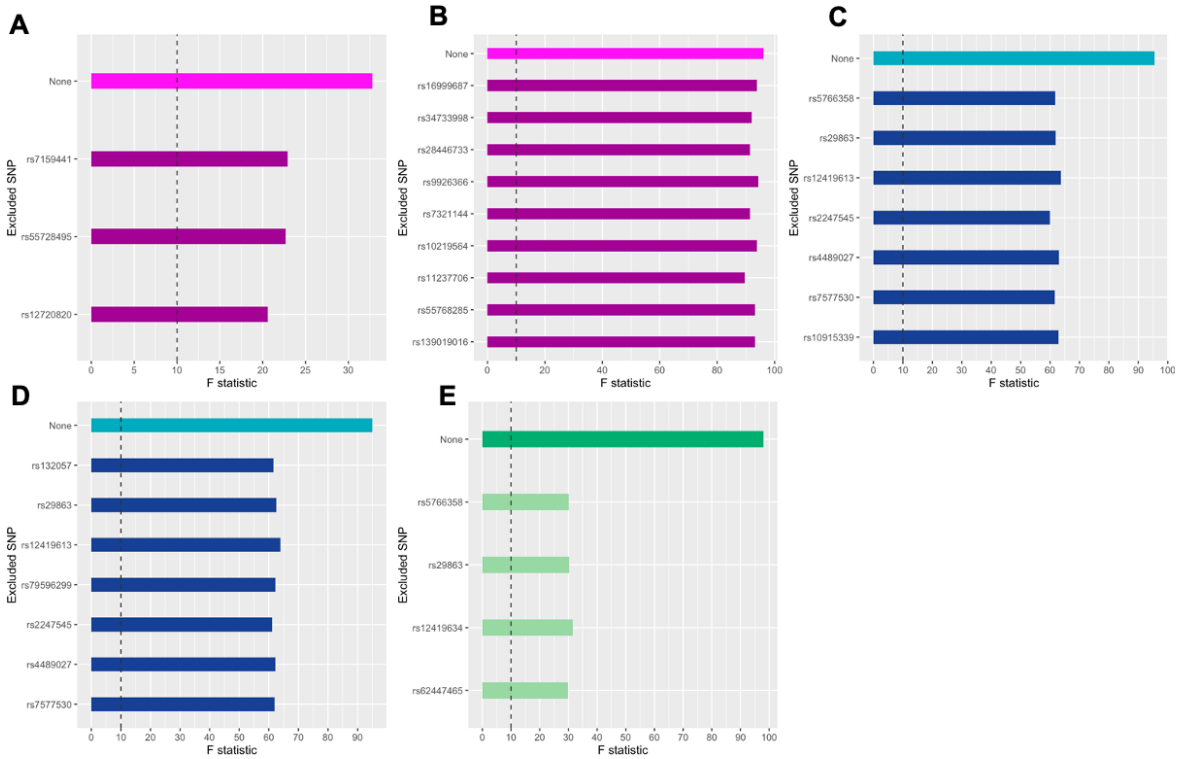
A: Histogram of number of SNPs identified for each metabolite class at the suggestive level ($p \leq 1 \times 10^{-5}$). **B:** Histogram of the number of SNPs remaining after thinning by LD (R^2>0.2).**C**: Histogram of the F-statistics for each instrument. Dashed line shows an F statistic of 10. An F statistic < 10 is an indicator of weak instrument bias. Blue: South Asians. Red: white Europeans.

Supplementary Figure 14: Forest plots of β values following leave-one-out analyses for identified associations in the analysis of metabolite class.



Forest plots showing β values and 95% CIs following the leave-one-out analyses of each SNP in each instrument. Dashed line represents no effect. Associations shown in purple indicate associations with 2-hour post glucose while associations shown in blue indicate associations with fasting glucose. **A:** Fatty acid class, SAs. **B:** XL-HDL class, WEs. **C:** M-LDL class, WEs. **D:** All LDL class, WEs. **E:** S-LDL class, WEs.

Supplementary Figure 15: F-statistics following leave-one-out analyses for identified associations in the analysis of metabolite classes.



Dashed line indicates an F statistic of 10, below which an instrument is classified as weak. Associations shown in purple indicating associations with 2-hour post glucose while associations shown in blue indicate associations with fasting glucose. Green bars indicate metabolite measures associated with both fasting glucose and 2-hour post glucose. **A:** Fatty acid class, SAs. **B:** XL-HDL class, WEs. **C:** M-LDL class, WEs. **D:** All LDL class, WEs. **E:** S-LDL class, WEs.

Supplementary Table 9: Investigation of pleiotropy in the GWAS Catalog and Phenoscanner databases for identified metabolite classes.

Ethnicity	Metabolite group	SNP	GWAS Catalog	Phenoscanner	Gene
White Europeans	XL-HDL	rs16999687	-	-	-
		rs34733998	-	-	-
		rs28446733	-	-	-
		rs9926366	-	-	-
		rs7321144	-	-	-
		rs10219564	-	-	-
		rs11237706	-	-	-
		rs55768285	Waist circumference adjusted for BMI	Height, sitting height, comparative height at age 10, trunk fat-free mass, trunk predicted mass, whole body fat-free mass, arm predicted mass (left and right), whole body water mass, arm fat-free mass (left), forced vital capacity	LINC01621, ELOVL4
		rs139019016	-	-	-
	M-LDL	rs5766358	-	-	-
		rs29863	-	-	-
		rs12419613	-	-	-
		rs2247545	-	-	-
		rs4489027	-	-	-
		rs7577530	-	-	-
		rs10915339	-	-	-
	S-LDL	rs5766358	-	-	-
		rs29863	-	-	
		rs12419634	Primary biliary cholangitis	-	POU2AF1
		rs62447465		treatment with Berroca effervescent tablet	CDCA7L
	All LDL classes	rs132057	-	Tinnitus severity or nuisance	-
		rs29863	-	-	-
		rs12419613	-	-	-
		rs79596299	-	-	-
		rs2247545	-	-	-
		rs4489027	-	-	-

		rs7577530	-	-	-
South Asians	Fatty Acids	rs7159441	-	-	-
		rs55728495	-	-	-
		rs12720820	-	Self-reported high cholesterol	APOB

Supplementary Table 10: *Post-hoc* power analysis.

A

							Power (α = 0.05)			Power (α = 0.01)	
Analysis	**Group**	**True β**	**Obs**	**R^2**	**Variance of exposure**	**Variance of Outcome**	**β from Unadjusted linear regression**	**β from MR**	**β from MR in significant ethnicity**	**β from MR**	**β from MR in significant ethnicity**
FAw3	SA	0.00662	-0.00646	0.00552	0.0257	0.0130	0.05	-	-	-	-
Fasting	WE	0.00327	0.0368	0.0139	0.0260	0.00819	0.05	0.98	0.99	0.93	0.97
FAw6	SA	-0.000584	-0.0121	0.0176	0.0322	0.0623	0.05	-	-	-	-
2-hour	WE	0.00341	0.0303	0.00633	0.0261	0.0566	0.05	0.07	0.25	-	-
LA	SA	-0.00128	-0.00948	0.0113	0.0292	0.0624	0.05	-	-	-	-
2-hour	WE	0.00335	0.0231	0.0125	0.0264	0.0567	0.05	0.34	0.47	-	-
M-VLDL-L	SA	-0.00333	-0.00212	0.0363	1.34	0.0617	0.05	-	-	-	-
2-hour	WE	0.00540	0.00623	0.237	0.457	0.0564	0.07	0.95	0.35	0.85	0.16
IDL-C	SA	-0.00365	-0.0772	0.00556	0.0129	0.0623	0.05	-	-	-	-
2-hour	WE	0.0223	0.0646	0.0191	0.0127	0.0564	0.05	0.11	0.97	0.03	0.9
IDL-CE	SA	-0.00615	-0.0966	0.00571	0.00925	0.0623	0.05	-	-	-	-
2-hour	WE	0.0315	0.0767	0.0311	0.00908	0.0566	0.05	0.05	1	0.01	0.99
L-LDL-P	SA	5.51E+03	22.9	0.0220	3.27E-08	0.0625	1	-	-	-	-
2-hour	WE	4.76E+03	17.3	0.0605	3.46E-08	0.0565	1	1	1	1	1
S-LDL-PL	SA	-0.0179	-0.183	0.00549	0.00130	0.0624	0.05	-	-	-	-
2-hour	WE	0.145	0.221	0.03193	0.00127	0.0566	0.06	0.38	1	0.18	1
L-HDL-PL	SA	0.0187	0.0372	0.0227	0.00867	0.0624	0.05	-	-	-	-

2-hour	WE	0.0163	0.0311	0.0408	0.00912	0.0566	0.05	0.33	0.88	0.15	0.71
S-HDL-L	SA	0.00781	0.0105	0.00599	0.0116	0.0624	0.05	-	-	-	-
2-hour	WE	-0.00316	-0.0415	0.0546	0.0115	0.0565	0.05	0.24	1	0.09	1
S-HDL-C	SA	-0.0126	0.00359	0.0176	0.00166	0.0623	0.05	-	-	-	-
2-hour	WE	0.0370	0.0419	0.0245	0.00164	0.0566	0.05	0.05	0.33	-	-

B

							Power (α = 0.05)			**Power (α = 0.01)**	
Analysis	**Group**	**True β**	**Obs**	**R^2**	**Variance of exposure**	**Variance of Outcome**	**β from Unadjusted linear regression**	**β from MR**	**β from MR in significant ethnicity**	**β from MR**	**β from MR in significant ethnicity**
Fatty Acids	SA	0.000693	-0.000612	0.0175	6.86	0.01308	0.047	-	-	-	-
2-hour	WE	0.00380	0.00650	0.0518	1.19	0.00814	0.028	1	1	1	1

C

							Power (α = 0.05)			Power (α = 0.01)	
Analysis	**Group**	**True β**	**Obs**	**R^2**	**Variance of exposure**	**Variance of Outcome**	**β from Unadjusted linear regression**	**β from MR**	**β from MR in significant ethnicity**	**β from MR**	**β from MR in significant ethnicity**
Leucine	WE	-0.129	-0.0220	0.0428	0.0161	0.00813	1	-	-	-	-
	SA	-0.0672	-0.0117	0.00195	0.0170	0.0131	0.05	0.04	0.08	-	-
Fasting	WE	-0.00698	-0.0143	0.0428	0.0161	0.0566	1	-	-	-	-
2-hour	SA	-0.0881	-0.0334	0.00195	0.0170	0.0621	0.05	0.3	0.09	-	-
HDL_D	WE	-0.00144	0.00367	0.0747	0.0415	0.00816	0.05	-	-	-	-
	SA	0.00591	0.00626	0.0229	0.0395	0.0131	0.05	0.05	0.25	-	-
Fasting	WE	-0.00133	0.0211	0.0747	0.0415	0.0565	0.05	-	-	-	-
2-hour	SA	0.0117	0.00682	0.0229	0.0395	0.0621	0.05	0.08	0.28	-	-
HDLC	WE	0.00654	0.0140	0.0515	0.0155	0.0565	0.05	-	-	-	-
2-hour	SA	0.00592	0.0223	0.0171	0.0148	0.06241	0.05	0.11	0.2	-	-
HDL2C	WE	0.00636	0.0142	0.0584	0.0183	0.0566	0.05	-	-	-	-
2-hour	SA	0.00617	0.00823	0.0200	0.0966	0.0623	0.05	0.08	0.82	0.02	0.62
HDL3C	WE	0.0931	0.183	0.0381	0.000698	0.0566	0.05	-	-	-	-
2-hour	SA	0.0598	0.0676	0.0676	0.000658	0.0622	0.05	0.92	0.71	0.79	0.47
XS-VLDL-TG	WE	0.0625	0.0984	0.0501	0.00242	0.0564	0.05	-	-	-	-
2-hour	SA	0.0230	-0.0335	0.0296	0.00244	0.0619	0.05	0.33	0.4	-	-
S-LDL-P	WE	8.31E+04	1.42E+02	0.0243	3.01E-09	0.0565	1	-	-	-	-
2-hour	SA	-1.74E+04	-139.5	0.0114	3.06E-09	0.0624	1	0.05	0.05	-	-
XL-HDL-CE	WE	0.0525	0.0745	0.0603	0.00506	0.0567	0.06	-	-	-	-
2-hour	SA	0.0314	-0.0102	0.00600	0.00477	0.0618	0.05	0.45	0.11	-	-
L-HDL-P	WE	4.76E+03	17.3	0.0605	3.46E-08	0.0565	1	-	-	-	-
2-hour	SA	5.51E+03	22.9	0.0220	3.27E-08	0.0625	1	0.31	0.29	-	-
L-HDL-PL	WE	0.00732	0.0225	0.0722	0.0226	0.0565	0.05	-	-	-	-
2-hour	SA	0.00857	0.0297	0.0218	0.0214	0.0625	0.05	0.31	0.28	-	-
L-HDL-C	WE	0.0129	0.0214	0.0669	0.0138	0.0565	0.05	-	-	-	-

2-hour	SA	0.00149	0.0296	0.0223	0.0130	0.0623	0.05	0.2	0.21	-	-
M-HDL-C	WE	-0.0357	-0.0592	0.06553	0.00503	0.00814	0.07	-	-	-	-
2-hour	SA	-0.0184	-0.0253	0.00614	0.00991	0.01304	0.05	0.15	0.12	-	-
M-HDL-CE	WE	-0.0450	-0.0680	0.0362	0.00398	0.00814	0.06	-	-	-	-
2-hour	SA	-0.0240	-0.0314	0.0114	0.00629	0.0130	0.05	0.16	0.3	-	-
S-HDL-CE	WE	0.0370	0.0419	0.0245	0.00164	0.0566	0.05	-	-	-	-
2-hour	SA	-0.01257	0.00359	0.01757	0.00166	0.06233	0.05	0.08	0.6	-	-

D

							Power (α = 0.05)			**Power (α = 0.01)**	
Metabolite Class	**Group**	**True β**	**Obs**	**R^2**	**Variance of exposure**	**Variance of Outcome**	**β from Unadjusted linear regression**	**β from MR**	**β from MR in significant ethnicity**	**β from MR**	**β from MR in significant ethnicity**
XL-HDL	WE	-0.00536	-0.00886	0.0585	0.4555	0.0565	0.06	-	-	-	-
2-hour	SA	0.00192	0.00120	0.0109	2.57	0.0623	0.05	0.45	1	0.23	1
S-LDL Fasting	WE	0.000358	-0.000274	0.0227	1.57	0.00814	0.05	-	-	-	-
	SA	0.000358	0.000672	0.0171	0.964	0.0131	0.05	0.92	1	0.78	7
2-hour	WE	-0.00363	-0.00694	0.0227	1.57	0.0565	0.05	-	-	-	-
	SA	-0.00215	0.00662	0.0171	0.964	0.0622	0.05	0.19	1	0.78	0.98
M-LDL	WE	-1.39E-04	0.000253	0.0409	2.69	0.00814	0.05	-	-	-	-
Fasting	SA	0.000672	0.000405	0.00572	2.25	0.0131	0.05	0.58	0.58	-	-

All LDL	WE	-1.26E-04	0.000219	0.0409	5.19	0.00814	0.05	-	-	-	-
Fasting	SA	-0.000492	-0.000285	0.00138	4.10	0.0131	0.05	0.05	0.06	-	-

Tables showing *post-hoc* MR calculation from mRND (https://shiny.cnsgenomics.com/mRnd/). Observational associations were obtained from linear regression models adjusted for maternal age (years), BMI (continuous), smoking status, multiple pregnancy, parity, and gestational age. Initial true β estimates were obtained from unadjusted linear regression models. Additional power analyses were performed in the non-significant ethnicity to determine the power to predict the β estimate obtained from the MR analyses and the power to detect the β estimate from the significant model in the alternative ethnicity. If the power from either analysis exceeded 80% then power was also calculated for α = 0.01. Obs: β from adjusted observational studies **A:** Individual metabolite analysis in South Asians. **B**: Analysis of metabolite classes in South Asians. **C:** Analysis of Individual metabolites in White Europeans. **D:** Analysis of metabolite classes in White Europeans.