Genome-wide association and mendelian randomisation study of metabolites and pregnancy dysglycemia 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. In WEs cholesterols were most common metabolite class identified, whereas 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 in order to inform ethnically tailored GDM prevention strategies.

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

GWAS analysis

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

modelling (median gestational age SA = 184 days, IQR= 182-186.7, median gestational age WE = 184 days, IQR= 182-187). 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 samples from BiB originate from a different region of Pakistan (the Mirpur Region) from the 1000 Genome SA samples [27, 28]. This is of particular importance in SA populations as even geographically close populations can have differing allele frequencies due to differing Biraderi ('Brotherhood') membership between population subgroups. Biraderi membership is assigned at birth, is an indicator of male lineage as well as social-occupational status which largely governs partner choice and can result in higher levels of consanguinity in the population [22]. PCA plots were created using the *ggplot2* package in R studio (version 4.0.2) [36, 37]. No clear separation in SA BiB samples and SA 1000 Genome (1000G) samples was identified indicating that LD estimates obtained from 1000G was suitable for use in BiB (**Supplementary Figures 2-4**).

Analysis

Genetic Risk Scores (GRS) were created in PLINK (version 1.9) for each metabolite with each SNP receiving a weight based on its estimated effect size on the metabolite obtained from the GWAS [38]. One-sample MR was then performed by Two-Stage Least Squares regression (TSLS; *ivpack*, ivreg, and *AER* packages in R version 4.0.2) to obtain a causal estimate for the effect of each metabolite value on the log-normalised continuous measures of fasting glucose and 2-hour post glucose following a 75g oral glucose tolerance test (OGTT) [36, 39, 40]. Here, the level of a metabolite is regressed on its respective GRS and, subsequently, the outcome is

regressed onto these fitted GRS-metabolite values in the second stage. All MR results have been reported according to STROBE-MR guidelines [41].

When significant associations were identified leave-one-out analysis was performed. For this, SNPs were removed sequentially from the instrument and changes to the effect estimate and F-statistic was assessed. If the exclusion of a SNP was found to alter either the effect estimates or F-statistic (through the visualisation of forest plots) it is possible that the SNP is influencing the outcome via an alternative pathway to other SNPs, potentially highlighting a violation of the 2nd or 3rd MR assumption. To further test for violations of these assumptions, included SNPs were searched for in both the Phenoscanner and GWAS Catalog databases to identify previously identified associations [42-44]. In both databases a p-value $\leq 1 \times 10^{-5}$ was interpreted as indicative of an additional association. Differences between MR and linear regression results were also evaluated via the Wu-Hausman statistic to assess deviation of the instrumental variable estimate from the ordinary least squares (OLS) estimate [45]. Deviations in these two measures can indicate either confounding in the OLS estimate (indicating a need for MR) or violations of the MR assumptions due to pleiotropy.

Post-hoc power analyses

For metabolites that associated with a measure of postprandial glucose in only one ethnic group, *post-hoc* power analyses were performed using the mRnd CNS genomics tool (https://shiny.cnsgenomics.com/mRnd/) to assess whether the absence of an association in the alternate ethnicity was due to limited power [46]. Observational and 'true' associations required by the tool were obtained by performing linear regression of the outcome on the metabolite and obtaining

unadjusted and adjusted estimates (adjusted for maternal age (years), BMI (continuous), smoking status, multiple pregnancy, parity, and gestational age) respectively. Due to the *post-hoc* nature of this analysis, additional power analyses could be conducted assuming the MR estimate to be the true causal effect in the MR calculation. This analysis was performed in the non-significant population for each metabolite associated in only one ethnicity. If power was found to be adequate (80%) at the 5% level ($\alpha = 0.05$) power was also assessed at the 1% level ($\alpha = 0.01$).

Results

GWAS of Metabolite Measures

A total of 6184 SNPs were associated with at least one metabolite in WEs at the suggestive level (1×10^{-5}), with 2616 (42.3%) SNPs being associated with a single metabolite measure. However, no SNPs were identified below the genome-wide significant level (p-value $<5 \times 10^{-8}$) in WEs.

Fewer SNPs were identified at the suggestive level in SAs, with 3685 SNPs SNP-metabolite associations in total, of which 1544 (41.9%) SNPs being associated with only one metabolite measure. SNP associations were identified for 138/146 (94.5%) metabolite exposures in SAs (**Supplementary Table 5**). No SNP was identified as being associated at the suggestive level in both ancestries, although shared genomic regions were identified between ancestries (**Supplemental Excel**).

To evaluate the possibility of shared genomic predictors of metabolites, a pooled meta-analysis of effect estimates in both ethnicities was performed. For 90 metabolite values, no associations were found to exceed the genome-wide

suggestive level ($p<10^{-5}$) following meta-analyses of both ethnicities. SNP associations were identified at the suggestive level for four metabolite measures (concentration of XL-HDL, total lipids in M-VLDL, mean density of VLDL and citrate) despite these differing in direction of effects in SAs and WEs. In addition, 4 SNPs were associated with alanine, despite these SNPs initially being associated with alanine only in the SA population. These SNPs (rs12256633, rs17121228, rs7096521, rs12240368) are all found on chromosome 10 in the receptor gene *SORCS1* and have not been associated with alanine levels previously [47, 48].

MR results

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

White Europeans

Two metabolite values, leucine and mean density of HDL lipoproteins (HDL_D), were associated with both fasting glucose and 2-hour post glucose (**Table 1**). Specifically, a 1mmol/L increase in blood leucine associated with lower fasting

glucose (-0.193 mmol/L, 95% CI -0.069, -0.319) and 2-hour post glucose (-0.443 mmol/L, 95% CI -0.113, -0.774). Likewise, a 1nm increase in mean diameter of HDL associated with lower fasting glucose (-0.082 mmol/L, 95% CI 0.026, 0.138) and 2-hour post glucose (-0.191, 95% CI 0.043, 0.339 mmol/L) **(Supplementary Figures 4-5)**. No other metabolites were associated with both measures of glucose in WEs.

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

Table 1: Significant MR results in white Europeans.

	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

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.

White Europeans: Sensitivity analyses

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

Three of the metabolites (leucine, L-HDL-L, L-HDL-C) that were associated with postprandial glucose in WEs included a SNP that previous studies have associated ($p \leq 1 \times 10^{-5}$) with at least one potential confounder (BMI, hypertension or waist circumference) **(Supplementary Table 7**). The removal of these SNPs from the instrument did not impact the significance of the associations identified for leucine or L-HDL-L **(Table 2)**. However, for L-HDL-C instrument, the exclusion of two SNPs (rs5576825 and rs6811162) previously associated with a potential confounder (waist circumference and hypertension respectively) resulted in non-significant association between L-HDL-C and 2-hour post glucose. Importantly, for both SNPs, it is conceivable that the confounders could reside on their causal pathway (i.e., vertically pleiotropic, where L-HDL-C effects 2-hr post-prandial glucose through its

effect on weight gain) rather than be in horizontal pleiotropy and may, therefore, violate the 2nd 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.

South Asians

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

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

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

No metabolite found to be associated with postprandial glucose measures in WEs was found to be associated with postprandial glucose in SAs. However, in both populations, members of the S-HDL and L-HDL class were found to be associated with increased postprandial glucose.

	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

Table 3: Significant MR results in South Asians

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

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

138 SNPS remained after LD thinning in WEs, 87 (63.04%) of which were unique to a single metabolite class exposure. 19/20 (95%) of the metabolite classes examined in WEs had an F-statistic ≥10 (the only exception being the MHDL class). 54 SNPS remained after LD thinning in SAs, 42 (77.78%) of which were unique to a single metabolite class. Despite the lower number of SNPs identified in SAs,17/20 (85%) of the metabolite classes examined had a F-statistic ≥10 indicating that instrument strength was still sufficient, for all classes except for the non-branched amino acids, LLDL class, and all VLDL classes. On average, the mean F-statistic of

metabolite class instruments was 19.83%. On average genetic instruments of composite measures of the metabolite classes were weaker than the instruments for individual metabolites in both WEs and SAs (**Supplementary Figure 13**).

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

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

Discussion

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

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

Interestingly, 3 of the 15 SNPs associated with tyrosine are found in a transmembrane transporter gene, *SLC13A2*. Moreover, an additional 10 of the newly identified 15 SNPs associated with tyrosine were found in the *FOXN* gene, a transcription factor that has previously been identified to be associated with ceramide levels (a lipid metabolite) in a GWAS from a Chinese cohort [50]. Moreover, ceramide has been shown to induce tyrosine phosphorylation in membrane proteins meaning it is plausible that a gene associated with ceramide is also associated with tyrosine levels in an Asian population [51]. Interestingly, ceramide has been proposed as a mediator of the interaction between saturated fat and insulin resistance and has been associated with T2D and cardiovascular disease [52]. To the best of our knowledge tyrosine levels have not previously been associated with either *FOXN SLC13A2*. The remaining 2 of the 15 SNPs identified as being associated with tyrosine in SAs, are currently not in any known genes. All 15 SNPs identified as being associated with tyrosine in SAs are in LD with each other in 1000G SA populations (all $R^2 \geq 0.38$).

Identified associations in WEs

Leucine

Branched chain amino acids (BCAAs), including leucine, are predominantly metabolised in skeletal muscles where they regulate protein synthesis and

mitochondrial functions [53]. In addition, BCAAs are hormonal signalling regulators and are expected to module insulin resistance (IR) through increasing insulin secretion in human pancreatic β-cells [54, 55]. Our study found leucine to be negatively associated with both fasting glucose and 2-hour post glucose levels during pregnancy in WEs; with 1 mmol/L of leucine associated with a decrease of 0.193 mmol/L in fasting glucose and 0.327 mmol/L 2-hour post glucose respectively. Although few studies have investigated the role of leucine in glucose regulation during pregnancy, interestingly the ratio of leucine/isoleucine was similarly found associated with reductions in fasting glucose in the HAPO study, a multi-ethnic cohort of pregnant women of Afro-Caribbean, Mexican American, Northern European, and Thai ancestry [56]. Common dietary sources of leucine include meat products and cheese, with smaller amounts also being present in other dairy products (such as dairy and yoghurt), fish and in certain legumes and nuts, such as dried raw broad beans and pine nuts [53]. Hence, dietary interventions aimed at increasing leucine levels during pregnancy, possibly through a dietary intervention promoting the consumption of lean animal protein, low-fat dairy and nuts, may help improve pregnancy hyperglycaemia in WEs.

Cholesterols

HDL cholesterol is colloquially described as ‘good cholesterol’ due to its role in the removal of cholesterols from atherosclerotic plaque, thereby reducing an individual’s risk of CVD [57]. Furthermore, low HDL levels have commonly been associated with diabetes in humans, with HDL shown to increase insulin secretion and β-cell survival [58, 59]. We identified four associations between HDL cholesterol and postprandial glucose measures in WEs. Herein, 1 mmol/L increase in S-HDL-CE

confers a 1.78 mmol/L increase (95% CI 0.49 – 3.11) in 2-hour post glucose. This is consistent with previous evidence from a Finnish sample of overweight and obese women where cholesterol esters in S-HDL were higher in the serum samples of GDM cases at ~14 weeks gestation [60]. Discrepancies in the direct effect of HDL cholesterol on dysglycemia have also been identified in the genetic literature [59], with a recent review highlighting that while a genetic study utilising linear relation analysis did find HDLs to have a protective effects against T2D (n cases = 2,447) [61], the same effect was not been replicated in an MR setting (n cases= 47,627) [62]. When considering LDL cholesterols, only S-LDLs was found to be significantly associated with a postprandial glucose measure (fasting glucose) in WEs. Additionally, in our composite analysis of metabolite classes, S-LDLs were associated with fasting glucose and 2-hour post glucose in WEs, whereas the M-LDL and all LDL (a combined measure of S-LDLs, M-LDLs and HDLs) classes were associated with fasting glucose. Unfortunately, because composite scores were comprised of PC1 coordinates the direction of effect of these associations could not be evaluated. To our knowledge, no previous study has conducted an MR of metabolites on dysglycemic predictors of GDM.

Triglycerides

Triglycerides are an abundant class of lipid particles found in the blood, originating from either from the consumption of dietary fats or as a result of hepatic metabolism [63, 64]. Once in the blood, triglycerides can be incorporated into HDL and LDL cholesterol particles. In addition to dietary triglyceride consumption, dietary fatty acids can be converted into triglycerides before they enter circulation,

highlighting the complex relationship between triglyceride, cholesterol, and fatty acid levels [63].

Our results suggest triglycerides in XSVLDL (XS-VLDL-TG) associate with increased 2-hour post glucose (0.841 mmol/L) in WEs. In agreement with these findings, increased triglycerides in XSVLDL levels have also previously been associated with increased likelihood of GDM in a Finnish population [60]. No other triglyceride was found to be associated with in WEs. One explanation no additional associations were detected could be due to the average BMI of the WEs in BiB. For example, an analysis of a prospective Irish cohort (~94% WE) found that triglyceride levels were only associated with GDM in obese individuals, a higher average BMI than that observed in the BiB cohort [65]. Further confirmation of these findings of increased triglycerides in XS-VLDLs would suggest that this association is, at least in part, responsible for the identified associations between diets high in fats and increased prevalence of GDM in WEs [11].

Identified associations in SAs

Fatty acids

Polyunsaturated acids (PUFAs) are consumed in the diet and can be converted to long-chain PUFAs (LC-PUFAs) through a process of desaturation and elongation reactions that predominately occur in the liver [66]. Changes in dietary patterns can have a large impact on fatty acid composition in the body and with-it disease risk. For example, a western dietary pattern, which has high levels of n-6 fatty acids, has been associated with GDM risk [11, 67]. In a cohort of Chinese adults, total n-6 fatty acids and 18:2 n-6 levels at baseline in venous blood samples

were both found to associate with an increased risk of T2D after ~8 years of follow up, while increased n-3 fatty acid levels were protective [68]. However, a recent two-sample MR suggested only a negligible effect of n-6 PUFA synthesis on T2D in a predominantly WE cohort [69]. Moreover, the relationship between n-3, n-6, n-9 fatty acids and GDM remains inconclusive, and in a recent (2021) systematic review none of the identified studies (n=15) was conducted in a SA population [67], highlighting the need for more studies exploring the role of fatty acids in GDM development in Asians [70].

This study provides evidence of an association between LA and total FAw6 levels and an increase in 2-hour post glucose levels during pregnancy in SAs. In addition, the fatty acid class associated with 2-hour post glucose in SAs. Through a leave-out-one sensitivity analyses for the FAw6 and LA instruments, the removal of the SNP rs12720820 (found within the *APOB* gene) resulted in non-significant associations for both exposures, indicating that this rs12720820 was the largest contributor to the identified associations and has previously been associated with cholesterol levels and the use of cholesterol lowering drugs [71]. Interestingly, FAw3 also associated with increased 2-hour post glucose in SAs; however, with only one SNP potential pleiotropy could not be explored.

It is well established that fatty acid profiles can impact blood cholesterol levels [72-74]. In addition, increased dietary cholesterol has previously been associated with an increased risk of GDM in a systematic review of observational studies [75]. Taken together, our data confirm that fatty acids and cholesterol metabolites are in vertically pleiotropy and are likely impacting gestational dysglycemia via the same causal pathway. Unlike horizontal pleiotropy, vertical pleiotropy does not result in a violation of the 2nd MR assumption as cholesterol is not acting as a confounder,

meaning MR estimates are still valid (**Figure 1**). Furthermore, it is also possible that this interaction between fatty acids and cholesterols may be ethnic-specific due to the absence of associations identified between fatty acids and postprandial glucose measures in WEs. In addition to possible variations in cholesterol metabolism, it is plausible that variations in fatty acid synthesis are also partially responsible for the increased GDM risk experienced by SAs. For example, variants within the *FADS* genes impact LC-PUFA conversion [76, 77]. Current evidence suggests that SAs are likely to synthesise LC-PUFA more quickly than WEs, which could contribute to elevated risk of prolonged exposure to elevated LC-PUFA levels (namely, w6) and risk of dysglycemia [76, 77]. If these ethnic differences in fatty acid metabolism are confirmed to be linked to disease risk, it would aid in the development of tailored GDM prevention strategies 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].

A

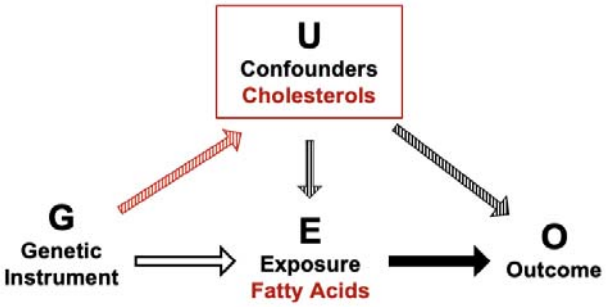

Figure 1: Schematic of potential horizontal and vertical pleiotropy in relation to fatty acid and cholesterol metabolites and postprandial glucose measures.

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

Nonetheless, this study has some limitations. Metabolites are highly correlated meaning it is not possible to confidently interpret that an individual metabolite is independently associated with a postprandial outcome measure. To account for this limitation MR analyses were performed on composite measures of each metabolite class (when PC1 explained ≥70% of the variation in the metabolite class) in order to assess the overall impact of each metabolite class on pregnancy dysglycemia. Furthermore, MR also assumes the level of genetically conferred exposure from conception to the time of measurement is constant, which is unknown when studying metabolites. The limited sample size of this study also means that

further adjustment could not be made at the GWAS stage since missing data in certain variables would further reduce sample size (eg, age) and some associations may have been underpowered to detect an effect. However, a *post hoc* power analyses found that for some metabolite values significant effects only identified in one ethnicity were possible to detect in the alternate ethnicity. In addition, some genetic instruments included only one SNP meaning it was not possible to evaluate the impact of pleiotropy for any 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 fatty acids 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:

775 10.3390/jcm7060120. PubMed PMID: 29882903; PubMed Central PMCID:
776 PMCPMC6024961.
777 7. Menard V, Sotunde OF, Weiler HA. Ethnicity and Immigration Status as Risk Factors
778 for Gestational Diabetes Mellitus, Anemia and Pregnancy Outcomes Among Food Insecure
779 Women Attending the Montreal Diet Dispensary Program. Can J Diabetes. 2020;44(2):139-
780 45 e1. Epub 2019/08/21. doi: 10.1016/j.jcjd.2019.05.004. PubMed PMID: 31427254.
781 8. Gadve SS, Chavanda S, Mukherjee AD, Aziz S, Joshi A, Patwardhan M. Risk of
782 Developing Type 2 Diabetes Mellitus in South Asian Women with History of Gestational
783 Diabetes Mellitus: A Systematic Review and Meta-Analysis. Indian J Endocrinol Metab.
784 2021;25(3):176-81. Epub 2021/11/12. doi: 10.4103/ijem.IJEM_57_21. PubMed PMID:
785 34760669; PubMed Central PMCID: PMCPMC8547406.
786 9. Fuller H, Iles M, Moore JB, Zulyniak MA. Unique Metabolic Profiles Associate with
787 Gestational Diabetes and Ethnicity in Low and High-Risk Women Living in the UK. J Nutr.
788 2022. Epub 20220726. doi: 10.1093/jn/nxac163. PubMed PMID: 35883228.
789 10. Taylor K, Ferreira DLS, West J, Yang T, Caputo M, Lawlor DA. Differences in
790 Pregnancy Metabolic Profiles and Their Determinants between White European and South
791 Asian Women: Findings from the Born in Bradford Cohort. Metabolites. 2019;9(9):190. doi:
792 10.3390/metabo9090190. PubMed PMID: 31540515.
793 11. Fuller H, Moore JB, Iles MM, Zulyniak MA. Ethnic-specific associations between
794 dietary consumption and gestational diabetes mellitus incidence: A meta-analysis. PLOS
795 Global Public Health. 2022;2(5):e0000250. doi: 10.1371/journal.pgph.0000250. PubMed
796 PMID: pub.1147817197.
797 12. Griffith RJ, Alsweiler J, Moore AE, Brown S, Middleton P, Shepherd E, et al.
798 Interventions to prevent women from developing gestational diabetes mellitus: an overview
799 of Cochrane Reviews. Cochrane Database Syst Rev. 2020;6:CD012394. Epub 2020/06/12.
800 doi: 10.1002/14651858.CD012394.pub3. PubMed PMID: 32526091; PubMed Central PMCID:
801 PMCPMC7388385.
802 13. de Souza RJ, Shanmuganathan M, Lamri A, Atkinson SA, Becker A, Desai D, et al.
803 Maternal Diet and the Serum Metabolome in Pregnancy: Robust Dietary Biomarkers
804 Generalizable to a Multiethnic Birth Cohort. Curr Dev Nutr. 2020;4(10):nzaa144-nzaa. doi:
805 10.1093/cdn/nzaa144. PubMed PMID: 33073162.
806 14. Wang QY, You LH, Xiang LL, Zhu YT, Zeng Y. Current progress in metabolomics of
807 gestational diabetes mellitus. World J Diabetes. 2021;12(8):1164-86. Epub 2021/09/14. doi:
808 10.4239/wjd.v12.i8.1164. PubMed PMID: 34512885; PubMed Central PMCID:
809 PMCPMC8394228.
810 15. Sheehan NA, Meng S, Didelez V. Mendelian Randomisation: A Tool for Assessing
811 Causality in Observational Epidemiology. In: Teare MD, editor. Genetic Epidemiology.
812 Totowa, NJ: Humana Press; 2011. p. 153-66.
813 16. Liu J, van Klinken JB, Semiz S, van Dijk KW, Verhoeven A, Hankemeier T, et al. A
814 Mendelian Randomization Study of Metabolite Profiles, Fasting Glucose, and Type 2
815 Diabetes. Diabetes. 2017;66(11):2915-26. Epub 2017/08/30. doi: 10.2337/db17-0199.
816 PubMed PMID: 28847883.
817 17. Carreras-Torres R, Johansson M, Gaborieau V, Haycock PC, Wade KH, Relton CL, et al.
818 The Role of Obesity, Type 2 Diabetes, and Metabolic Factors in Pancreatic Cancer: A
819 Mendelian Randomization Study. J Natl Cancer Inst. 2017;109(9). Epub 2017/09/28. doi:
820 10.1093/jnci/djx012. PubMed PMID: 28954281; PubMed Central PMCID: PMCPMC5721813.

821 18. Wright J, Small N, Raynor P, Tuffnell D, Bhopal R, Cameron N, et al. Cohort Profile:
822 the Born in Bradford multi-ethnic family cohort study. Int J Epidemiol. 2013;42(4):978-91.
823 Epub 20121012. doi: 10.1093/ije/dys112. PubMed PMID: 23064411.
824 19. Bird PK, McEachan RRC, Mon-Williams M, Small N, West J, Whincup P, et al. Growing
825 up in Bradford: protocol for the age 7-11 follow up of the Born in Bradford birth cohort.
826 BMC Public Health. 2019;19(1):939. Epub 20190712. doi: 10.1186/s12889-019-7222-2.
827 PubMed PMID: 31300003; PubMed Central PMCID: PMCPMC6626420.
828 20. Lawlor DA, West J, Fairley L, Nelson SM, Bhopal RS, Tuffnell D, et al. Pregnancy
829 glycaemia and cord-blood levels of insulin and leptin in Pakistani and white British mother-
830 offspring pairs: findings from a prospective pregnancy cohort. Diabetologia.
831 2014;57(12):2492-500. Epub 2014/10/03. doi: 10.1007/s00125-014-3386-6. PubMed PMID:
832 25273345; PubMed Central PMCID: PMCPMC4218974.
833 21. Josse J, Husson F. missMDA: A Package for Handling Missing Values in Multivariate
834 Data Analysis. Journal of Statistical Software. 2016;70(1). doi: 10.18637/jss.v070.i01.
835 22. Arciero E, Dogra SA, Malawsky DS, Mezzavilla M, Tsismentzoglou T, Huang QQ, et al.
836 Fine-scale population structure and demographic history of British Pakistanis. Nat Commun.
837 2021;12(1):7189. Epub 2021/12/12. doi: 10.1038/s41467-021-27394-2. PubMed PMID:
838 34893604; PubMed Central PMCID: PMCPMC8664933.
839 23. Marees AT, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, et al. A
840 tutorial on conducting genome-wide association studies: Quality control and statistical
841 analysis. Int J Methods Psychiatr Res. 2018;27(2):e1608. Epub 2018/02/28. doi:
842 10.1002/mpr.1608. PubMed PMID: 29484742; PubMed Central PMCID: PMCPMC6001694.
843 24. Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, et al. Genome-
844 wide association studies. Nature Reviews Methods Primers. 2021;1(1). doi: 10.1038/s43586-
845 021-00056-9.
846 25. Sheridan E, Wright J, Small N, Corry PC, Oddie S, Whibley C, et al. Risk factors for
847 congenital anomaly in a multiethnic birth cohort: an analysis of the Born in Bradford study.
848 The Lancet. 2013;382(9901):1350-9. doi: https://doi.org/10.1016/S0140-6736(13)61132-0.
849 26. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. Advantages and pitfalls in the
850 application of mixed-model association methods. Nat Genet. 2014;46(2):100-6. Epub
851 2014/01/30. doi: 10.1038/ng.2876. PubMed PMID: 24473328; PubMed Central PMCID:
852 PMCPMC3989144.
853 27. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a
854 tool set for whole-genome association and population-based linkage analyses. Am J Hum
855 Genet. 2007;81(3):559-75. Epub 20070725. doi: 10.1086/519795. PubMed PMID: 17701901;
856 PubMed Central PMCID: PMCPMC1950838.
857 28. Purcell S. PLINK (Version 1.9).
858 29. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex
859 trait analysis. Am J Hum Genet. 2011;88(1):76-82. Epub 2010/12/21. doi:
860 10.1016/j.ajhg.2010.11.011. PubMed PMID: 21167468; PubMed Central PMCID:
861 PMCPMC3014363.
862 30. van den Berg S, Vandenplas J, van Eeuwijk FA, Lopes MS, Veerkamp RF. Significance
863 testing and genomic inflation factor using high-density genotypes or whole-genome
864 sequence data. J Anim Breed Genet. 2019;136(6):418-29. Epub 2019/06/20. doi:
865 10.1111/jbg.12419. PubMed PMID: 31215703; PubMed Central PMCID: PMCPMC6900143.

866 31. Rivadeneira F, Uitterlinden AG. Chapter 18 - Genetics of osteoporosis. In: Dempster
867 DW, Cauley JA, Bouxsein ML, Cosman F, editors. Marcus and Feldman's Osteoporosis (Fifth
868 Edition): Academic Press; 2021. p. 405-51.
869 32. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide
870 association scans. Bioinformatics. 2010;26(17):2190-1. Epub 2010/07/10. doi:
871 10.1093/bioinformatics/btq340. PubMed PMID: 20616382; PubMed Central PMCID:
872 PMCPMC2922887.
873 33. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-
874 specific haplotype structure and linking correlated alleles of possible functional variants.
875 Bioinformatics. 2015;31(21):3555-7. Epub 2015/07/04. doi: 10.1093/bioinformatics/btv402.
876 PubMed PMID: 26139635; PubMed Central PMCID: PMCPMC4626747.
877 34. Myers TA, Chanock SJ, Machiela MJ. LDlinkR: An R Package for Rapidly Calculating
878 Linkage Disequilibrium Statistics in Diverse Populations. Front Genet. 2020;11:157. Epub
879 2020/03/18. doi: 10.3389/fgene.2020.00157. PubMed PMID: 32180801; PubMed Central
880 PMCID: PMCPMC7059597.
881 35. Genomes Project Consortium. A global reference for human genetic variation.
882 Nature. 2015;526(7571):68-74. Epub 2015/10/04. doi: 10.1038/nature15393. PubMed
883 PMID: 26432245; PubMed Central PMCID: PMCPMC4750478.
884 36. R Development Core Team. R: A language and environment for statistical computing
885 Vienna, Austria: R Foundation for Statistical Computing; 2010.
886 37. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag;
887 2016.
888 38. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation
889 PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015;4(1). doi:
890 10.1186/s13742-015-0047-8.
891 39. Budu-Aggrey A, Brumpton B, Tyrrell J, Watkins S, Modalsli EH, Celis-Morales C, et al.
892 Evidence of a causal relationship between body mass index and psoriasis: A mendelian
893 randomization study. PLoS Med. 2019;16(1):e1002739. Epub 2019/02/01. doi:
894 10.1371/journal.pmed.1002739. PubMed PMID: 30703100.
895 40. Fox J, Kleiber, C., Zeileis, A. _ivreg: Instrumental-Variables Regression by '2SLS',
896 '2SM', or '2SMM', with Diagnostics_. . R package version 0.6-1 ed2021.
897 41. Skrivankova VW, Richmond RC, Woolf BAR, Davies NM, Swanson SA, VanderWeele
898 TJ, et al. Strengthening the reporting of observational studies in epidemiology using
899 mendelian randomisation (STROBE-MR): explanation and elaboration. BMJ.
900 2021;375:n2233. Epub 2021/10/28. doi: 10.1136/bmj.n2233. PubMed PMID: 34702754.
901 42. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The
902 NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays
903 and summary statistics 2019. Nucleic Acids Research. 2019;47.
904 43. Staley JR, Blackshaw J, Kamat MA, Ellis S, Surendran P, Sun BB, et al. PhenoScanner: a
905 database of human genotype-phenotype associations. Bioinformatics. 2016;32(20):3207-9.
906 Epub 2016/06/19. doi: 10.1093/bioinformatics/btw373. PubMed PMID: 27318201; PubMed
907 Central PMCID: PMCPMC5048068.
908 44. Kamat MA, Blackshaw JA, Young R, Surendran P, Burgess S, Danesh J, et al.
909 PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations.
910 Bioinformatics. 2019;35(22):4851-3. Epub 2019/06/25. doi: 10.1093/bioinformatics/btz469.
911 PubMed PMID: 31233103; PubMed Central PMCID: PMCPMC6853652.

45. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ. 2018;362:k601. Epub 2018/07/12. doi: 10.1136/bmj.k601. PubMed PMID: 30002074; PubMed Central PMCID: PMCPMC6041728.
46. Brion M-JA, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. International journal of epidemiology. 2012;42(5):1497-501.
47. Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, et al. Ensembl 2021. Nucleic Acids Research. 2020;49(D1):D884-D91. doi: 10.1093/nar/gkaa942.
48. Ensembl. Gene: SORCS1 e! Ensembl: Ensembl release 105; 2021 [cited 2022 01/04/2022]. Available from: https://www.ensembl.org/Homo_sapiens/Gene/Summary?g=ENSG00000108018;r=10:106573663-107164706.
49. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med. 2008;27(8):1133-63. doi: 10.1002/sim.3034. PubMed PMID: 17886233.
50. Chai JF, Raichur S, Khor IW, Torta F, Chew WS, Herr DR, et al. Associations with metabolites in Chinese suggest new metabolic roles in Alzheimer's and Parkinson's diseases. Hum Mol Genet. 2020;29(2):189-201. Epub 2019/10/20. doi: 10.1093/hmg/ddz246. PubMed PMID: 31628463.
51. Gulbins E, Szabo I, Baltzer K, Lang F. Ceramide-induced inhibition of T lymphocyte voltage-gated potassium channel is mediated by tyrosine kinases. Proc Natl Acad Sci U S A. 1997;94(14):7661-6. Epub 1997/07/08. doi: 10.1073/pnas.94.14.7661. PubMed PMID: 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.

957 59. Wong NKP, Nicholls SJ, Tan JTM, Bursill CA. The Role of High-Density Lipoproteins in
958 Diabetes and Its Vascular Complications. International journal of molecular sciences.
959 2018;19(6):1680. doi: 10.3390/ijms19061680. PubMed PMID: 29874886.
960 60. Mokkala K, Vahlberg T, Pellonperä O, Houttu N, Koivuniemi E, Laitinen K. Distinct
961 Metabolic Profile in Early Pregnancy of Overweight and Obese Women Developing
962 Gestational Diabetes. J Nutr. 2020;150(1):31-7. Epub 2019/09/19. doi: 10.1093/jn/nxz220.
963 PubMed PMID: 31529056.
964 61. Qi Q, Liang L, Doria A, Hu FB, Qi L. Genetic predisposition to dyslipidemia and type 2
965 diabetes risk in two prospective cohorts. Diabetes. 2012;61(3):745-52. Epub 20120207. doi:
966 10.2337/db11-1254. PubMed PMID: 22315312; PubMed Central PMCID: PMCPMC3282815.
967 62. Haase CL, Tybjærg-Hansen A, Nordestgaard BG, Frikke-Schmidt R. HDL Cholesterol
968 and Risk of Type 2 Diabetes: A Mendelian Randomization Study. Diabetes. 2015;64(9):3328-
969 33. Epub 20150513. doi: 10.2337/db14-1603. PubMed PMID: 25972569.
970 63. Alexopoulos A-S, Qamar A, Hutchins K, Crowley MJ, Batch BC, Guyton JR.
971 Triglycerides: Emerging Targets in Diabetes Care? Review of Moderate Hypertriglyceridemia
972 in Diabetes. Curr Diab Rep. 2019;19(4):13-. doi: 10.1007/s11892-019-1136-3. PubMed
973 PMID: 30806837.
974 64. Laufs U, Parhofer KG, Ginsberg HN, Hegele RA. Clinical review on triglycerides. Eur
975 Heart J. 2020;41(1):99-109c. doi: 10.1093/eurheartj/ehz785. PubMed PMID: 31764986.
976 65. O'Malley EG, Reynolds CME, Killalea A, O'Kelly R, Sheehan SR, Turner MJ. Maternal
977 obesity and dyslipidemia associated with gestational diabetes mellitus (GDM). Eur J Obstet
978 Gynecol Reprod Biol. 2020;246:67-71. Epub 2020/01/22. doi: 10.1016/j.ejogrb.2020.01.007.
979 PubMed PMID: 31962258.
980 66. Gonzalez-Soto M, Mutch DM. Diet Regulation of Long-Chain PUFA Synthesis: Role of
981 Macronutrients, Micronutrients, and Polyphenols on Δ-5/Δ-6 Desaturases and Elongases
982 2/5. Adv Nutr. 2021;12(3):980-94. doi: 10.1093/advances/nmaa142. PubMed PMID:
983 33186986; PubMed Central PMCID: PMCPMC8166571.
984 67. Hosseinkhani S, Dehghanbanadaki H, Aazami H, Pasalar P, Asadi M, Razi F.
985 Association of circulating omega 3, 6 and 9 fatty acids with gestational diabetes mellitus: a
986 systematic review. BMC Endocrine Disorders. 2021;21(1):120. doi: 10.1186/s12902-021-
987 00783-w.
988 68. Bragg F, Kartsonaki C, Guo Y, Holmes M, Du H, Yu C, et al. Circulating Metabolites
989 and the Development of Type 2 Diabetes in Chinese Adults. Diabetes Care. 2021;45(2):477-
990 80. doi: 10.2337/dc21-1415.
991 69. Zulyniak MA, Fuller H, Iles MM. Investigation of the Causal Association between
992 Long-Chain n-6 Polyunsaturated Fatty Acid Synthesis and the Risk of Type 2 Diabetes: A
993 Mendelian Randomization Analysis. Lifestyle Genomics. 2020;13(5):146-53. doi:
994 10.1159/000509663.
995 70. Pan X-F, Huang Y, Li X, Wang Y, Ye Y, Chen H, et al. Circulating fatty acids and risk of
996 gestational diabetes mellitus: prospective analyses in China. European Journal of
997 Endocrinology. 2021;185(1):87-97. doi: 10.1530/eje-21-0118.
998 71. MedlinePlus. APOB gene National Library of Medicine (US): NIH; 2021 [cited 2022
999 01/04/2022]. Available from: https://medlineplus.gov/genetics/gene/apob/.
1000 72. Mensink RP, Zock PL, Kester AD, Katan MB. Effects of dietary fatty acids and
1001 carbohydrates on the ratio of serum total to HDL cholesterol and on serum lipids and
1002 apolipoproteins: a meta-analysis of 60 controlled trials. The American Journal of Clinical
1003 Nutrition. 2003;77(5):1146-55. doi: 10.1093/ajcn/77.5.1146.

73. Jacobson TA, Glickstein SB, Rowe JD, Soni PN. Effects of eicosapentaenoic acid and docosahexaenoic acid on low-density lipoprotein cholesterol and other lipids: a review. J Clin Lipidol. 2012;6(1):5-18. Epub 2012/01/24. doi: 10.1016/j.jacl.2011.10.018. PubMed PMID: 22264569.
74. Ooi EM, Ng TW, Watts GF, Barrett PH. Dietary fatty acids and lipoprotein metabolism: new insights and updates. Curr Opin Lipidol. 2013;24(3):192-7. Epub 2013/04/27. doi: 10.1097/MOL.0b013e3283613ba2. PubMed PMID: 23619368.
75. Schoenaker DA, Mishra GD, Callaway LK, Soedamah-Muthu SS. The Role of Energy, Nutrients, Foods, and Dietary Patterns in the Development of Gestational Diabetes Mellitus: A Systematic Review of Observational Studies. Diabetes Care. 2016;39(1):16-23. Epub 2015/12/24. doi: 10.2337/dc15-0540. PubMed PMID: 26696657.
76. Koletzko B, Reischl E, Tanjung C, Gonzalez-Casanova I, Ramakrishnan U, Meldrum S, et al. FADS1 and FADS2 Polymorphisms Modulate Fatty Acid Metabolism and Dietary Impact on Health. Annu Rev Nutr. 2019;39:21-44. Epub 2019/08/23. doi: 10.1146/annurev-nutr-082018-124250. PubMed PMID: 31433740.
77. Ameur A, Enroth S, Johansson A, Zaboli G, Igl W, Johansson AC, et al. Genetic adaptation of fatty-acid metabolism: a human-specific haplotype increasing the biosynthesis of long-chain omega-3 and omega-6 fatty acids. Am J Hum Genet. 2012;90(5):809-20. Epub 2012/04/17. doi: 10.1016/j.ajhg.2012.03.014. PubMed PMID: 22503634; PubMed Central PMCID: PMCPMC3376635.
78. Hu J, Gillies CL, Lin S, Stewart ZA, Melford SE, Abrams KR, et al. Association of maternal lipid profile and gestational diabetes mellitus: A systematic review and meta-analysis of 292 studies and 97,880 women. EClinicalMedicine. 2021;34:100830. Epub 2021/05/18. doi: 10.1016/j.eclinm.2021.100830. PubMed PMID: 33997732; PubMed Central PMCID: 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.

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

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

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

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.

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.

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

Supplementary Table 10: Post-hoc power analysis.

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.