# The Role of the Metabolome in the Development of Gestational Diabetes Mellitus in High-Risk Minority Women: A Causal Investigation.

Harriett Fuller

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
Faculty of Environment
School of Food Science and Nutrition

May, 2022

Supervised by Dr Michael A. Zulyniak [1], Dr Mark M. Iles [2-3] and Dr J. Bernadette Moore [1]

1. School of Food Science and Nutrition, Faculty of Environment, University of Leeds
2. School of Medicine, University of Leeds, Leeds, UK
3. Leeds Institute for Data Analytics, University of Leeds, Leeds, UK

The candidate confirms that the work submitted is his/her own and that appropriate credit has been given where reference has been made to the work of others.

## Acknowledgements

First and foremost, I would like to express my upmost gratitude to my primary supervisor, Dr Michael Zulyniak for his constant guidance, support, expertise, and enthusiasm throughout the duration of my project, despite the trying times of the Covid-19 pandemic. I would also like to extend this thanks to my co-supervisors, Dr Mark Iles and Dr J. Bernadette Moore for their guidance, support, and expertise also. Through this joint mentorship I have learnt a lot and improved greatly as a scientist, for which I am forever grateful for.

I also give my sincerest thanks to the participants of Born in Bradford for providing the data used throughout this project as well as all researchers in the Born in Bradford team that have aided in the development of the cohort. I would also like to thank the Advanced Research Computing (ARC) team at the University of Leeds for proving essential training on the ARC4 high performance computing (HPC) facilities that were utilised in this project. I would also like to thank the School of Food Science and Nutrition at the University of Leeds for funding my project and the Nutritional Epidemiology group at the University of Leeds for providing ample opportunities for friendly and intellectual discussions throughout.

Finally, I would also like to thank my family, partner, friends, and pets for all of their continual love and support throughout my PhD journey. Most importantly, I would like to thank my mum for all of her unwavering support throughout my education and beyond, as none of this would be possible without her.

# List of Abbreviations

| Abbreviation | Definition |
|---|---|
| 1000G | 1000 Genomes Project |
| AHEI | Alternate Healthy Eating Index |
| aOR | Adjusted odds ratio |
| BCAAs | Branched chain amino acids |
| BiB | Born in Bradford |
| CI | Confidence interval |
| CVD | Cardiovascular disease |
| DL | DerSimonian and Laird |
| EA | East Asian |
| FAw3 | n-3 fatty acids |
| FAw6 | n-6 fatty acids |
| GDM | Gestational diabetes mellitus |
| GWAS | Genome wide association study |
| GWG | Gestational weight gain |
| HKSJ | Hartung-Knapp-Sidik-Jonkman |
| IADSPG | International Association of Diabetes on Pregnancy Study Groups |
| IQR | Inter quartile range |
| IV | Instrumental variable |
| LA | 18:2 Linoleic Acid |
| LC-PUFA | long chain PUFA |
| LD | Linkage disequilibrium |
| LMICs | Low-middle income countries |
| LogL | Log liklihood |
| NHS | National Health Service |
| NST | Normal score transformation |

| | |
|---|---|
| MAF | Minor allele frequency |
| MCA | Multiple correspondence analysis |
| MR | Mendelian randomisation |
| MUFA | Monounsaturated fatty acids |
| MW | Mann-Whitney test |
| NMR | Nuclear magnetic resonance |
| NST | Normal score transformation |
| OGTT | Oral glucose tolerance test |
| OLS | Ordinary least squares |
| OR | Odds ratio |
| oPLSDA | Orthogonal partial least squares discriminatory Analysis |
| PC | Principal component |
| PCA | Principal component analysis |
| PLSDA | Partial least squares discriminatory analysis |
| PK | Pakistani |
| PUFAs | Polyunsaturated fatty acids |
| RCTs | Randomised controlled trials |
| RMSEE | Root mean squared error of estimation |
| ROC | Receiver operator curve |
| RR | Risk ratio |
| SA | South Asian |
| SAC-H | South Asian cases High |
| SAC-N | South Asian cases normal |
| SANC-N | South Asian non-case normal |
| SANC-H | South Asian non-case high |
| SE | Standard error |
| SFAs | Saturated fatty acids |
| sPLSDA | Sparse partial least squares discriminatory analysis |

| | |
|---|---|
| SUS | Shared and unique structure |
| T2D | Type 2 diabetes |
| TE | Treatment effect |
| TotFA | Total fatty acids |
| TSLS | Two stage least squares |
| VIF | Variance inflation factors |
| VIP | Variable importance in the projection |
| WE | White European |
| WEC-H | White European cases high |
| WEC-N | White European cases normal |
| WENC-H | White European non-cases high |
| WENC-N | White European non-cases high |
| WHO | World Health Organisation |

## Abstract

Gestational Diabetes Mellitus (GDM) is the most common pregnancy complication worldwide. However, GDM prevalence is substantially lower in white Europeans (WEs) compared to other ethnicities, especially South Asians (SAs) who experience the highest risk. Globally, healthy diet promotion is the mainstay in GDM prevention, however current guidelines are predominantly based on evidence from WEs. Furthermore, metabolic factors responsible for the disparities in prevalence are unknown but may offer guidance for improved prevention and management.

This project aimed to (i) assess the association between diet and GDM across ethnic groups, (ii) determine if distinct metabolic profiles characterise GDM in SAs and WEs, and (iii) evaluate the presence of ethnic-specific causal associations between metabolites and gestational dysglycemia. Aims (ii) and (iii) utilised data from the Born in Bradford cohort (mean gestational age 26.1 weeks).

First, through a systematic review of observational and randomised studies, pre-pregnancy diet was found to associate with GDM in WEs, but not in Asians. Secondly, the multivariate analyses of metabolites identified 7 metabolites that were characteristic of GDM in both ethnicities, with an additional 6 characteristic in WEs only. Finally, through Mendelian Randomisation (MR) analyses, 14 metabolites associated with pregnancy dysglycemia in WEs and 11 in SAs. No metabolites were identified in both ethnicities. Cholesterols and fatty acids were the most commonly identified classes identified in WEs and SAs, respectively.

This project demonstrated (i) inconsistencies in the association between diet and GDM across ethnicities (ii) distinct metabolic profiles that associate with GDM in WEs and SAs and offers and supports the need for ethnic-specific manage GDM management strategies. In high-risk SAs, fatty acids may be the most important predictors of GDM. Future work should evaluate the role of pre-pregnancy fatty acid intake in GDM development in SAs to aid in the development of culturally tailored dietary interventions.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1: Literature Review

## 1.1 Gestational Diabetes Mellitus

Gestational Diabetes Mellitus (GDM) is defined as the development of hyperglycaemia during pregnancy (typically at 24-28 weeks) (McIntyre et al., 2019; Lorenzo-Almorós et al., 2019). Affecting approximately 1 in 6 pregnancies globally and 20.4 million live births in 2019, GDM is the most common pregnancy complication to occur worldwide and encompasses 70-90% of all incidences of hyperglycaemia during pregnancy (Newman and Dunne, 2021; McIntyre et al., 2020; Nielsen et al., 2018; Yuen et al., 2018). Despite this already substantial disease burden, GDM prevalence is on the continual rise, in part due to the global obesity epidemic (Wang et al., 2021; Kuller and Catov, 2017; Dalfrà et al., 2020). For example, in a serial cross-sectional analysis in the US, age-standardised rates of GDM per 1000 live births increased from 47.6 (95% confidence interval (CI) 47.1- 48.0) in 2011 to 63.5 (95% CI 63.1 - 64.0) in 2019 (Shah et al., 2021). Australia also experienced a 2.87-fold increase in GDM between 2010-2019, making GDM the fastest growing type of diabetes in this country. This increase is expected to continue, with an additional 500,000 GDM cases expected to occur in Australia over the next decade (Diabetes Australia, 2020). The UK has also experienced a substantial rise in GDM prevalence since 2010, with prevalence estimates for increasing from 1-3% pre-2010 to 8-24% post-2010 (Farrar et al., 2016).

The implementation of the International Association of Diabetes on Pregnancy Study Groups (IADSPG) GDM diagnostic criteria now recommended for use by the World Health Organisation (WHO) is thought to be partially responsible for this increased prevalence globally (Coustan et al., 2010; World Health Organization Consultation, 2014; Farrar et al., 2016). These criteria state that a single measure of fasting plasma glucose between 5.1-6.9 mmol/l, 1-hour post 75g oral glycose load ≥ 10.0 mmol/l, or 2-hour post glucose levels between 8.5-11.0 mmol/l following a 75g oral glucose tolerance test (OGTT) is sufficient to be indicative of GDM. These are more permissive

criteria than those previously utilised, resulting in a rise in GDM diagnoses globally (Coustan et al., 2010; World Health Organization Consultation, 2014). Indeed, a 2019 study of 51 population-based studies including 5,349,476 pregnant women found the pooled global estimate of GDM to be 4.4% (95% CI 4.3 - 4.4) when historically used GDM diagnostic criteria were utilised. This estimate increased drastically to 10.6% (95% CI 10.5 - 10.6) when only studies that utilised the IADSPG diagnostic criteria were included (Behboudi-Gandevani et al., 2019). Likewise, a 2021 systematic review and meta-analysis of cohort and cross-sectional studies conducted in 2010 - 2018 found that the utilisation of the IADPSG criteria resulted in a 75% increase in the number of women diagnosed with GDM (Risk Ratio (RR) 1.75, 95% CI 1.53 - 2.01) worldwide (Saeedi et al., 2021).

The utilisation of these new criteria also results in high heterogeneity of GDM prevalence estimates, particularly in countries where the uptake of these new criteria is inconsistent (Kanguru et al., 2014). For example, a recent review of Indian studies found prevalence estimates to range from 0% - 49%, making it difficult to untangle the extent to which the rise in prevalence is due to new diagnostic criteria, demographic changes, including an increase in maternal age and BMI (both risk factors of GDM), or increasing urbanisation (McIntyre et al., 2019; Thanawala et al., 2021). Importantly, despite the lower uptake of the IADPSG criteria in low-middle income countries (LMICs), 87.6% of GDM cases are still believed to occur in LMICs (Nielsen et al., 2018). Considering that a GDM pregnancy is predicted to cost healthcare services an additional 25% compared to a non-GDM pregnancy, GDM can place a great strain onto healthcare services, particularly in LMICs (Xu et al., 2017). Indeed, in 2015 GDM cost the Chinese economy an estimated $5.59 billion (Xu et al., 2017). As GDM rates continue to rise, so will the burden of GDM on healthcare systems globally, highlighting the urgent need for effective methods of GDM prevention.

### 1.1.1 Health consequences of GDM

In addition to its economic burden, the health consequences of GDM can also be severe in both mothers and children, even in mild cases. This is in spite of the fact that initial symptoms of GDM are relatively minor and common

in general pregnancy (e.g., tiredness and increased urination) (Chong et al., 2014; International Diabetes Federation, 2019). Short-term, the mother's likelihood of suffering a miscarriage or birth complications is substantially increased (Plows et al., 2018). These birth complications include, but are not limited to increased likelihood of pre-eclampsia (~30% increased odds), vaginal candidiasis (~6 times the risk) and the need for caesarean section (~50% increase in risk). (Odar et al., 2004; Kanguru et al., 2014; McIntyre et al., 2020; Weissgerber and Mudd, 2015; Gorgal et al., 2012) In addition to the physical implications of a GDM diagnosis, the quality of life and mental health of mothers with GDM are known to be impacted, with GDM increasing the risk of post-natal depression nearly 2-fold (McIntyre et al., 2019; Barakat et al., 2014; Kunasegaran et al., 2021). In addition, a women with GDM is at an increased risk of hypertension and is 2.3 times more likely to develop cardiovascular disease (CVD) in the decade following their diagnosis. (Kuller and Catov, 2017; Kramer et al., 2019; Schiavone et al., 2016). Despite these risks, the key concern of a GDM diagnosis is the increased risk of type 2 diabetes (T2D) experienced by the mother in later life. Following a GDM diagnosis, a women is 7 times more likely to develop T2D, often in 3-6 years of their GDM diagnosis (International Diabetes Federation, 2019). This increased risk in a reasonably short period can result in a GDM mother experiencing T2D in middle age, which can have life-long health and quality of life consequences for the mother. From a public health perspective, this places a large burden on healthcare systems: an estimated one third of women with T2D (who have been pregnant) expected to have had GDM during at least one pregnancy, making GDM the strongest historical predictor of T2D (McIntyre et al., 2020; Damm et al., 2016). Furthermore, loss of follow up in a healthcare setting after a GDM diagnosis and pregnancy delivery is very common and can make it challenging to manage future diabetes risk, particularly in communities that have less healthcare access (Moon et al., 2017). Concerningly, the risk of stillbirth is also significantly increased in pregnant women with GDM: in a multi-ethnic retrospective US cohort the risk of stillbirth in women with GDM compared to in women without GDM is also increased (RR = 1.34, 95% CI 1.20 – 1.50) (Rosenstein et al., 2012).

Reoccurrence rates of GDM are also high, with 70% of cases developing GDM again in a future pregnancy (Diabetes Australia, 2020).

Offspring of a GDM mother also have increased health risks in both the short and long term. This relationships between maternal glycemia and adverse neonatal outcomes due to intrauterine hyperglycaemia is thought to be linear, with increased glucose levels increasing the offspring's risk of 3 of macrosomia 3-fold, (itself increasing the likelihood of nerve damage and fracture during birth), hyperbilirubinemia 2-fold and congenital heart abnormalities 2-fold. (Kanguru et al., 2014; McIntyre et al., 2020; Rai et al., 2021; Dalfrà et al., 2020; Kc et al., 2015; Chen, L. et al., 2019b; Kouhkan et al., 2021). Most importantly, the offspring's risk of T2D and obesity are also increased, the latter due to increases in adiposity tissue as a result of GDM increasing the offspring's fat mass (Popova et al., 2021; Plows et al., 2018; Kanguru et al., 2014; Kandasamy et al., 2021; International Diabetes Federation, 2019). This risk of T2D in GDM offspring has been suggested to be 8 times of that in the general population in a Danish cohort (Damm et al., 2016). Likewise, A 2018 systematic review and meta-analysis found the BMI z-score of GDM offspring in childhood to be higher compared to offspring from non-GDM mothers (mean difference$_{\text{BMI z-score}}$ 0.14, 95% CI 0.04 - 0.24), while also 2-hour post glucose levels in early adulthood were also higher in GDM offspring (mean difference$_{\text{2-hour post glucose}}$ 0.43mmol/L, 95% CI 0.18 - 0.69) (Kawasaki et al., 2018). Worryingly, as both an increased BMI and a family history of T2D are risk factors for GDM, an inter-generational disease cycle can become embedded in a population that is difficult to break (Unnikrishnan et al., 2016; Plows et al., 2018). In addition, CVD risk is also increased in offspring of GDM mothers (RR = 1.19, 95% CI 1.07 – 1.32) (Yu et al., 2019). Likewise, a 2020 systematic review also found GDM offspring to have a significantly higher systolic blood pressure (mean difference 1.75 mmHg, 95% CI 0.57 – 2.94) compared to in non-GDM offspring (Pathirana et al., 2020).

### 1.1.2 GDM aetiology

To date the biological mechanisms responsible for the development of GDM remain unknown. However, they are expected to be complex and multifactorial, involving a range of pathways including hormonal, inflammatory, autoimmune, metabolomic and genetic processes that are triggered by metabolic changes that occur during pregnancy (Schiavone et al., 2016; Clish, 2015; Dalfrà et al., 2020). Furthermore. It is thought that the early stages of GDM, namely insulin resistance and β-cell dysfunction, are likely asymptomatic and may present before conception and then manifest during pregnancy; suggesting that some individuals, including those living with obesity, are predisposed to GDM before pregnancy begins (McIntyre et al., 2019).

During pregnancy, there is a natural increase in catabolism to ensure that the foetus' energy demands for growth and development are met (Plows et al., 2018; Taylor et al., 2019; Mills et al., 2019). In the early stages of pregnancy, both maternal lipogenesis and insulin sensitivity increase to facilitate maternal energy storage in preparation for the ~14-fold increase in energy demand required for foetal growth and development during the 2$^{nd}$ and 3$^{rd}$ trimesters of pregnancy, resulting in β-cell expansion and increased lipid storage (Kuller and Catov, 2017; Smith, 2010; Lorenzo-Almorós et al., 2019). As the pregnancy advances and enters the 2$^{nd}$ trimester, the mother progresses into a state of hyperinsulinemia and insulin resistance in peripheral tissues, as well as a state of elevated lipolysis, hepatic gluconeogenesis, and a 50-60% reduction in insulin sensitivity (Schiavone et al., 2016; Layton et al., 2019; Lorenzo-Almorós et al., 2019; McIntyre et al., 2019; Kampmann et al., 2019). Insulin resistance is defined as the reduced ability of target tissues, the liver and adipose tissue, to respond to circulating insulin (Sonagra et al., 2014). Increases in insulin resistance occur in all pregnancies in order to facilitate glucose uptake by the foetus by limiting glucose uptake in the mother (Kampmann et al., 2019). This increase in maternal blood glucose and insulin resistance in peripheral tissues is governed by increasing maternal adiposity and the insulin-agnostic paracrine action of placental hormones and placental factors, including human placental lactogen, oestrogen, progesterone, leptin,

cortisol, and placental growth hormone (Popova et al., 2021; Plows et al., 2018; Schiavone et al., 2016; Clish, 2015; Dalfrà et al., 2020).

To compensate for this decreased level of insulin sensitivity and increased insulin resistance occurring in later stages of pregnancy, the body undergoes a 2-3-fold increase in insulin secretion (McIntyre et al., 2019). Although essential, an inability to effectively control this increase in insulin resistance is suspected to be the cause of 80% of GDM diagnoses, with the remaining 20% possibly being attributable to pancreatic diseases, chemically-induced diabetes (for example, as a result of organ transplant or HIV infection) or autoimmune disorders (Plows et al., 2018). These defects in insulin resistance are likely a result of the additive effect of increased insulin resistance as a result of pregnancy occurring alongside β-cell dysfunction and a pre-existing background of chronic insulin resistance. This β-cell dysfunction can occur as a result of insulin-producing pancreatic β-cells being overburdened by increased blood glucose levels and, when occurring alongside increased insulin resistance observed during pregnancy, can result in pregnancy dysglycemia and with it GDM (22). The biological mechanisms that lead to β-cell dysfunction are unclear but are expected to affect all stages of insulin synthesis and excretion, partially as a result of the metabolic changes that occur throughout even a healthy pregnancy that act as β-cell stressors (McIntyre et al., 2019).

In addition to hyperglycaemia, hyperlipidaemia is also a characterising factor of GDM (Mauro et al., 2022; Layton et al., 2019). The foetus is unable to synthesise fatty acids that are required for growth and development and is, therefore, completely reliant on the mother during pregnancy (Mills et al., 2019). Lipid metabolites have been associated with GDM and the physiological changes in lipid metabolism that accompany a normal pregnancy are expected to be amplified in GDM cases, although the precise mechanism by which hyperlipidaemia (elevated blood lipids) can influence GDM is unclear. However, it is suspected to be dependent on aetiology of insulin resistance (i.e., the extent to which an insulin sensitivity defect is present alongside an insulin resistance defect) and the interactions between distinct lipid subtypes (Layton et al., 2019). For example, in a European cohort of 805 individuals, nonesterified fatty acids were elevated in 8-hour fasting

serum samples in all cases, whereas those with increased insulin resistance had higher levels of triglycerides and lower levels of HDL than individuals compared to individuals' insulin secretion defects (Layton et al., 2019). Moreover, individuals with insulin secretion defects had comparable lipid profiles to individuals without GDM (Layton et al., 2019). This illustrates the complexity of the metabolic changes that characterise GDM and the heterogeneity of GDM cases.

In summary, the aetiology of GDM is multifaceted with numerous potential pathways leading to the single prognosis of elevated maternal blood glucose levels. If not managed, the condition will lead to prolonged exposure of the mother and foetus to elevated glucose that elevates their risk for future health conditions.

### 1.1.3 GDM epidemiology

Numerous well-recognised risk factors of GDM have been identified (**Figure 1.1**)**.** Firstly, BMI is a major, well-established risk factor for GDM. Pregnant women with a BMI ≥30 kg/m$^2$ experience 3 times the odds (adjusted odds ratio (aOR) = 3.07, 95% CI 2.10 – 4.49) of GDM compared to women with a healthy BMI (≤25 kg/m$^2$) (El-Chaar et al., 2013). This increase in GDM odds increases as BMI increases, with pregnant women with a BMI ≥ 40 kg/m$^2$ being at 5.7 times the odds (aOR = 5.70, 95% CI 3.73 – 8.70) of the disease compared to individuals in the healthy BMI range (El-Chaar et al., 2013). Alongside increased BMI, increased maternal age (specifically ≥35 years of age) is a major GDM risk factor that is becoming ever-more important as childbearing age rises globally (McIntyre et al., 2019; Menard et al., 2020). Additional GDM risk factors include multiple pregnancies (i.e., twins or triplets), carrying a male offspring, the number of previous children (parity), smoking, or having a family history of T2D. Additional environmental risk factors for GDM have also been proposed, including exposure to endocrine-disruptors in the form of organic pollutants, although these are less well established (McIntyre et al., 2019).

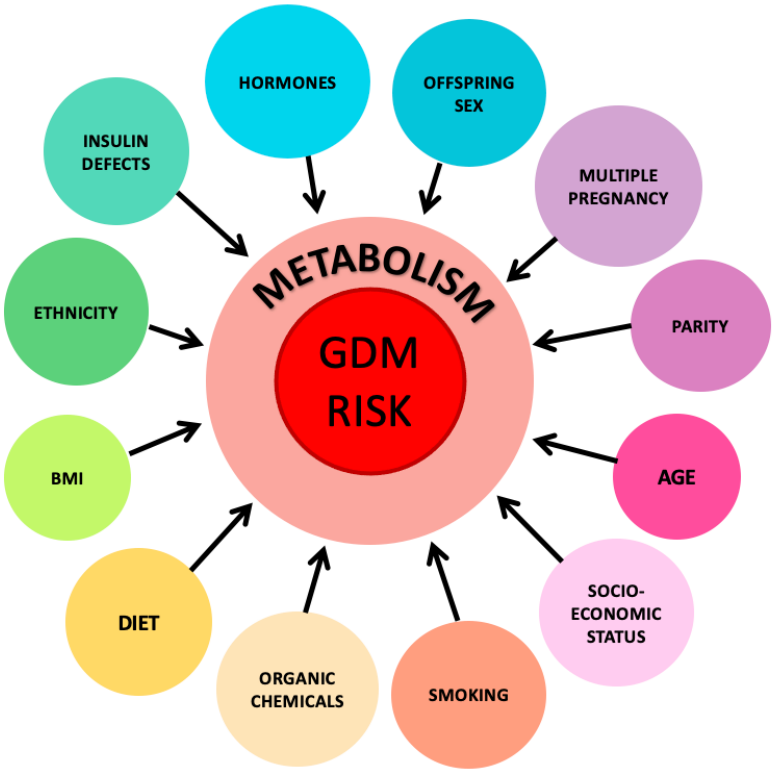



**Figure 1.1: Schematic of known GDM risk factors.**

To date, one of the strongest known risk factors for GDM is ethnicity, with white European (WE) women consistently being found to have a lower risk of GDM compared to women of non-white European ancestries. The median GDM prevalence estimate from 2015 - 2018 in the European WHO region (assumed to be majority WE) was 6.1% (inter quartile range (IQR) 1.8% – 31%), was lower than all other global regions. North America and the Caribbean experienced the second lowest prevalence estimate at 7% (IQR 6.5% - 11.9%), while the prevalence estimates in the South-East Asian region (formed of 11 countries including India, Bangladesh, Nepal, Sri Lanka and Indonesia) were more than double the European prevalence at 15% (IQR 9.6% - 18.3%). The region with the highest prevalence estimates during this period was the Middle East and North Africa region (formed of 20 countries/territories including Pakistan, Iran, Israel, Saudi-Arabia, Oman and Egypt) where prevalence reached 15.2% (IQR 8.8% - 20%) (McIntyre et al., 2019). Moreover, although the IADSPG diagnostic criteria are well accepted across Asia, limited resources often prevent the uptake of these criteria,

potentially resulting in as many as 50% of GDM cases being missed across Asia (Xu et al., 2017; Chong et al., 2014). This means that the above discrepancies in prevalence may be underestimated meaning that disparities in prevalence may be even more pronounced (McIntyre et al., 2019).

Notable trends in prevalence estimates also exist across continental Asia, with East Asian (EA) countries (e.g., China, South Korea, Japan and Taiwan), and South-East Asian countries (e.g., Thailand, Cambodia, Vietnam, Malaysia and the Philippians) experiencing a lower level of prevalence compared to their South Asian (SA) (India, Pakistan, Sri Lanka and Bangladesh) neighbours (Schiavone et al., 2016). Indeed, the estimated prevalence of GDM in the WHO Western Pacific region from 2015-2018 was 10.3% (4.5% – 20.3%), 4.7% lower than that experienced in the South-East Asian WHO region (which encompasses both SA and South-East Asian countries) (McIntyre et al., 2019). In the Asian subcontinent, intra-country variation in GDM prevalence also exists. For example, although a third of pregnancies in India are expected to be impacted by GDM, prevalence estimates for the disease vary from 3.8% to 16.2% across the country (Thanawala et al., 2021; Unnikrishnan et al., 2016).

It has been repeatedly shown that women of SA descent outside of SA are also at increased risk of GDM, even after emigrating and accounting for immigrant status (Menard et al., 2020; Sanchalika and Teresa, 2015). When compared to WEs, SAs living in Canada are at almost 3 times the odds (aOR = 2.88, 95% CI 1.03 – 8.07) of GDM after accounting for BMI, income, infant sex, education, parity, age, weight gain and marital status (Menard et al., 2020). Likewise, in the US, Bangladeshis, Indians, Sri Lankans, and Pakistanis (PKs) have 4.3, 3.9, 3,7 and 3.4 times the odds of GDM, respectively, after adjusting for maternal age, parity, smoking and measures of socioeconomic status, compared to WEs in the US (Sanchalika and Teresa, 2015). This trend can also be seen in Europe, where SAs account for 9.2% of UK pregnancies, but 25.6% of UK GDM diagnoses (Read et al., 2019; Greenhalgh et al., 2015).

Moreover, the absolute aggregate risk (i.e., the risk of an individual with a set of risk factors developing a disease) has increased over time in immigrant populations, highlighting the increased burden of GDM in non-white

communities (Gail, 2008). For example, in the US between 2011- 2019, the absolute aggregate risk of GDM in nulliparous individuals was highest in Asian individuals and increased significantly from 90.8 (95% CI 85.9- 95.9) per 1000 live births in 2011 to 129.1 (95% CI 124.1 - 134.2) per 1000 live births in 2019 (Shah et al., 2021). This increase resulted in Asian-Indian individuals having 2.24 times the rate (95% CI 2.15 - 2.33) compared to non-Hispanic white individuals in 2019: a larger disparity to that seen in 2011, where Asian Indian women experienced a 2.06 times greater rate of GDM (95% CI 1.94 - 2.18) compared to non-Hispanic whites. This increase in GDM in the Asian Indian strata occurred despite lower BMI levels and higher rates of educational attainment compared to non-Hispanic whites (Shah et al., 2021).

Furthermore, the difference in GDM risk between EAs and SAs is also observed in immigrant populations. In a 2021 Norwegian cohort, prevalence of GDM was higher in all immigrant populations compared to non-immigrant populations, with the highest odds of GDM being observed within SA immigrants (OR = 5.45, 95% CI 5.05 – 5.89) (Strandberg et al., 2021). In addition, odds of GDM were also increased in other immigrant populations, with the combined odds of GDM in immigrants from Southeast Asia, East Asia and Oceania immigrants and the combined odds of immigrants from North Africa and the Middle East being 4.03 and 3.09 respectively (Strandberg et al., 2021). There is also evidence to suggest that GDM is increasing at a higher rate in SA immigrant populations compared to EA immigrant populations. Compared to 2011, 2019 rates of GDM were not significantly different in the Japanese stratum of a US cohort but were significantly higher across this period in the Asian-Indian stratum (Shah et al., 2021).

In Asian immigrant populations the increased risk of GDM also remains after accounting for acculturation. Acculturation is the process of assimilating into a different culture and can involve the processes of regaining financial security, developing new social ties and habits, and embedding oneself into the new culture, including modification of diet, physical activity, and leisure activities. For example, a 2019 study from Los Angeles found that only 15.9% of the increased risk experienced by Asian immigrants could be attributed to factors related to acculturation (Chen, L. et al., 2019a). Herein, the majority of the risk remained, suggesting a role of robust biological and genetic factors

driving elevated GDM risk. Likewise, in a study of 231,618 pregnant women who immigrated to Canada after 1985, those carrying their first pregnancy were at a greater risk of GDM compared to long-term residents after accounting for confounding (RR = 1.73, 95% CI 1.66 - 1.81) (Read et al., 2021). This further suggests that ethnicity increases GDM risk in ways other than acculturation, possibly because immigrant women from minority backgrounds have more risk factors for GDM, including lower education levels, cultural obligations, lower-income, higher levels of dysglycemia preconception, and an increased likelihood of having family members with diabetes (Read et al., 2019). Taken together, these data show that SAs are at a substantially higher risk of GDM compared to WEs, independent of an individual's geographical location or immigration status. Hence, for GDM burden to be reduced globally, it is crucial that GDM prevention and intervention efforts are targeted at SAs.

Factors driving these ethnic disparities in GDM risk are unclear, although behavioural, cultural, and biological factors are thought to be involved (Read et al., 2019; McDonald et al., 2015). Despite known differences in adiposity between WEs and SAs, with SAs having a higher level of body fat at a lower BMI compared to WEs, it has been demonstrated that adiposity alone is not responsible for the increased risk of GDM in SAs (Read et al., 2021; McIntyre et al., 2019; Chong et al., 2014). In addition, even at a lower BMIs, SA women are at a greater risk of GDM compared to WEs. In a Canadian prospective cohort, GDM prevalence exceeded 5% in SAs when the average BMI was 21.5 $kg/m^2$ (i.e., a healthy BMI), while in the general population, the prevalence only reached 5% when the average BMI reached 29.5 $kg/m^2$ (i.e., a near obese BMI) (Read et al., 2021). Likewise, it was shown that in a multi-ethnic UK population the risk of GDM in SAs and EAs at a BMI of 21 $kg/m^2$ was equivalent to the risk of GDM experienced by WE women with a BMI of 30 $kg/m^2$ (Nishikawa et al., 2017).

In addition to BMI, several physiological drivers have been identified which could increase GDM risk, but as with BMI no factor has been found to be responsible for the vast disparity in prevalence estimates observed between the ethnicities. These drivers include increased insulin resistance during pregnancy in SAs compared to WEs (after accounting for age, pregnancy

weight gain and a history of diabetes) and an increased likelihood of a SA newborn baby being born small for gestational age while also having increased adiposity levels, which could increase their own future GDM risk (McDonald et al., 2015; Menard et al., 2020; Lorenzo-Almorós et al., 2019; McIntyre et al., 2019). In summary, although numerous factors are suspected to increase GDM risk in SAs, no individual factor has been identified to markedly contribute to the disparity in risk observed between the ethnicities, meaning additional avenues need to be explored to understand this increased risk.

#### 1.1.3.1 Impact of GDM in non-white ethnicities

In addition to being at an increased risk for GDM, non-white ethnicities are also more likely to experience detrimental health outcomes following their GDM diagnosis. Non-white ethnicities were found to have a higher rate of GDM reoccurrence compared to WEs in a 2015 meta-analysis: 38% (95% CI 33% - 44%) of WEs experienced GDM reoccurrence compared to 55% (95% CI 44% - 68%) of non-WEs (Schwartz et al., 2015). When non-White ethnicities were stratified into distinct ethnic groups, Hispanic women were found to have the highest rate of GDM reoccurrence (59%, 95% CI 38% - 80%). However, heterogeneity was greater in Hispanics compared to Asian women, who experienced a slightly lower reoccurrence rate (54%, 95% CI 34% - 74%) but a lower heterogeneity ($I^2$ = 57.2% in SAs vs 96.1% in Hispanics), indicating a greater certainty regarding the rate of reoccurrence in SAs (Schwartz et al., 2015).

Furthermore, not only do non-white individuals have an increased risk of GDM reoccurrence, but they are also at increased risk of T2D following their GDM diagnosis (30). A 2020 meta-analysis found the cumulative incidence of T2D to be 15.4% (95% CI 13.3% – 17.9%) in non-white ethnicities compared to 9.9% (95% CI 9.4% - 10.4%) in white ethnicities, despite the fact that mean follow up was shorter in WE studies (7.28 years vs 6 years) (Vounzoulaki et al., 2020). This estimate is slightly lower than that obtained in a separate meta-analysis of SA women (from both western countries and the Asian sub-continent), which found the cumulative incidence of T2D to be 17.3% (95% CI 12 – 23.8) at 5 years and 33% (95% 31.3 – 34.8) at 10 years post GDM

diagnoses (Gadve et al., 2021). This study found previous SA GDM cases to be at 10.8 times the risk of developing T2D (95% CI 7.6- 15.4) compared to SA women without GDM. In summary, this evidence suggests that SAs are at a greater risk of GDM complications compared to WEs, with a stronger evidence base supporting this increased risk in SAs compared to other non-white ethnic groups.

### 1.1.4 Diet and GDM

Consuming a healthy diet during pregnancy is important for the health of both mother and child. The WHO defines a healthy diet during pregnancy as one that "contains adequate energy, proteins, vitamins, and minerals, obtained through the consumption of a variety of foods, including green and orange variables, meat, fish, beans, nuts and pasteurized dairy products and fruits" (Super and Wagemakers, 2021). A similar but more detailed definition of a healthy diet during pregnancy is provided by the UK's National Health Service (NHS), who state that a healthy diet during pregnancy does not need to be 'special' but should be balanced (NHS, 2020). In addition, NHS guidelines suggest that pregnancy diets should be high in fibre and over a third of the diet should be comprised of carbohydrates, which should ideally be wholegrains and not refined carbohydrates. These guidelines also suggest limiting oily fish to two portions week (because fish can contain pollutants) and limiting the intake of unpasteurised cheese and saturated fats. Finally, low-fat dairy varieties are recommended as a replacement of other dairy products. Interestingly, although some expectant mothers may unknowingly not satisfy current nutritional guidelines, women may be more motivated to engage in diet modification during pregnancy due to the desire to foster a heathy baby (Nielsen et al., 2018; Super and Wagemakers, 2021; Zinsser et al., 2020; Herring et al., 2012). Furthermore, as pregnancy itself is a risk factor for new or persistent obesity, providing nutritional intervention during pregnancy can mitigate risk of maternal obesity as well as improving offspring health and health behaviours of the whole family unit (Menard et al., 2020).

Given the clear link between diet and glycaemia, promotion of healthy dietary guidelines during pregnancy is at the forefront for GDM prevention strategy globally (Newman and Dunne, 2021; McIntyre et al., 2019). Dietary

guidelines for GDM prevention are similar to generic guidelines for a healthy diet during pregnancy, however GDM guidelines have a greater focus on glycaemic control. For example, the NHS states that during pregnancy low glycaemic foods should be consumed three times a day alongside plentiful fruit and vegetables, while the intake of sugary foods and drinks should be limited to aid in the prevention of GDM (NHS, 2019). Very similar advice is also provided by Australian and American diabetes organisations (Diabetes Australia, 2021; UCSF Health, 2022).

In addition to dietary guidelines, pharmaceutical methods for preventing GDM are utilised globally; however dietary guidelines remain the preferential treatment due to potential side-effects of pharmaceuticals. For example, Metformin (a glucose lowering drug) may reduce birth weight while simultaneously accelerating offspring growth in later childhood, contributing to a reluctance to prescribe pharmaceuticals during pregnancy unless essential (Tarry-Adkins et al., 2021). Furthermore, the evidence base supporting the effectiveness of metformin in preventing GDM is still debated, with a 2021 meta-analysis concluding that Metformin may not prevent GDM in 46% of women (Tarry-Adkins et al., 2021). Due to these potential side effects and unknown effectiveness of therapeutic interventions in preventing GDM as well as an increased motivation to make lifestyle changes during pregnancy, dietary interventions are the preferred method of GDM prevention and treatment. However, for these dietary interventions to be effective, it is crucial that dietary interventions targeted at non-white ethnicities are culturally sensitive, considering individual preferences, cultural beliefs, and lifestyles, between and in ethnic groups (Super and Wagemakers, 2021; Bandyopadhyay, 2021; Valentini et al., 2012). To summarise, dietary guidelines focusing on the consumption of a healthy diet during pregnancy are at the forefront of GDM prevention, which could lead to long-term health benefits as a result of individuals being more amendable to lifestyle changes during pregnancy if these guidelines are effective.

#### 1.1.4.1 Evidence from Randomised Controlled Trials (RCTs).

Although healthy dietary guidelines are embedded in routine clinical care, current evidence regarding the effectiveness of such interventions for

GDM prevention across global populations is inconsistent and uncertain (McIntyre et al., 2019; Griffith et al., 2020). The most recent Cochrane Review (2020) evaluated methods of GDM prevention failed to identify a single lifestyle or medical intervention that significantly prevented GDM, and concluded the effect of diet on GDM prevention as being of 'unknown benefit or harm' (Griffith et al., 2020). Moreover, the quality of the evidence upon which 'healthy dietary' guidelines were based was found to be 'very-low' while the evidence supporting 'low-glycaemic index' diets was found to be 'low'. Interestingly, a recent (2022) systematic review of interventions reported that diet was effective against GDM (OR: 0.61, 95% CI 0.45 – 0.82) but was limited only to studies/diets focussed on the management of gestational weight gain (GWG) (Teede et al., 2022). This suggests that a relationship between diet and GDM exists but that it may be best mediated through gestational weight gain rather than attempting to directly moderate dysglycemia. Importantly, in these reviews, very few included studies with majority non-WE populations and none presented analyses stratified by ethnicity, despite non-white populations being at highest risk of GDM and have differing dietary habits. Indeed, throughout the field of non-communicable disease epidemiology, nutritional approaches aimed at reducing disease incidence rarely considers ethnic variability in food metabolism and dietary habits (Franzago et al., 2020). Currently the evidence does not confidently demonstrate the effectiveness of dietary interventions commonly tested in WE populations on GDM prevention in non-WEs (McIntyre et al., 2019).

To date, all meta-analyses evaluating the effectiveness of healthy dietary interventions in preventing GDM have focussed on randomised control trials (RCTs). RCTs are considered the 'gold standard' of epidemiological study designs. When recruitment and randomisation are carefully considered, RCTs allow causality to be inferred to an extent that is not possible in other designs more susceptible to confounding and biases (Hariton and Locascio, 2018). Despite these strengths, RCTs also have their limitations: (i) they are costly which often restricts both how long the study can last and the sample size; (ii) they are often not generalizable due to limited sample size and restricted recruitment strategies (Hariton and Locascio, 2018); (iii) it is not often possible to assess the long-term effect of a diet on a health outcome

(Davies, G. et al., 2018); (iv) it is not possible to assess the impact of non-healthy diets on a disease due to ethical concerns.

In summary, out of all epidemiological study designs RCTs are the gold standard; when carefully conducted, the randomisation process allows for associations to be determined while confounding is simultaneously minimised meaning the true effect of an intervention can be more accurately determined. Despite this clear strength, the practicality of performing RCTs over long-time periods means it is often not possible to determine the long-term effect of diet on an outcome in a randomised setting, particularly in large samples. Instead, an alternative approach involving observational studies needs to be utilised.

### 1.1.4.2 Evidence from observational studies

Observational studies aim to estimate the impact of an exposure on an outcome without intervening on a participants lifestyle or behaviours. Although observational studies are susceptible to a range of biases (including recall bias and performance bias) and confounding, they can be conducted in much larger samples and over a more prolonged period of time than RCTs, thereby improving the generalisability of findings (Viswanathan et al., 2013). Additionally, because they can be prospective in nature, the effects of self-reported habitual 'healthy' or 'unhealthy' dietary habits on disease risk can both be evaluated.

To date, one systematic review has been conducted to summarise current evidence regarding the impact of dietary patterns on GDM in an observational setting, but no meta-analyses were performed. This review reported that increased consumption of fat (1-5% of energy intake of carbohydrates replaced with fat), cholesterol (≥300 mg/day), heme iron (≥ 1.1 mg/day), red/processed meat (1 serving/day) and egg (≥ 7 per week) were associated with increased risk of GDM risk (Schoenaker et al., 2016). Conversely, diets high in fruit, vegetables, whole grains, and high-fat dairy protected against GDM. Nevertheless, the authors acknowledge some limitations of this study. Firstly, results were not stratified by ethnicity and the majority of studies included in the analysis focussed on a single cohort (Nurses' Health Study II); therefore, the results may not be generalisable.

Secondly, both cohort and case-control studies were included meaning temporality could not be inferred.

When considering data solely from the Nurses' Health Study II, it has been suggested that 45% of GDM cases are potentially preventable through the uptake of a healthy diet alongside additional lifestyle modifications, such as smoking cessation and ≥ 30 mins/day of exercise, although it is not known whether these lifestyle changes would need to be adopted before pregnancy (McIntyre et al., 2019; Zhang et al., 2014). In addition, the limited number of non-white individuals in the Nurses' Health Study II means that future studies are needed to analyse the relationship between lifestyle modification and GDM prevention in non-WE populations to determine if the association is generalisable (Zhang et al., 2014). In fact, various discrepancies exist in the literature regarding the variation in magnitude and direction of effect sizes of pre-pregnancy dietary patterns on GDM between ethnicities. For example, in a WE cohort, a meat-based dietary pattern increased the odds of GDM (OR 1.38, 95% CI 1.14 – 1.68) while a comparable diet in a Chinese cohort reported no association (OR 0.89, 95% CI 0.58 – 1.36) (Mak et al., 2018; Bao et al., 2013). This absence of association in an Asian population could be due to numerous factors.

Firstly, even if dietary patterns are similar between populations, differences in the type of meat product consumed along with preparation and cooking methods may vary between ethnicities. This supports the notion that for dietary assessment to be effective in non-white populations, culturally sensitive assessment tools need to be utilised that account for these factors (Super and Wagemakers, 2021; Bandyopadhyay, 2021; Valentini et al., 2012). Secondly, it is possible that either the demographic characteristics of the two cohorts are not comparable, or that there is insufficient statistical power in the smaller Asian population to detect any effect. However, it is also possible that biological differences exist between the ethnicities which result in dietary patterns having differential effects. Indeed, there are multiple examples in the literature identifying ethnicity as a mediator of associations between diet and disease outcomes. For instance, associations between the Alternate Healthy Eating Index (AHEI) and diabetes in postmenopausal women have been observed in WEs but not in individuals of Asian or African descent (Qiao et

al., 2014) whereas the Mediterranean diet has been associated with a lower risk of T2D in WEs but not in Japanese-American or Native Hawaiian individuals (Sotos-Prieto and Mattei, 2018). Additionally, the multi-ethnic NUTRIGEN consortium reported that maternal adherence to a plant-based diet during pregnancy was associated with increased birth weight in SAs but reduced birth weight in WEs (Zulyniak et al., 2017).

Taken together, this evidence shows that, although diet has been found to be associated with GDM in observational settings, how this association differs between ethnicities has not yet been assessed. Given that diet has been shown to have varying impacts on other disease outcomes across ethnicities, and that different ethnicities consume different foodstuffs and prepare food in different ways, it is possible that the impact of diet on GDM may be distinct between different ethnic groups.

#### 1.1.4.3 Importance of culturally sensitive interventions

Factors responsible for the ethnic-specific differences in the associations between diet and disease outcomes assessed in the observational setting are still unknown. Although they may be a result of biological factors and/or differences in dietary consumption between populations, differenced may also be a result of methodological limitations. Due to the differences in dietary consumption between populations, it is possible that the tools used to assess the intake of dietary patterns typically characterised in WEs (e.g., the Mediterranean diet) do not directly translate to other ethnicities (Kunasegaran et al., 2021; Sotos-Prieto and Mattei, 2018; Mchiza et al., 2010). Likewise, the methodological decision to broadly categorise ethnic groups (e.g., ‘Asian’ rather than SA or EA), often as a result of smaller sample sizes and numbers, can obscure associations in individual ethnic sub-groups due to differing socio-economic, lifestyle or dietary habits in immigrant populations. Furthermore, considerable differences exist in how medical nutritional therapy is utilised in different cultures which may impact the success of dietary interventions. It has previously been shown that dietary interventions or nutritional guidelines prescribed in predominantly WE countries may not take into account cultural differences or individual preferences between the WE community and immigrant communities which can limit the success of the

intervention in immigrant populations (Yuen et al., 2018; Read et al., 2019; Bandyopadhyay, 2021; Super and Wagemakers, 2021; Valentini et al., 2012). For example, immigrant women can often be reluctant to give up foods traditional to their culture or consume different foods to which they are used to and may also lack the skills required to prepare the foods in alternative ways. These foods are often vegetarian, and higher in carbohydrates (often in the form of roti and rice) and fats due to preparation methods using oil and ghee meaning they can negatively impact glycaemic control. First-generation immigrants may also be less likely to have a family support network throughout their pregnancy to help them engage in healthy behaviours and dietary changes (Kandasamy et al., 2021). Even when these support networks do exist women often report prioritising the needs of family members over their own health needs and can also feel pressure from family members to consume traditional foods that negatively impact glycaemic control, particularly at social gatherings (Bandyopadhyay, 2021; Kandasamy et al., 2021). These social pressures can also influence mealtimes, which can further hinder glycaemic control, as can Ramadan for SAs who are of the Islamic faith, even though pregnant women are excused from fasting (Yuen et al., 2018).

Although difficulties in balancing work and family responsibilities during pregnancy are likely experienced by most women, advice from family members that conflicts with that of medical professionals has been identified as a particular issue for SAs (Kandasamy et al., 2021; Read et al., 2019; Nielsen et al., 2018; Thanawala et al., 2021). In SA communities, it has been reported that advice from female family members often emphasises the need to increase food consumption during pregnancy to ‘eat for two’, while also decreasing physical activity to facilitate the baby’s growth (Kandasamy et al., 2021). Furthermore, beliefs surrounding the idea that exercise can increase the risk of miscarriage are often commonly reported by this community. (Kandasamy et al., 2021; Read et al., 2019; Nielsen et al., 2018; Thanawala et al., 2021). Regarding the ‘eating for two’ belief, it has been found that when women did restrict their food intake, this was often linked to negative pregnancy symptoms, including headaches, fainting and lack of energy, with food being viewed as medication: a belief that can be counter-productive when

trying to regulate glycaemic levels (Kandasamy et al., 2021). This can be particularly problematic as SA women can be more reluctant to share sensitive information with their healthcare provider, which can hinder the flow of information between both parties and hence prevent the uptake of those dietary modifications needed to improve glycaemia (Kandasamy et al., 2021). It is also important to note that in western countries women from minority ethnic backgrounds are more likely to experience systemic barriers that prevent them from maintaining a healthy pregnancy and can impact a woman's ability to consume a healthy diet during pregnancy (Phonyiam and Berry, 2021).

These cultural barriers may be responsible for some of the differences in effects seen in the literature regarding diet and GDM development. To begin to assess this, the effect of diet on GDM development in distinct ethnic groups is required to investigate inconsistencies, compare the effect of diet pre- and post-conception, and offer direction. If ethnic-specific associations between diet and GDM are demonstrated, it would support the use of maternal dietary guidelines that are tailored to a mother's ethnicity in GDM prevention. Moreover, if diet is shown to have a differential impact on GDM between ethnic groups, it may shed light on biological pathways that contribute to the disparity in GDM risk between the ethnicities. However, to draw conclusions regarding the absence/presence of an association between diet and GDM in SAs it is crucial that SA women can engage fully in nutritional care and that culturally sensitivity interventions and assessment methods are utilised, otherwise these factors could confound the true relationship between diet and GDM in SAs.

## 1.2 Metabolomics

One potential biological factor that may contribute to the disparity in GDM risk between SAs and WEs is differences in underlying metabolism. The metabolome, defined as 'the comprehensive analysis of metabolites in a biological specimen', provides a snapshot of the biological molecules in the individual at the time of sample collection and can provide insights into the biological mechanisms of metabolic diseases (Wang et al., 2021).

Downstream from the genotype, transcriptome and proteome, the metabolome reflects both genetic and environmental factors making it particularly valuable when investigating the relationship between diet and disease progression (McIntyre et al., 2019; McIntyre et al., 2020). Indeed, the plasma metabolome is known to be influenced by a range of factors including diet, environmental exposures, medication, an individual's health status and the microbiome (Li et al., 2020). As a result of this, metabolomics (the study of metabolites) is an ideal tool to study GDM, a disease linked to metabolomic dysregulation. Indeed, in 2018 the Diabetic Pregnancy Study Group specifically called for more research into the metabolome in relation to GDM (Wang et al., 2021; Kuller and Catov, 2017; Clish, 2015; Schaefer-Graf et al., 2018). Furthermore, the group also empathised the importance of considering how GDM risk factors influence the metabolome, including ethnicity (Schaefer-Graf et al., 2018).

There are numerous advantages to using metabolomics to study disease progression. Firstly, the rapid development and decreasing costs of quantification techniques now permit larger metabolite panels to be examined concurrently, facilitating the advancement of biomarker discovery and precision medicine (Kettunen et al., 2012). As the metabolome is downstream of the genome, changes in the genome are magnified at the level of the metabolome, making it easier to detect biological changes, which can also show temporal tissue specificity and provide more detailed insight into disease mechanisms (Clish, 2015). Additionally, as there are far fewer metabolites (~ 4,000 metabolites in human serum) compared to genes (~30,000), mRNA, and proteins (~90,000), it is more manageable and easier to identify unique features and responses (Dalfrà et al., 2020; Wang et al., 2021; NIH, 2020a; NIH, 2020b). Numerous sample types can be examined via metabolomics (including urine, blood, amniotic fluid, and breast milk) via a range of quantification techniques (including nuclear magnetic resonance (NMR), mass spectroscopy, liquid chromatography-mass spectroscopy, and gas chromatography-mass spectroscopy) depending on the sample and metabolites of interest. As such, metabolomics is a valuable tool to investigate disease aetiology and identify early metabolic perturbations (Franzago et al., 2020; Wang and Hu, 2018).

In addition, metabolomics can aid in the development of precision nutrition, particularly when combined with other 'omics techniques, including proteomics and transcriptomics (de Toro-Martin et al., 2017; Wang and Hu, 2018). The goal of precision nutrition is to facilitate the development of tailored and dynamic dietary guidelines recommended to an individual (or group) based upon internal and external environmental exposures and demographic factors, such as age, sex and ethnicity (de Toro-Martin et al., 2017). Although yet to be realised, it has been proposed that customizing dietary interventions to an individual's specific needs may be cost-effective when compared to generalised interventions (Tebani and Bekri, 2019). Taken together, when considering both the increasing ease of quantifying metabolites in biological samples, their use in precision medicine and the known metabolic dysregulation that characterises a GDM diagnosis, investigations into the metabolome have the potential to shed light on the disparities in GDM risk seen between WEs and SAs and may allow for a better understanding of GDM pathology and preventative options.

### 1.2.1 Diet and the metabolome

Diet has been shown to influence serum metabolites more robustly than clinical parameters (age and sex), the microbiome, diurnal variation, and other lifestyle factors (e.g., smoking, stress, exercise, and sleep) (Bar et al., 2020). In total, dietary factors explain almost 50% of the observed variation in serum metabolite levels. In light of this, distinct metabolite signatures have been identified that are associated with common dietary patterns. For example, a 2020 study identified a metabolomic signature comprised of 67 metabolites that characterised adherence to a Mediterranean diet in a cohort of Spanish and US participants (Li et al., 2020). Unsurprisingly, this signature was predominantly characterised by metabolites involved in the metabolism of polyunsaturated fatty acids (PUFAs) and lipids, key components of the Mediterranean diet. In addition to being associated with the Mediterranean dietary pattern, this signature was also associated with genetic loci linked to fatty acid metabolism and CVD risk, highlighting a link between diet, the metabolome, the genome, and disease risk (Li et al., 2020). Moreover, it has been established that an individual's metabolome can alter both their dietary

requirements and their response to diet (Tebani and Bekri, 2019). Therefore, the metabolome offers a unique opportunity to investigate why the effect of diet on GDM is heterogenous in nature as well as potentially providing an opportunity to prevent GDM by allowing for the development of tailored dietary interventions and advances in the field of precision nutrition.

### 1.2.2 GDM and the metabolome

Numerous metabolites have been associated with GDM, however few reproducible results have been obtained. In a review of 9 studies that investigated associations between metabolites and GDM, numerous metabolites assessed in the blood (plasma or serum) were not replicated in more than one study, sometimes due to identified associations not being replicated across studies or due to the fact that individual metabolites were not analysed in more than one study (Chen et al., 2018). Examples of metabolites that have been associated with GDM in serum samples include lipoproteins, amino acids, lactate, pyruvate, carbohydrates, carboxylic acids, alcohols, bile acids and various organic compounds, although often these were only identified in a single cohort (Wang et al., 2021; Clish, 2015; Mauro et al., 2022). A summary of the current evidence regarding the role of macronutrients in GDM development can be found in **Table 1.1**.

Numerous factors likely contribute to poor reproducibility between metabolomics studies, including small sample sizes, heterogeneous methods of sample quantification, varied sample types, varied metabolite panels, differences in time of sample collection (in relation to both pregnancy trimester and pre/post GDM diagnosis) and population heterogeneity (Chen et al., 2018). Indeed, samples taken in early pregnancy may represent potential GDM therapeutic targets, helping the identification of biomarkers, while samples taken during late pregnancy may represent changes in the metabolome that have occurred as a result of GDM (Wang et al., 2021). Finally, levels of certain metabolites are also strongly correlated meaning that they may only impart an effect in the presence/absence of additional metabolites. For the above reasons, it is difficult to compare metabolomic results from different studies in order to investigate the role of ethnicity in the metabolome. Instead, to limit confounding a multi-ethnic population should be

utilised where samples are taken at the same time in comparable conditions, utilising multivariate statistical techniques to account for the correlation structure present in metabolomics data.

**Table 1.1: Summary of associations between macronutrients and GDM.**

| Biological molecule | Examples | Mechanism linking biological molecule to dysglycemia during pregnancy | GDM association |
|---|---|---|---|
| **Carbohydrates** | Hexose, tricarboxylic acid metabolites | Dysregulation of glycolysis<br>Promotion of lipid synthesis | Increased in GDM cases during early pregnancy and post-partum |
| **Amino Acids** | Branched chain amino acids | Increased insulin resistance<br>Inhabitation of β-cell function | Elevated in early pregnancy in GDM cases |
| **Lipids** | Fatty acid metabolites | Increased hyperglycaemia<br>Increased Insulin sensitivity<br>Increased Oxidative stress<br>Increased Inflammation | Differing effects on GDM status dependent on dosage and timing. Strongest associations observed during the 3rd trimester. |

Interestingly, only one study has examined the association of the metabolome with GDM risk in SAs (Taylor et al., 2019). In this study, multivariable statistical analyses were conducted in the multi-ethnic Born in Bradford (BiB) cohort, containing an almost equal proportion of SAs and WEs. Herein, differences in the levels of a range of metabolite values were identified when comparing SA and WE GDM cases, with SA GDM cases having higher levels of fatty acids, amino acids and glucose (Taylor et al., 2019) compared to WE GDM cases. Some evidence of ethnic-specific associations between fatty acids and GDM were also identified, with positive associations only present in WEs, although adjusting for education and parity reduced this difference. Furthermore, differences between the metabolomes of SAs and WEs during pregnancy irrespective of GDM were also observed (Taylor et al., 2019). This highlights the possibility that differences in metabolism during

pregnancy may contribute to differences in risk for GDM between ethnic groups.

### 1.2.3 Genetics and the metabolome

In addition to the environmental factors such as diet that influence the metabolome, it is also well-established that an individual's metabolite levels are partially determined by genetics (Rhee, 2020). Differences in metabolism between ethnicities may be supported by genetic differences between ethnicities. To determine differences at a genetic level, genomic data needs to be utilised. Genomics is the study of an individual's genome, its interaction with the environment, and its effect on health and disease (NIH, 2020a). The field of genomics has developed rapidly over the past few decades following the initial sequencing of the human genome and subsequent advances in sequencing technology. Through these advances, it has become possible to determine genetic risk factors for a range of diseases and to study how these risk factors interact with environmental factors to influence certain traits (Burton et al., 2014; NIH, 2020a). Furthermore, genomics can also help identify biological mechanisms associated with a disease, aiding in the development of personalised medical interventions and more advanced diagnostics and screening procedures (Burton et al., 2014). Indeed, through the use of genomics, it is possible to gain a better understanding of how an individual's genets impact their metabolism and risk of GDM.

In a genome-wide association study (GWAS) of 8,330 Finnish individuals, heritability estimates (i.e., the proportion of the variance of a trait in a population attributable to genetics) of >0.6 were found for almost 40% of tested serum metabolites (Kettunen et al., 2012; Tebani and Bekri, 2019; Visscher et al., 2008). In this study heritability estimates were highest for lipoproteins (range 0.5 - 0.76) and lipids (range 0.48 - 0.62), and were lowest for the amino acids, where estimates were often ~ 0.2 (Kettunen et al., 2012). Similar heritability estimates were identified for lipids in a 2020 meta-analysis of heritability estimates of blood metabolites (serum and blood) (Hagenbeek et al., 2020). Although the heritability of amino acids was found to be greater when considering plasma samples and additional cohorts of mixed ethnicity (median heritability of amino acids ~40%) (Hagenbeek et al., 2020). None of

the studies identified in this meta-analysis were conducted in SA populations. Therefore, differences in the heritability of metabolites could not be compared between SAs and WEs, which could help determine the extent in which genetic predictors or metabolism vary between SAs and WEs.

As a consequence of this high heritability, variation in the genome has been linked to variation in an individual's metabolome, with lipid, carbohydrate and folate metabolism all being shown to have strong genetic components (Franzago et al., 2020). This highlights the need to integrate metabolomics and genomics data to better understand the combined role of diet, ethnicity, and the metabolome in GDM development (Tebani and Bekri, 2019; Dalfrà et al., 2020). Moreover, genetic variation in metabolite levels has also been linked to historical geographical differences in dietary consumption, highlighting how historical diets have acted as a selective pressure, impacting metabolite levels in modern-day populations (Koletzko et al., 2019).

For example, variability in the *FADS* genes has been shown to influence PUFA levels (especially n-6 PUFAs) highlighting the importance of genetic variance in influencing metabolite levels (Koletzko et al., 2019). Furthermore, stark differences between the global distribution of the common *FADS* genotypes across the globe are thought to be influenced by varying dietary patterns acting as distinct selective pressures (Ameur et al., 2012). For example, only 1% of Africans have haplotype A (comprised of a distinct pattern of 28 SNPs across the *FADS* region), associated with a lower rate of long-chain PUFA (LC-PUFA) conversion. Meanwhile, 50% of EAs and >95% of Native Americans have this haplotype (Koletzko et al., 2019), Instead, African populations are more likely to have the D haplotype, associated with a higher rate of LC-PUFA conversion. These disparities in allele frequency are believed to be a result of the variations in the LC-PUFA content of historical diets resulting in distinct selective pressure on genotypes; populations that consume large quantities of LC-PUFAs (i.e., the protein and seafood-rich diets consumed by Native Americans) more likely to be homozygous for allele A, due to the lesser requirement of LC-PUFA conversion. Additional ways in which diet can impact the genome have also been suggested, including epigenetic changes such as methylation which can have upstream effects on genome instability, miRNA production and protein production (Franzago et al.,

2020; Tebani and Bekri, 2019; Dalfrà et al., 2020). This illustrates the close interplay between diet and genetics and highlights how diet can act as a distinct selection pressure to metabolite related genes, which may have an impact on disease risk. Furthermore, a 2018 study identified a positive correlation between genes associated with lipid polymorphisms and genes associated with GDM in WEs (Franzago et al., 2018). This correlation highlights how dietary metabolism may be associated with GDM risk and highlights how nutritional interventions aimed at altering levels of specific metabolites may reduce the incidence of GDM in genetically suspectable populations.

### 1.2.4 Genetics and GDM

The stark differences in GDM burden between ethnicities indicate that the development of GDM may have a genetic component. Indeed, Insulin sensitivity and secretion (key factors that drive the development of GDM) have heritability estimates of 53% and 75% respectively indicating that these factors are greatly influenced by genetics (Ding et al., 2018). Differences in genetic risk factors for GDM have also been reported between populations, for example, a 2018 meta-analysis found the *GCKR* gene to be associated with an increased risk of GDM in WEs but not in African or Korean populations (Guo et al., 2018). Additionally, a 2020 meta-analysis found adiponectin, a hormone that contributes to the regulation of carbohydrate and fatty acid metabolism, to increase the risk of GDM in Asians and WEs but to decrease the risk in North American and South American populations (Bai et al., 2020), indicating that biological molecules closely related to metabolism may have an ethnic-specific impact on GDM development. To date, no study has directly utilised genomic data to establish how metabolites impact GDM in an ethnic-specific manner. Through the determination of ethnic-specific metabolite-gene associations it may be possible to highlight distinct metabolic profiles that drive GDM in SAs compared to WEs, while also allowing for causal associations to be tested via Mendelian Randomisation (MR); an instrumental variable (IV) approach that utilises genetic data as proxies for an exposure to test for associations in the absence of confounding.

## 1.3 Mendelian Randomisation

Traditional observational studies are subjected to confounding, both known and unknown, that cannot be fully accounted for. Therefore, in an observational setting, even when an exposure is consistently associated with an outcome, the presence of causality is not assured. Biases imparted on associations as a result of confounding are especially problematic for both dietary exposures and pregnancy-related health outcomes due to the large number of factors that influence diet, maternal health and offspring health; including genetics, lifestyle factors and socioeconomic status (Diemer et al., 2021).

Over the last decade to resolve the issue of confounding in observational studies the IV approach of MR has been adopted in the field of genetic epidemiology where a single or set of genetic variables act as a proxy for the exposure. MR biologically mimics a RCT as an individual's genetic variants are randomised during independent assortment in meiotic cell division and fertilisation, in the absence of confounders (Sheehan et al., 2011). MR exploits this non-confounded relationship by utilising those SNPs associated with the exposure of interest as a genetic proxy for the exposure. In this way the SNPs are employed as IVs. This means that the lifetime effect of an exposure on an outcome, or the effect of an exposure at a distinct period of life, can be assessed in the absence of confounding (Sheehan et al., 2011; Liu et al., 2017; Carreras-Torres et al., 2017). For these causal associations to be valid, the selected IVs must satisfy 3 assumptions: (i) they are associated with the exposure, (ii) they are not associated with potential confounders, and (iii) there are no alternative pathways that link the IV-exposure to the outcome (iii) (Sheehan et al., 2011; Liu et al., 2017; Carreras-Torres et al., 2017). These assumptions are represented in **Figure 1:2**.

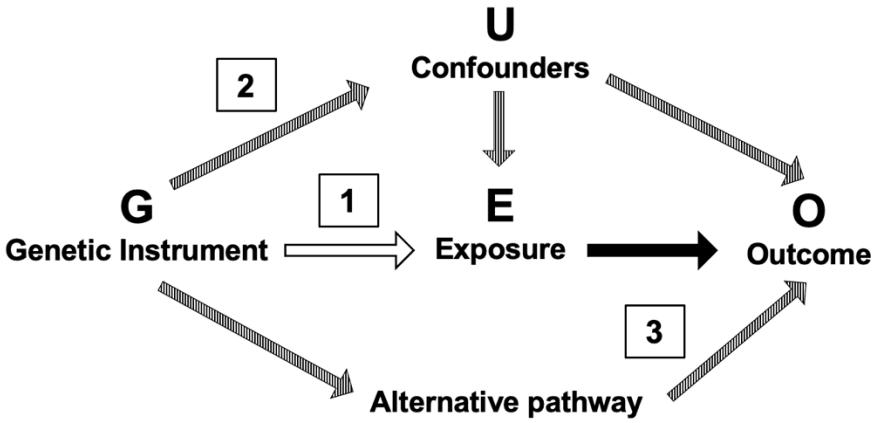



**Figure 1.2: Directed acyclic graph showing the MR assumptions.** Bold arrow illustrates investigated association. 1st assumption: Genetic instruments must be associated with the exposure. 2nd assumption: IVs must not be associated with a confounder. 3rd assumption: IVs must not be associated with the outcome via any pathway other than through the exposure. Figure adapted from Davey Smith and Hemani, 2014.(Davey Smith and Hemani, 2014).

To date no MR study has been performed with either GDM or postprandial glucose measures as the outcome (Diemer et al., 2021), likely due to the limited number of published GWASs for these outcome variables. Furthermore, few MR studies have been performed in SA populations, likely due to the relatively lower proportion of non-WE GWAS data compared to WE data and the abundance of WE data in publicly available resources often utilised for MR. In 2016, only 19% of all DNA samples taken were from individuals of non-WE descent, 14% of which were obtained from EAs. (Cooke et al., 2020; Popejoy and Fullerton, 2016). Despite these limited data, MR has been used to highlight ethnic-specific causal associations in a SA population. A 2021 study found that the loss of function variant rs138326449 in the *APOC3* gene (associated with circulating triglyceride levels) was causally associated with coronary artery disease in WEs but not in a SA population, although how this difference could impact the prevention or treatment of coronary artery disease was not explored (Goyal et al., 2021). This study is an example of how performing MR in distinct ethnicities can make it possible to infer whether differences in causality between ethnic groups exist and how metabolites may be causally associated with disease outcomes in an ethnic-

specific manner. In turn, this highlights the potential of utilising MR to identify ethnic-specific associations between metabolites and GDM.

The most common type of MR design utilised is a two-sample design. The use of a two-sample MR design requires summary level data, typically from previously published GWAS studies. However, despite the popularity of GWASs, only two published GWAS studies have been conducted with GDM as an outcome to date, both of which were conducted in EA populations (one in a population of 1710 overweight Korean women and the other in a smaller Chinese cohort (n=218)) (Tam et al., 2019; Kwak et al., 2012). In the absence of published GWAS studies in both WE and SA populations, a one-sample study design (that relies on individual level data) in an ethnically diverse cohort with genomic and metabolomic data and information on GDM status (i.e., the outcome) is required to investigate the causal role of ethnic-specific metabolic perturbations on GDM. An example of such a cohort is the multi-ethnic Born in Bradford (BiB) cohort.

## 1.4 Born in Bradford

BiB is a large multi-ethnic prospective longitudinal birth cohort, established in 2007 and based in Bradford, a city in the northeast of England. Bradford is also one of the most deprived cities in the UK, with 60% of babies born in babies being in the poorest 20% of the population of England and Wales (Wright et al., 2013). BiB was established with the aim of characterising maternal, foetal and childhood disease determinants in a multi-ethnic, economically deprived population. Bradford has one of the largest populations of SAs and the largest proportion of PKs individuals in the UK, meaning the rates of GDM in Bradford are also high, making it an ideal cohort to study determinants of GDM. (Wright et al., 2013; Bradford Metropolitan District Council, 2022).

The cohort aimed to recruit all mothers giving birth at the Bradford Royal Infirmary (the largest hospital in Bradford) between March 2007 and December 2010 (Wright et al., 2013). In total, 13,858 babies (including stillbirths and multiple pregnancies) from 12,453 mothers, ~45% of PK descent, are included in the cohort and continue to be monitored to this day. Upon recruitment detailed questionnaires were completed, permission to

access medical records granted, and over 250,000 fasting serum blood and urine samples were taken and stored, with further biological samples being taken after 2010 (Wright et al., 2013). Following fasting serum sample collection, a panel of 227 metabolite values was quantified by high-throughput NMR (Nightingale Health © (Helsinki, Finland)) (Taylor et al., 2019). DNA was also collected from participants and genotyping was conducted on two genome-wide arrays, one containing ~550K SNPs and the other ~640K SNPs (Arciero et al., 2021). The richness of the data and the large proportion of SA individuals makes the BIB cohort the ideal setting to investigate differences in risk factors and disease outcomes between SA and WE women.

## 1.5 Summary

GDM places a large burden on healthcare systems due to its high prevalence and associated health complications for both mother and child. Metabolism is thought to play a key role in determining GDM risk, although the underlying mechanisms behind these associations are unclear and reproducibility in the field is limited. Furthermore, little is known about how ethnic-specific metabolic characteristics may be associated with increased GDM risk. Through the integration of genomics and metabolomics data, it may be possible to achieve a better understanding of how metabolism impacts GDM risk through the utilisation of MR, a causal inference technique. BiB, a large multi-ethnic birth cohort with a large proportion of SA mothers and offspring, provides an ideal setting to test whether ethnics-specific differences in metabolism contribute to the elevated risk for GDM experienced by SAs. Furthermore, because metabolites can be moderated by diet, a better understanding of the metabolites that are associated with GDM may aid in the future development of more effective intervention strategies to reduce GDM risk.

This thesis aims to contribute to these evidence gaps and aid in the understanding of the ethnic specific role of metabolism in characterising GDM by i) evaluating the current evidence base regarding the role of diet in GDM prevention in distinct ethnicities, ii) determining metabolite profiles characteristic of GDM in SAs and WEs, and iii) assessing the causality of relationships between pregnancy metabolites and GDM in SAs and WEs.

# Chapter 2: Aims and Hypothesis

## 2.1 Hypothesis

SA women are at 3 times the risk of GDM compared to WEs yet the factors driving this discrepancy are unclear (Menard et al., 2020). Both pregnancy and GDM are characterised by metabolic dysregulation meaning that ethnic differences in metabolism may be partially responsible for the disparity in risk experienced between WEs and SAs. This research project aims to test the hypothesis that ethnic differences in metabolism are causally associated with GDM.

## 2.2 Aims

### Aim 1 (Chapter 3)

1 To determine whether common dietary patterns are associated with GDM incidence similarly across ethnic populations before and during pregnancy.

### Aim 2 (Chapter 4)

2 To investigate whether metabolite profiles during pregnancy differ between SAs and WEs.

2.a To examine whether metabolic profiles associated with GDM differ between SAs and WEs.

### Aim 3 (Chapter 5)

3 To assess whether metabolites are causally associated with postprandial glucose measures in SAs and WEs.

3.a To identify genetic variants associated with serum metabolites in SAs and WEs.

# Chapter 3: Ethnic-specific Associations Between Diet and Gestational Diabetes Mellitus Incidence: a Meta-Analysis

Authors original report of the study published as Fuller H, et al. PLOS Global Public Health. 2022. https://doi.org/10.1371/journal.pgph.0000250.

## 3.1 Abstract

GDM is the most common pregnancy complication to occur worldwide and confers short-term and long-term health risks on both the mother and the child. Although the implementation of healthy dietary interventions during pregnancy is the cornerstone of GDM prevention, the evidence supporting the effectiveness of these interventions in non-white ethnicities is lacking. To better inform prevention strategies, this review aims to summarise the effects of unhealthy and healthy diets on GDM risk in distinct ethnic groups.

PubMed, Scopus, Cochrane and OVID were systematically searched to identify randomised controlled trials (RCTs) and observational studies that investigated diet and GDM. Grouped analyses of common 'healthy' and 'unhealthy' diets were performed first, before analysing individual dietary patterns. Random effect models, sensitivity analyses and dose-response analyses were performed where possible (PROSPERO: CRD42019140873).

Thirty-eight publications encompassing 5 population groups (white European, Asian, Iranian, Mediterranean and Australian Nationals) were included in this review. No associations were identified between healthy diets and GDM incidence in RCTs in any ethnicity. However, in observational studies, healthy diets were found to reduce the odds of GDM by 24% (OR=0.76, 95% CI 0.70 - 0.89, p value <0.0001, $I^2$ = 75%), while unhealthy diets were found to increase GDM odds by 59% (OR= 1.59, 95% CI 1.41 – 1.81, p value <0.0001, $I^2$=0) in WEs. No evidence of consistent effects in other ethnicities were observed, despite adequate power (≥80%) to detect the magnitude of effects identified in WEs in Asians. In conclusion, pre-pregnancy diet was consistently associated with GDM risk in WEs only, despite non-white ethnicities being at an increased risk of GDM and the common utilisation of culturally sensitive assessment tools in non-white populations.

## 3.2 Background

Globally, healthy dietary recommendations are the most common tool utilised in the prevention and treatment of GDM. Despite this, the evidence of the effectiveness of dietary interventions in preventing GDM is limited, with a 2020 Cochrane review concluding that the evidence supporting the association between diet and GDM was of 'low quality', and that there was an 'unknown benefit or harm' of dietary interventions on GDM (Griffith et al., 2020). However, this review did not investigate whether ethnicity impacts on the association between diet and GDM, despite the substantial variation in GDM prevalence between ethnicities (McIntyre et al., 2019). Likewise, a 2018 systematic review found fewer than 50% of included RCTs mentioned the ethnicity of participants (Yamamoto et al., 2018), while other previously published reviews had failed to report the ethnic-specific effects of dietary intervention on GDM development (Griffith et al., 2020; Schoenaker et al., 2016; Mijatovic-Vukas et al., 2018; Guo et al., 2019).

Data supporting the association between diet and GDM obtained in observational settings are also limited, despite the fact that observational studies can capture the long-term effects of an exposure in typically larger populations compared to RCTs. A single systematic review of observational studies exploring the association between diet and GDM was conducted in 2016 (Schoenaker et al., 2016). The authors found evidence of associations between a range of dietary factors and GDM, including positive associations with heme and cholesterol and negative associations with vegetables and refined grains. However, in this study, no meta-analysis was conducted so the effects of these dietary factors were not quantified. Furthermore, their results were not stratified by ethnicity. It is possible that the association between diet and GDM may differ between ethnicities, either due to the cultural appropriateness of dietary measurement tools used or variation in biological effects between the ethnicities. To assess this, and help aid in the development of ethnic-specific GDM prevention strategies, in this chapter I aimed to evaluate the impact of diet on GDM incidence in distinct populations through a systematic review and meta-analyses of both RCTs and observational studies.

## 3.3 Hypotheses

i. Associations between diet and GDM are ethnic specific and could partially explain the disparities in disease risk observed between different ethnicities.

ii. Associations between diet and GDM incidence assessed in observational studies will be stronger than those assessed in randomised trials due to the larger sample sizes and typically longer time spans of observational studies.

## 3.4 Methods

### 3.4.1 Search strategy

The Ovid (AMID, CAB abstracts, EBM, EMBASE, Global Health, Health Care Management Information Consortium, MIDRIS, OVID Medline R), Cochrane (including trial registries), Scopus and PubMed databases were searched from inception until 31st January 2021. Where possible, databases were limited to original human-based studies written in the English language. Searches were structured using PICO (Population, Intervention, Comparison, Outcome) and MESH indexing, and included key terms (and synonyms thereof) for pregnancy (P), diet (I), ethnicity (C) and gestational diabetes (O); along with the terms for the included study designs as shown below:

*'(Gestational diabetes OR "Diabetes, Gestational/diagnosis"[Mesh] OR "Diabetes, Gestational/epidemiology"[Mesh] OR "Diabetes, Gestational/diet therapy"[Mesh]) AND (ethnicit* or ethnic or native or minorit* or "high risk" or race or "Population Groups"[Mesh] OR "Ethnic Groups"[Mesh] OR "Race Factors"[Mesh] OR "Minority Groups"[Mesh]) AND (Diet* OR Dietary OR "Feeding behaviour*" OR "Eating behaviour*" OR Eating OR "Feeding habit*" OR "Nutritional habit*"OR "Eating habit*" OR Food OR "food consumption*" OR "Nutritional consumption" OR Calorie OR "Calorie consumption" OR Caloric intake OR "Diet*pattern*" OR "Eating pattern*" OR "Nutritional intake" OR Cooking OR "Diet"[Mesh] OR "Protective Factors"[Mesh] OR "Drinking"[Mesh] OR "Energy Intake"[Mesh] OR "Diet Therapy"[Mesh] OR*

*"Nutrition Therapy"[Mesh] OR "Feeding Behavior"[Mesh] OR "Nutritional Status"[Mesh] OR "Food"[Mesh] OR "Nutrients"[Mesh] OR "Eating"[Mesh:NoExp] OR "Cooking"[Mesh:NoExp])AND (Prospective OR Retrospective OR Longitudinal OR Cohort OR follow-up OR observational or Randomised Controlled Trial* or controlled trial*, or clinical trial*, or clinical stud* OR "Epidemiologic Studies"[Mesh] OR "Nutrition Surveys"[Mesh] OR AND "Health Surveys"[Mesh:NoExp] OR "Diet Records"[Mesh] OR "Nutritional Sciences"[Mesh] OR "Clinical Studies as Topic"[Mesh] OR "Observational Studies as Topic"[Mesh])AND ("pre-pregnancy" or "pre pregnancy" or "early pregnancy" or "prenatal" or "pre-natal" or "pregnant" or "pregnancy" or "Prenatal Nutritional Physiological Phenomena"[Mesh] OR "Prenatal Care"[Mesh] OR "Preconception Care"[Mesh]). Limited to humans/ English'*

Prior to screening, PROSPERO registration was obtained for this review (PROSPERO registration number CRD42019140873). Following full text screening, citation lists of included studies were searched for additional relevant studies until no further studies were identified.

### 3.4.2 Inclusion and exclusion criteria

RCTs and observational studies published in English were eligible for inclusion in this review (with the exception of non-nested case control studies due to the inability to infer temporality). During the abstract screening stage, studies that explored the association between diet and GDM (reporting ORs, RRs or raw data) were included. During full text screening, studies were deemed eligible if they reported details of the ethnicity of the participants. Studies were excluded if: they commenced in the third trimester, followed up for less than 1 trimester, combined diet with other lifestyle interventions, included unhealthy participants (i.e., individuals with pre-existing diabetes or individuals living with morbid obesity), or did not report participant ethnicity or nationality. Where effect estimates were adjusted for ethnicity without stating the ethnicity, the baseline group was assumed to be the ethnic majority (≥60% of the population). Where this was not possible to confirm, corresponding authors were contacted. If no additional data were obtained the study was

excluded. Abstracts were screened in duplicate, with disagreements mediated by a third reviewer.

### 3.4.3 Data extraction

The following variables were extracted from all studies when provided: the numbers of participants, cases, control, unexposed and exposed individuals, summary statistics for participant's age and BMI, country of residence of study participants, main ethnicity of study participants (and the % of participants with this ethnicity),GDM diagnostic criteria utilised, and unadjusted/ adjusted effect estimates were extracted for all studies. For adjusted effect estimates, the covariates adjusted for were also extracted alongside information on whether studies had excluded multiparous women, women with a multiple pregnancy or previous miscarriage/stillbirth upon recruitment. Additionally, for RCTs the time of study initiation during pregnancy and the duration of the intervention were extracted and details of the intervention were extracted; while for observational studies, details of the dietary pattern assessed were extracted. 10% of included studies were randomly selected and data extraction was performed in duplicate by a second reviewer.

### 3.4.4 Data analysis

Identified RCTs and observational studies were analysed separately. Crude and adjusted effect estimates were obtained for all studies. When multiple effect estimates were presented, the estimate adjusted for the most confounders was utilised in the analysis. Studies were grouped based on their dietary exposures and then split into subgroups based on the participants' ethnicity. Healthy and unhealthy diet categories were defined based on: (i) study authors' definition, or (ii) common definitions according to major health bodies (e.g., WHO, WCRF, NHS UK, ADA) (World Health Organisation (WHO), 2020; National Health Service (NHS), 2019; Centre for Disease Control (CDC), 2019; American Diabetes Association (ADA), 2020). In general, 'healthy' diets were characterised by fruit and vegetables, wholegrains, fish, lean meats, and unsaturated fats; while 'unhealthy' diets

were characterised by red/ processed meats, fried foods, confectionary, sugar sweetened beverages (SSBs), saturated fats, and added sugars. Where it was difficult to group with confidence, diets were unclassified. Unclassified diets required ≥2 studies to be considered. When exposure data was presented categorically, highest consumers were compared to lowest consumers.

RRs were converted to ORs with 95% confidence intervals before running generic inverse variance weighted random-effect meta-analysis using the DerSimonian-Laird (DL) approach performed in Review Manager 5.3 (RevMan) from the Cochrane collaboration (The Cochrane Collaboration, 2020). Review Manager 5.3 was also used to assess differences in effect estimates between ethnic groups. Due to reports of inflated type 1 error leading to low reliability of the DL approach when study numbers are low (n<5) a sensitivity analyses using the Hartung-Knapp-Sidik-Jonkman (HKSJ) random effects model was performed through the use of the *meta* package (V4.9-6) in the R Studio environment (version 1.2.5019) (Jackson et al., 2017; Inthout et al., 2014; Tobias et al., 2012). To account for multiple testing a Bonferroni correction was applied as necessary.

### 3.4.5 Risk of Bias (ROB) assessment

For each study, ROB was assessed using a modified version of the 2016 Academy of Nutrition and Dietetics tool 'Evidence Analysis Manual: Steps in the Academy Evidence Analysis Process', a tool translatable to both RCTs and observational studies and designed for nutritional research (Academy of Nutrition Dietetics., 2016). The tool contains 10 validity questions that can be answered with either a positive score, a neutral score, or a negative score when the source of bias is likely to influence study conclusions. To aid the user in assigning a score, each validity question is composed of weighted 'important' sub-questions, and where ≥ 6 validity questions are answered 'negative' a study is determined to have a high risk of bias. Despite these guidelines, the protocol provides no strict guidelines for translating subgroup question scores to overall scores for each validity question. To ensure a systematic approach, a protocol was developed and followed to translate the scores of the sub-questions to the scores of the validity

questions. When an 'important' sub-question and/or ≥ 50% of all sub-questions was assessed to be negative, the overall validity question was assigned as negative. When all sub-questions were assigned to be positive, the overall validity question was assigned to be positive. In situations meeting neither of the above criteria, the validity question was assigned to be neutral. The tool was also modified to exclude question 5.1 as it is unfeasible to blind participants to their food intake. For the purpose of this chapter, the meaning of blinding was modified to refer only to the blinding of data collection for exposure collection and the blinding of a participant's diet when assessing the outcome (i.e., GDM). When ≥ 10 studies were present for an exposure, publication bias was also assessed via funnel plot asymmetry in STATA (StataCorp, 2017).

### 3.4.6 Sensitivity analyses

Sensitivity analyses were performed on all exposures classified as having 'considerable' or 'substantial' heterogeneity based upon $I^2$ values (i.e., $I^2 \geq 40$) as defined by Cochrane (Higgins et al., 2019). Sensitivity analyses to account for GDM confounders were performed by subgroup analyses on the following variables: (i) studies of dietary intervention during pregnancy; (ii) studies of older mothers; (iii) studies of mothers with an underweight/ healthy BMI; (iv) studies including overweight/ obese mothers (BMI ≥ 25 kg/$m^2$ in non-Asian populations or ≥23 kg/$m^2$ in Asian populations 29); and (v) studies conducted in an east Asian (EA) vs south Asian (SA) population. The average age of mothers in each study was classed as young or old if the average age of mothers in the study was below or above the average age of a woman during her first pregnancy in that country/region. Cut-off values for this analysis were obtained from the Office of National Statistics (UK), CIA (USA, Spain, Australia, Iceland, Finland, Japan, France, Italy, Singapore), OECD and EU average (Malaysia and multinational studies), China (Pulitzer Center), Iran (Worldbank) and India (Times of India). No sensitivity analysis was performed if all studies in the meta-analyses were classified similarly for the same confounder (e.g., all were studies of young mothers). Likewise, the obstetric adjustment sensitivity analyses were not run if all studies examining that exposure had accounted for obstetric risk factors. Where possible,

estimates unadjusted for age and BMI were calculated in the sensitivity analyses.

### 3.4.7 Dose-response analysis

A dose response relationship was examined for observational studies reporting multiple levels of consumption of a dietary factor using linear, quadratic or cubic spline models. This analysis was done using the *dosresmeta* package (version 1.2.5019) in R Studio (Crippa and Orsini, 2016). To minimise heterogeneity and maximise power, the existence of a dose response relationship between dietary exposures and GDM was explored in observational studies that reported multiple measures of adherence to a dietary pattern or levels of consumption of a macronutrient or foodstuff. For this, the required data (number of cases, number of controls, ORs and 95% CIs) were obtained for all stated doses. To allow studies to be combined, the scales of all studies for an exposure were transformed onto the same scale (i.e., a study measuring adherence in quartiles changed to represent units of 1.25, 2.5, 3.75, 5 if combined with a study measuring adherence in quintiles). Studies using binary classification for exposure adherence (exposed and unexposed) were excluded before the development of quadratic spline and cubic spline models, due to the intractability of the algorithm when these studies were included.

To minimise this limitation for quadratic models, an exponential curve was plotted using the ORs for the exposed and unexposed groups as the maximum and minimum effect sizes. From this, the median dose was calculated which could be included in a new quadratic model (hereafter referred to as quadratic*). The sample size for the calculated OR data point was assumed to be the same as the sample size of exposed individuals. These quadratic* models were utilised to minimise the data loss that would have occurred by excluding studies with only two adherence effect sizes presented and was possible due to the ease of extracting an additional point from an exponential curve. 3k models (knots at 0.1, 0.5, 0.9) and 4k models (knots at 0.05, 0.35, 0.65, 0.95) were run when possible (Perperoglou et al., 2019). Knots indicate percentile of the dose distribution where the linear trend deviates and the trend curves allowing for flexible trends in the data to be

modelled. Between each pair of successive knots cubic splines are formed using cubic polynomials (Orsini et al., 2012). In order to fit either a 3K or 4K spline model it is essential that all included studies contained at least as many doses as the number of splines, meaning that all studies included in a 3K models must present at least 3 effect sizes (i.e., 3 measures of adherence) and all studies included in 4K models must present at least 4 effect sizes (i.e., 4 measures of adherence). Wald test estimates were used to determine whether the cubic spline relationship provided evidence of deviation from the linear model. Log likelihood (LogL) values were used to select the best fitting model overall.

### 3.4.8 Power analysis

The motivation driving meta-analysis studies is the increase in power that comes from combining studies. However, it is possible for a random effects analysis to be underpowered even if the total number of participants is high, either if the total number of studies is small or if the study effect estimates are highly dispersed (Borenstein et al., 2009). *Post-hoc* power analyses were undertaken using fixed-effects ($tau^2=0$) or random-effects ($tau^2>0$) methods. Exposed and unexposed were calculated as an average of the exposed/ unexposed sizes of all studies for the exposure. Power ≥80% was considered adequate while a meaningful change in effect size was considered as 20% (i.e., a change in OR of ± 0.2) based upon recent meta-analyses investigating the association between a range of diets and GDM in European, American and Asian populations (Mijatovic-Vukas et al., 2018; Hassani et al., 2020).

## 3.5 Results

### 3.5.1 Summary of included studies

Abstract screening identified 3,393 studies after the removal of duplicates; of which, 57 of these were retained for full text screening and 38 (6 RCTs and 32 observational studies) were kept for the final analyses (**Figure 3.1)**. Identified studies included a total of 251,778 participants encompassing 5 distinct ethnic groups: white European (WE) (83%), Asian (9%, East and South Asian), Mediterranean (i.e., southern European populations; 5%)

(Seldin et al., 2006). Australian nationals (Australian residents with an almost equal proportion of Asians and WEs; 3%) and Iranian (<1%). The maternal age of study participants ranged from 24–36 years, and 14/38 studies reported an average BMI of overweight or obese **(Appendix Figures A.1, A.2 and A.3)**.

#### 3.5.1.1 Exposure assessment

Twenty-five studies reported on one or more of the following healthy dietary patterns: healthy recommendations (n=13), Mediterranean diet (n=6), prudent diet (n=4), plant-based diet (n=6) or a healthy snack-based intervention (n=1). Thirteen studies also reported on a dietary pattern that could be considered unhealthy: Western diet (n=6), fried/fast food (n=4), sweet and seafood dietary pattern (n=2) and an unhealthy dietary score (n=1). Additional dietary patterns included in this review that were classified as neither healthy nor unhealthy, there were grouped as 'unclassified' (meat pattern, high protein diet, a traditional Asian diet, high-fish diets, high-fat diets, high-carbohydrate diets, high-animal protein diet and high-vegetable protein diet); all of which were included in ≥ 2 of the identified studies **(Appendix Figure A.4)**

#### 3.5.1.2 Outcome assessment

All included RCTs utilised a 75g OGTT for GDM diagnosis while 62.5% of observational studies utilised an OGTT with clearly defined diagnostic criteria. For both study types, IADPSG (2010/2013) guidelines were the most commonly utilised. Following a chi-squared test, no significant difference was found between the proportion of studies with and without stated GDM diagnostic criteria between ethnicities or the proportion of studies that utilised an OGTT test for GDM diagnosis between ethnicities. **(Appendix Table A.1)**.

### 3.5.2 RCTs

Six RCTs, including 3,041 individuals from 4 population groups (Asian, Australian nationals, Mediterranean and WE), evaluated the impact of healthy

dietary interventions on GDM incidence (**Table 3.1**). Four studies reported on the effect of generalised healthy dietary interventions. One study reported on the effect of the Mediterranean diet, and 1 study reported on the effectiveness of a healthy snack intervention during pregnancy. When analysed collectively, there was no evidence that these interventions were effective in reducing GDM incidence (**Figure 3.2**). Likewise, when stratified by intervention type, no evidence was found to support the effectiveness of healthy recommendations or the consumption of a healthy snack in reducing GDM incidence (**Appendix Figure A.5)**.

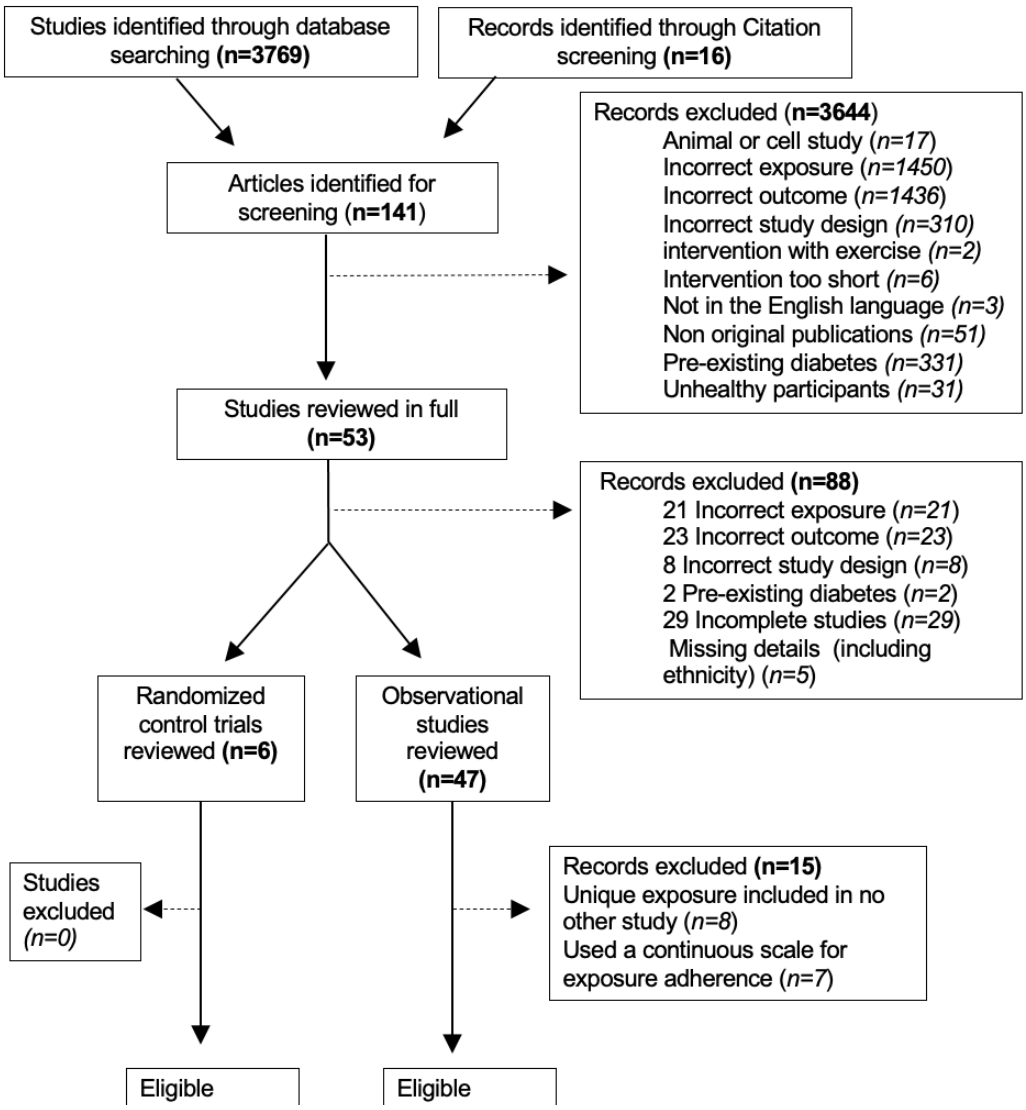



**Figure 3.1: PRISMA diagram highlighting systematic search.** Records were identified via searches in Pubmed, OVID, Scopus and Cochrane databases. Databases were searched from their inception through to the 31/01/2021.

**Table 3.1: Included RCTs.**

| Study | Country of study and participant ethnicity | GDM Diagnostic Criteria | Pre/ During pregnancy | Initiation (I) and Duration (D) | Sample size (Int/C) | Exposure | Intervention | Control |
|---|---|---|---|---|---|---|---|---|
| **Assaf-balut, 2017 (Assaf-Balut et al., 2017)** | Spain<br><br>Mediterranean | 75g OGTT, ≥1 of the following: Fasting ≥5.1 mmol/L 1-hr ≥10 mmol/L 2-hour ≥8.5mmol/L | During Pregnancy | **I:** 12th week gestation **D:** 24-26 weeks' gestation | 874 (434/440) | **Healthy**<br><br>Med diet | • 3 servings/day vegetables, fruit (not including juices), skimmed dairy, and wholegrains<br>• 2-3 servings/week legumes and fish consumption<br>• Limited red meat, refined sugars, and convenience/ processed foods<br>• Supplemented with olive oil and pistachios | Usual standard of care |
| **Markovic, 2016 (Markovic et al., 2016)** | Australia<br><br>Australian Nationals | 75g OGTT Fasting ≥ 5.5 mmol/L 1-hr ≥ 10 mmol/L 2-hour ≥ 8 mmol/L | During Pregnancy | **I**: 14-20 weeks' gestation **D**: Not reported | 125 (65/60) | **Healthy**<br><br>HR | • Low GI diet | High fibre, moderate-GI diet |
| **Opie, 2016 (Opie et al., 2016)** | Australia<br><br>White European | 75g OGTT Fasting ≥ 5.5mmol/L | During Pregnancy | **I**: 24-28 weeks' gestation | 153 (82/71) 58 (10/48) | **Healthy**<br><br>HR | • AGHE modified for pregnancy | Usual standard of care |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Asian | 2-hour ≥8 mmol/L | | **D**: Not reported | | | | |
| **Rono, 2018 (Rönö et al., 2018)** | Finland<br><br>White European | 75g OGTT<br>Fasting ≥5.3 mmol/L<br>1-hr ≥10 mmol/L<br>2-hour ≥8.6 mmol/L | During pregnancy | **I:** <20 weeks' gestation<br>**D:** Until OGTT test based upon | 335 (235/100) | **Healthy**<br><br>HR | • Structured and individually modified dietary advice based on the Nordic nutritional recommendations | Usual standard of care |
| **Sahariah, 2016 (Sahariah et al., 2016)** | India<br><br>Asian | 75g OGTT, ≥1 of the following:<br>Fasting ≥ 5.1 mmol/L<br>2-hour > 8.5mmol/L | Pre and During pregnancy | **I:** ≥ 90 days pre-pregnancy<br>**D:** Until delivery | 1008 (492/516) | **Healthy**<br><br>Healthy Snack | • Supplementation with leafy green vegetables, fruit and milk | Snacks made of low macro-nutrient vegetables including potatoes and onions |
| **Simmons, 2017 (Simmons et al., 2017)** | European countries<br><br>Mediterranean | 75g OGTT, ≥1 of the following:<br>Fasting ≥ 5.1 mmol/L<br>2-hour > 8.5mmol/ | During pregnancy | **I:** <20 weeks' gestation<br>**D:** Until delivery | 206 (106/100) | **Healthy**<br><br>HR | • Lower simple/complex carbohydrate and fat intake<br>• Higher fibre and protein intake. Calorie deficit focused on portion control | Usual standard of care |

Bolded exposures show overarching exposure. C: Control. D: Duration. HR: Healthy recommendations. I: Initiation. Int: Intervention.

Conversely, the consumption of a Mediterranean diet enriched with extra virgin olive oil and pistachios was found to be protective against GDM in a single study in a Mediterranean population (OR = 0.67, 95% CI 0.48 – 0.98) (Sahariah et al., 2016). Only 2/3 of dietary interventions were found to be culturally sensitive.

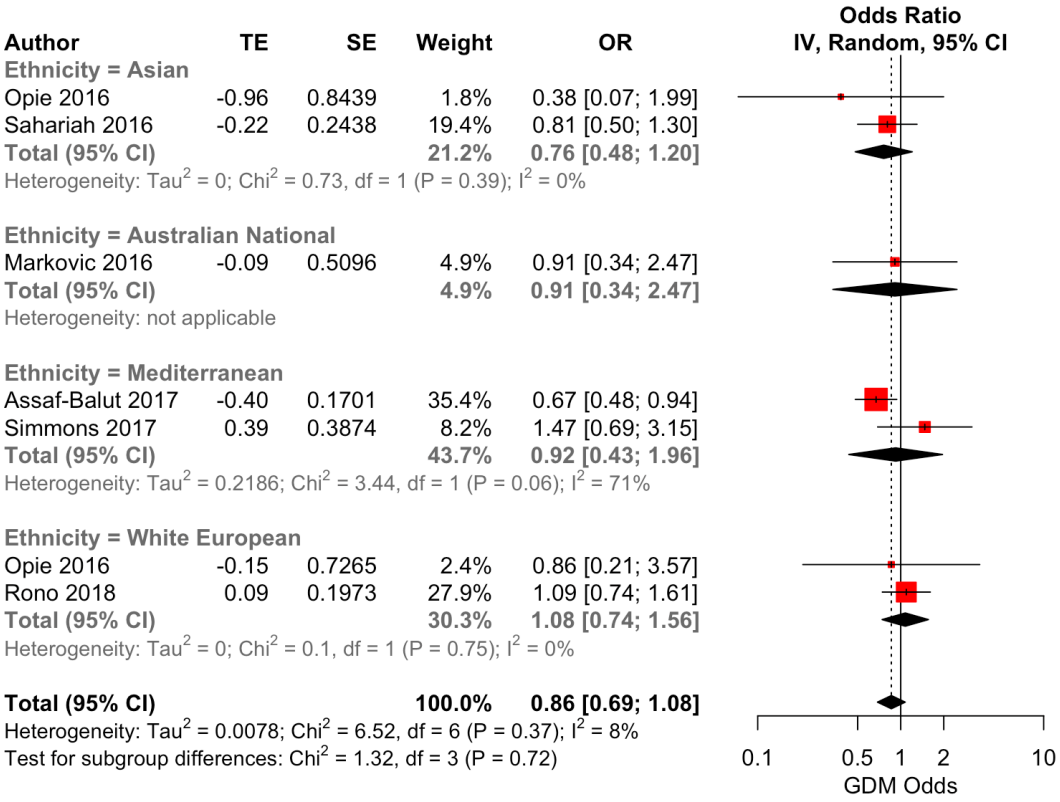



**Figure 3.2: Forest plot of healthy dietary interventions.** Results from a DerSimonian and Lard (DL) approach. CI: confidence interval. IV: inverse variance. OR: odds ratio. SE: standard error. TE: treatment effect.

### 3.5.3 Observational studies

Thirty-two observational studies, including 248,737 individuals from 5 ethnic groups (Asian, Australian Nationals, Iranian, Mediterranean and WEs) across 20 countries, were identified for this review. In total 17 dietary patterns were reported across these studies **(Table 3.2)**.

**Table 3.2: Included observational studies.**

| Study | Country of study and participant ethnicity | GDM Diagnostic Criteria | Method of dietary assessment | Timing of assessment | Exposure | Sample Size (High /low) | Dietary Pattern |
|---|---|---|---|---|---|---|---|
| **Assaf-Balut, 2018 (Assaf-Balut et al., 2018)** | Spain<br><br>Mediterranean | Not stated | FFQ | During pregnancy: 12-14 weeks | **Healthy**<br><br>Mediterranean diet | 759 (623/136) | Frequent consumption of vegetables, legumes, fruits, nuts, EVOO, oily fish, canned fish, wholegrains, cereals, pasta and skimmed dairy products |
| **Badon, 2016 (Badon et al., 2017)** | USA<br><br>White European | 100g OGTT<br>≥2 of the following:<br>Fasting ≥ 105 mg/dL<br>1-hr ≥190 mg/dL<br>2-hour ≥165 mg/dL<br>3-hr ≥ 145 mg/dL | FFQ- Diet over the previous year | During pregnancy: 15 weeks, +/- 3 weeks) | **Healthy**<br><br>HR | 3305 (611/2694) | • Modified version of AHEI-2010<br>• Increased consumption of vegetables, fruit, whole grains, nuts, long- chain (n-3) fatty acids, and PUFAs<br>• Limited intake of SSBs, red/processed meat, trans-fat and sodium |
| **Bao, 2013 (Bao et al., 2013)** | USA<br><br>White European | Self-reported | FFQ every 4 years | Prior to pregnancy: 2001 | High-protein diet | 5799 (2793/3006) | • Individuals' total protein intake, split into quintiles based on population (%TE) |
| | | | | | Vegetable protein | 6159 (2871/3288) | • Individuals' vegetable protein intake, split into quintiles based on population intake (%TE) |
| | | | | | Meat pattern/ Animal protein | 5877 | • Individuals' animal protein intake, split into quintiles |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | (2862/3015) | based on population intake. (%TE) |
| **Bao, 2014** **(Bao, W. et al., 2014)** | USA<br><br>White European | Self-reported | FFQ – Diet over the previous day (x4) | Prior to pregnancy: 2001 | **Healthy**<br><br>Plant-based pattern | 10860 (3685/7175) | • Determined on the basis of percentage of energy from carbohydrate, vegetable protein and vegetable fat<br>• Higher score reflects a higher intake of vegetable protein/ fat and a lower intake of carbohydrate |
| | | | | | Meat pattern | 10301 (4642/5659) | • Determined on the basis of percentage of energy from carbohydrate, total protein and total fat<br>• Higher score reflects a higher intake of protein/ fat and a lower intake of carbohydrate |
| | | | | | High Protein | 10864 (4591/6273) | • Determined on the basis of percentage of energy from carbohydrate, animal protein and animal fat<br>• Higher score reflects a higher intake of animal protein/ fat and a lower intake of carbohydrate |
| **Bao, 2014[2]** **(Bao, Wei et al., 2014)** | USA<br><br>White European | Self-reported | FFQ every 4 years | Prior to pregnancy: 2001 | **Unhealthy**<br><br>Fried/ fast food | 10866 (348/10518) | • Frequent consumption of fast food |
| **Bowers, 2012 (Bowers et al., 2012)** | USA<br><br>White European | Self-reported – no information on the criteria used | FFQ every 2 years – Diet over the previous year | Pre-pregnancy: Most recent questionaries | Fat | 6020 (2190/3110) | • Total fat (%TE) |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Domingues, 2014 (Domingues et al., 2014)** | Spain<br><br>Mediterranean | Self-reported and then confirmed via a panel of medical doctors. Most common criteria used National Diabetes Group Criteria and Carpenter and Coustan cut-offs | FFQ every 2 years | Pre-pregnancy: Most recent questionnaire | **Unhealthy**<br><br>Fried/ fast food | 1587 (971/616) | • Frequent consumption of hamburgers, sausages and pizza |
| **Donazar-Ezcurra, 2017 (Donazar-ezcurra et al., 2017)** | Spain<br><br>Mediterranean | Not stated | FFQ- Diet over the previous 24 hours | Pre-pregnancy: Most recent questionnaire | **Healthy**<br><br>Mediterranean diet | 1727 (863/864) | • High intake of poultry, olive oil, nuts, low-fat dairy products, whole grain bread, fish, fruit and vegetables |
| | | | | | **Unhealthy**<br><br>Western diet | 1727 (863/864) | • Frequent intake of high-fat processed meats, potatoes, commercial bakery goods, whole dairy products, fast food, sauces, pre-cooked foods, SSB's and confectionary |
| **Flynn, 2016 (Flynn et al., 2016)** | UK<br><br>White European | 75g OGTT ≥1 of the following: Fasting ≥ 5.1 mmol/L 1-hr ≥ 10 mmol/L 2-hour > 8.5 mmol/L | FFQ – Diet over the previous week | During pregnancy: 15-18 weeks until 27-28 weeks | **Unhealthy**<br><br>Western diet | 425 (206/219) | • High intake of potato, French fries, crisps, processed meat, fizzy drinks (SSB's and sugar free), root vegetables, green vegetables and chocolate |
| | | | | | **Healthy**<br><br>Plant-based pattern | 429 (213/216) | • Frequent intake of fresh/citrus/tropical fruits, green/root/salad |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | | vegetables, bananas and yoghurt |
| **Gicevic 2018 (Gicevic et al., 2018)** | USA<br><br>White European | Self-reported – No information on most common criteria used | FFQ- Diet over the previous year | Pre-pregnancy: Most recent dietary assessment (1991-2001) | **Healthy**<br><br>HR | 7274 (4206/3068) | • AHEI 2010 adherence score.<br>• Characterised by high intakes of nuts, long chain omega-3 fats, polyunsaturated acids, nuts and low intakes of red meat, refined sugar, refined grains and SSBs |
| **He, 2015 (He et al., 2015)** | China<br><br>Asian | 75g OGTT ≥1 of the following: Fasting ≥ 5.1 mmol/L 1-hr ≥ 10 mmol/L 2-hour > 8.5 mmol/L | FFQ- Diet over the previous week (x3) | During pregnancy: Week 16, Weeks 24-27, Weeks 35-38 | **Healthy**<br><br>Prudent diet | 369 (188/181) | • Frequent intake of dairy products, nuts, eggs, fish, soups and fruits.<br>• Infrequent intake of processed meats, SSBs, and processed vegetables |
| | | | | | **Healthy**<br><br>Plant-based pattern | 2040 (1019/1021) | • Frequent intake of, beans, mushroom, melon vegetables, seaweed, legumes, fruits, leafy/root/cruciferous vegetables, nuts and cooking oil |
| | | | | | High-protein diet | 2046 (1023/1023) | • High intake of poultry red/animal/processed/organ meat, grains, fish, soups, leafy/ cruciferous vegetables and eggs |
| | | | | | **Unhealthy**<br><br>Sweets and Seafood pattern | 2043 (1019/1024) | • Frequent intake of Cantonese desserts, molluscs, shellfish and SSBs |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Hrolfsdottir, 2020 (Hrolfsdottir et al., 2019)** | Iceland<br><br>White European | 75g OGTT<br>≥1 of the following:<br>Fasting ≥ 5.1 mmol/L<br>1-hr ≥ 10 mmol/L<br>2-hour > 8.5 mmol/L | FFQ – Diet over the previous month | During pregnancy: 11$^{th}$ -14$^{th}$ week | **Unhealthy**<br><br>Unhealthy diet score | 709 (302/407) | • Low intake of fruit, vegetables fish, dairy, wholegrains, beans, nuts, seeds, and vitamin D supplementation<br>• High intake of refined sugars and grains, and processed foods, dairy, and low-quality fat (i.e., using butter rather than oil). |
| **Hu, 2019 (Hu et al., 2019)** | China<br><br>Asian | 75g OGTT<br>≥1 of the following:<br>Fasting ≥ 5.1 mmol/L<br>1-hr ≥ 10 mmol/L<br>2-hour > 8.5 mmol/L | FFQ – Diet over the previous 2 months | During pregnancy: 22$^{nd}$ week | **Unhealthy**<br><br>Sweets and Seafood pattern | 508 (255/253) | • Frequent intake of pastries, candid, sweet beverages, shrimps, crabs, fruit, mussels and red meat |
| | | | | | Fish- Seafood | 508 (255/253) | • Frequent intake of marine fish, shrimp, crabs and mussels, freshwater fish and seaweed<br>• Infrequent intake of eggs, dairy products and rice |
| | | | | | Traditional Asian diet | 508 (255/253) | • Frequent intake of tubers, vegetables, fruit, rice, red meat, eggs and nuts |
| **Karamanos, 2014 (Karamanos et al., 2014)** | Mediterranean countries (Algeria, France, Greece, Italy, Lebanon, Malta, Morocco, Serbia, Syria, Tunisia) | 75g OGTT<br>≥2 of the following:<br>Fasting ≥ 5.3 mmol/L<br>1-hr ≥ 10 mmol/L<br>2-hour > 8.6 mmol/L<br><br>OR | FFQ | During pregnancy: Before OGTT (24$^{th}$-28 week) | **Healthy**<br><br>Mediterranean diet | 668 (334/334) | • High adherence to the Med Diet pyramid<br>• Characterised by frequent intake of bread, cereals, legumes, vegetables, fruits, meat, fish, eggs, potatoes, cheese and dairy products<br>• High ratio of olive oil to animal fat |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Mediterranean | ≥1 of the following:<br>Fasting ≥ 5.1 mmol/L<br>1-hr ≥ 10 mmol/L | | | | | |
| **Lamyian, 2017 (Lamyian et al., 2017)** | Iran<br><br>Irani | 75g OGTT<br>≥2 of the following:<br>Fasting 95≥ mg/dL<br>1-hr ≥180mg /dL<br>2-hour ≥155mg /dL<br>3-hr ≥140mg /dL | FFQ- Diet over the previous year | During pregnancy: Before 6th week | **Unhealthy**<br><br>Fried/ fast food | 513 (256/257) | • High intake of hamburger, bologna, pizza, sausage and French fries |
| **Le Donne, 2016 (Le Donne et al., 2016)** | Italy<br><br>White European | Not stated | FFQ- Diet over the previous week | During pregnancy: 34th week | Fish | 114 (104/10) | • Intake of tuna, swordfish, mackerel, salmon, anchovy, garfish, spatula, sardine, sea gilt-head bream, sea bass, cod, sea bream, perch and shellfish |
| **Liang, 2018 (Liang et al., 2018)** | China<br><br>Asian | 75g OGTT<br>≥1 of the following:<br>Fasting ≥ 5.1 mmol/L<br>1-hr ≥ 10 mmol/L<br>2-hour > 8.5 mmol/L | FFQ- Diet over the previous day | Pre-pregnancy: (Assessed at first routine ultrasound) | High-protein diet | 714 (342/372) | • High intake of meats, fish, shrimps, dairy products, soybeans and nuts |
| | | | | | Fish | - | • Intake of fish and shrimp |
| | | | | | Vegetable protein | 796 (439/357) | • Intake of Beans (soybeans and soybean products) and nuts |
| | | | | | Meat pattern/ Animal protein | 5877 (2862/3015) | • High intake of animal protein (% TE)<br>• Characterised by intake of meats, fish, shrimps and dairy products |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Looman, 2018 (Looman et al., 2018)** | Australia<br><br>Australian Nationals | Self-reported<br><br>Diagnostic criteria during study period<br><br>75g OGTT<br>Fasting ≥ 5.5mmol/L<br>2-hour ≥8.0 mmol/L<br><br>OR<br><br>75g OGTT<br>≥1 of the following:<br>Fasting ≥ 5.1 mmol/L<br>1-hr ≥ 10 mmol/L<br>2-hour > 8.5 mmol/L | FFQ- Diet over the previous year | Pre-pregnancy: Dietary information collected in 2003 used as baseline. Pregnancies reported in years 2006, 2009, 2012 and 2015. | **Healthy**<br><br>HR | 2431 (1529/902) | • Low GI diet<br>• Lowest group median glycaemic index 47.8. Highest group median glycaemic index 56.7 |
| | | | | | High protein diet | 2483 (1537/901) | • Low Carbohydrate Diet score |
| | | | | | Carbohydrate | 3051 (1510/1541) | • Carbohydrate intake (%TE) |
| **Mak, 2018 (Mak et al., 2018)** | China<br><br>Asian | 75g OGTT<br>≥1 of the following:<br>Fasting ≥ 5.1 mmol/L<br>1-hr ≥ 10 mmol/L<br>2-hour > 8.5 mmol/L | FFQ | During pregnancy: Between 15$^{th}$-20$^{th}$ week | Meat pattern | 892 446/446) | • Frequent intake of organ/processed meat, ox tripe, pig blood curd, squid, pork, and mushrooms |
| | | | | | **Healthy**<br><br>Plant-based pattern | 892 (446/446) | • High intakes of green leafy vegetable/ cruciferous/ gourd/melon family/red or orange/root/bean vegetables<br>• High intake of potatoes, bean products, mushrooms and fruits |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | | • Low intake of lean pork meat |
| | | | | | High-protein diet | 891 (445/446) | • High intake of eggs, milk, lean pork meat and fish<br>• Low intake of bread and sea vegetables |
| **Marí-Sanchis, 2018 (Sanchis et al., 2018)** | Spain<br><br>Mediterranean | Self-reported and then confirmed via an endocrinologist Most common criteria used National Diabetes Group Criteria and Carpenter and Coustan cut-offs | FFQ every 2 years | Pre- pregnancy | Meat pattern | 1649 (824/825) | • Frequent intake of red/processed/ unprocessed meats, poultry and rabbit |
| **Mohanty, 2015 (Mohanty et al., 2015)** | USA<br><br>White European | 100g OGTT ≥2 of the following: Fasting ≥ 105 mg/dL 1-hr ≥190 mg/dL 2-hour ≥165 mg/dL 3-hr ≥ 145 mg/dL | FFQ – Diet 3 months prior to conception and during the first trimester. | During pregnancy: 16$^{th}$ week | Fish | 2418 (2116/302) | • Intake of shellfish, lean fish and fatty fish |
| **Osorio-Yáñez, 2017 (Osorio-Yáñez et al., 2017)** | USA<br><br>White European | 100g OGTT ≥2 of the following: Fasting ≥ 95 mg/dL 1-hr ≥180 mg/dL 2-hour ≥155 mg/dL | FFQ- Diet over the previous 3 months | During pregnancy: 15$^{th}$ week | **Unhealthy**<br><br>Fried/ fast food | 4207 (3414/793) | • Intake of fried potatoes, fried chicken, fried fish, doughnuts and snack crisps |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | 3-hr ≥ 140 mg/dL<br>USA<br><br>White European | | | | | |
| **Pang, 2017 (Pang et al., 2017)** | Singapore<br><br>Asian | 75g OGTT<br>≥1 of the following:<br>Fasting ≥ 7 mmol/L<br>2-hour ≥ 7.8 mmol/L | FFQ- Diet over the previous 24 hours/ 3 days | During pregnancy: 26th-28th week | High-protein diet | 490 (245/245) | • Protein intake from animal and vegetable sources combined |
| | | | | | Vegetable protein | 490 (245/245) | • Protein intake from vegetables, rice, noodles, desserts and beans |
| | | | | | Meat pattern/ Animal protein | 490 (245/245) | • High intake of animal protein (% TE) |
| **Schoenaker, 2015 (Schoenaker et al., 2015)** | Australia<br><br>White European | During study period following recommendations used in Australia:<br><br>75g OGTT<br>Fasting ≥5.6 mmol/L<br>2-hour ≥ 8 mmol/L | FFQ- Diet over the previous 12 months | During pregnancy: Varied times | **Healthy**<br><br>Mediterranean diet | 4376 (2249/2127) | • High intake of vegetables, legumes, nuts, tofu, rice, pasta, rye bread, wine and fish |
| | | | | | **Unhealthy**<br><br>Western diet | 4365 (2137/2228) | • Frequent intake of red/ processed meat, snacks, sweets (including cakes, chocolate and biscuits), pizza and fruit juice |
| | | | | | **Healthy**<br><br>Plant-based pattern | 4706 (2125/2554) | • Frequent intake of carrots, peas, cauliflower, broccoli, potatoes, pumpkin, green beans and cabbage |
| **Tajima, 2017 (Tajima et al., 2017)** | Japan<br><br>Asian | 75g OGTT<br>≥1 of the following:<br>Fasting ≥ 5.1 mmol/L<br>1-hr ≥ 10 mmol/L | FFQ- Diet over the previous 3 days | During pregnancy: First prenatal visit | Fat | 125 (105/110) | • Total fat (%TE) |
| | | | | | Carbohydrate | 216 (108/108) | • Carbohydrate (%TE) |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | 2-hour > 8.5 mmol | | | | | |
| **Tobias, 2012 (Tobias et al., 2012)** | USA<br><br>White European | Self-reported, most commonly by the National Diabetes Data Group Criteria | FFQ | Pre-pregnancy: Most recent questionnaire (survey every two years) | **Healthy**<br><br>HR | 9637 (6141/3496) | • High intake of AHEI diet<br>• Increased intake of fruit, vegetables, cereal fibre, nuts and multivitamins<br>• A high white: red meat ratio and PUFA:SFA ratio.<br>• Moderate alcohol consumption<br>• Decreased consumption of trans-fat |
| | | | | | **Healthy**<br><br>Mediterranean diet | 8572 (5275/3297) | • Increased intake of fruits, vegetables, nuts, legumes, soy, fish and wholegrains<br>• Moderate intake of alcohol and MUFA: SFA servings/d<br>• Limited intake of red and processed meat |
| **Tryggvadottir, 2015 (Tryggvadottir et al., 2015)** | Iceland<br><br>White European | 75g OGTT<br>≥1 of the following:<br>Fasting ≥ 5.1 mmol/L<br>1-hr ≥ 10 mmol/L<br>2-hour > 8.5 mmol/L | FFQ – Diet over the previous 4 Days | During pregnancy: 20 weeks | **Healthy**<br><br>HR | 168 (56/112) | • High adherence to the Healthy Eating Index. |
| | | | | | **Healthy**<br><br>Prudent diet | 168 (56/112) | • Frequent intake of seafood, eggs, fruits, vegetables, vegetable oils, nuts, seeds, pasta, breakfast cereals, coffee, tea and cocoa powder<br>• Limited intake of soft drinks and French fries. |
| | | | | | Fish | NA | • Intake of fish, seafood and shellfish products |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Yi, 2017 (Yi et al., 2017)** | China<br><br>Asian | 75g OGTT ≥1 of the following: Fasting ≥ 5.1 mmol/L 1-hr ≥ 10 mmol/L 2-hour > 8.5 mmol/L | 2 x FFQ- 24hr recall on 3 occasions | During pregnancy: 1 questionnaire during 5-15-week period, 1 questionnaire during 24-28-week period | **Unhealthy**<br><br>Western diet | 352 (173/179) | • Frequent intake of dairy, baked/fried food and white meat |
| | | | | | Traditional Asian diet | 346 (179/167) | • Frequent intake of light-coloured vegetables, fine grain, red meat and tubers |
| | | | | | **Healthy**<br><br>Prudent diet | 351 (181/170) | • Frequent intake of dark coloured vegetables and deep-sea fish |
| **Yong, 2020 (Yong et al., 2020)** | Malaysia<br><br>Asian | 75g OGTT ≥1 of the following: Fasting ≥ 5.6mmol/L 2-hour ≥ 7.8mmol/L | FFQ – Diet over the previous 6 months | During pregnancy:<br><br>~10 weeks, ~12 weeks, ~27 weeks | **Healthy**<br><br>Plant-based† | 300 (150/150) | • High intake of vegetables, nuts, seeds, legumes, fruits, eggs and dairy products |
| | | | | | **Unhealthy**<br><br>Western diet† | 300 (150/150) | • High intake of poultry, meat, sweet foods, seafoods, oil, fat, rice, noodles and pasta |
| **Zhang, 2006 (Zhang et al., 2006)** | USA<br><br>White European | Self-reported following a previous diagnosis. Most common criteria used was National Diabetes Group criteria | FFQ- Diet over the previous year | Pre-pregnancy: 1-7 years | **Healthy**<br><br>Prudent diet | 5185 (2519/2666) | • High intake of fruit, green leafy vegetables, poultry and fish |
| | | | | | **Unhealthy**<br><br>Western diet | 5196 (2530/2666) | • Highest meat (poultry, red and processed), pizza, dessert, sweet, French fries, dairy products and refined grain intake |
| **Zhang, 2014 (Zhang et al., 2014)** | USA<br><br>White European | Self-reported following a previous diagnosis. Most common criteria used was National Diabetes Group criteria | FFQ- Diet over the previous year | Pre-pregnancy: Most recent questionnaire | **Healthy**<br><br>HR | 8236 (4219/4017) | • Modified version of AHEI-2010 (10/11 components, excluding alcohol)<br>• Higher intakes of vegetables, fruit, whole grains, nuts, long- chain (n-3) fatty acids, and PUFAs |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | | • Lower intakes of SSBs, red/processed meat, trans-fat, and sodium |
| **Zhou, 2018 (Zhou et al., 2018)** | China<br><br>Asian | 75g OGTT<br>≥1 of the following:<br>Fasting ≥ 5.1 mmol/L<br>1-hr ≥ 10 mmol/L<br>2-hour > 8.5 mmol/L | FFQ- Diet over the previous month | During pregnancy:<br>2 weeks before OGTT<br>(OGTT at 24-28 weeks) | **Healthy**<br><br>Plant-based pattern | 131 (60/71) | • Frequent intake of root vegetables, melon, solanaceous/leafy/cruciferous vegetables, mushrooms, algae, beans and bean products |
| | | | | | Meat pattern | 132 (79/53) | • Frequent intake of animal organs and blood, seafood and poultry |
| | | | | | High-protein diet†† | 1356 (668/668) | • Total protein intake (%TE) |
| | | | | | Animal protein†† | 1356 (668/668) | • Animal protein (%TE) |
| | | | | | Fat†† | 1356 (668/668) | • Total fat (%TE) |
| | | | | | Carbohydrate†† | 1356 (668/668) | • Carbohydrate (%TE) |
| | | | | | Vegetable protein†† | 1356 (668/668) | • Vegetable protein (%TE) |

Bolded exposures represent overarching grouped exposure FFQ: food frequency questionnaire. SSB: Sugar sweetened beverages. TE: Total energy intake. HR: Healthy recommendation. † Number in each tertial calculated as a third of the study population.††Number in each quartile calculated as a quarter of the study population.

### 3.5.3.1 Healthy diets

Twenty studies reported on the relationship between healthy dietary patterns and GDM incidence in observational studies. These studies encompassed four distinct dietary patterns (i.e., healthy dietary recommendations, Mediterranean diet, prudent diet and plant-based diet) in 4 population groups (Asian, Australian Nationals, Mediterranean and WEs). When analysed collectively, high adherence to a healthy dietary pattern was found to be associated with a decreased odds of GDM by 22% (OR = 0.78, 95% CI 0.70 – 0.88, $I^2$ = 74%) compared to those with the lowest level of adherence (**Figure 3.3**). When stratified by ethnicity, healthy diets were also found to lower the odds of GDM in WE (OR = 0.75, 95% CI 0.65 – 0.88, $I^2$ = 79%), however, significant associations were not found in other ethnicities. Likewise, when stratified by healthy diet type, healthy diet recommendations and the Mediterranean diet were found to be associated with a lower odds of GDM solely in WEs. However, when the more stringent HKSJ approach was applied only healthy dietary recommendations retained significance **(Appendix Table A.2)**. The prudent diet and plant-based diet were not associated with a lower odds of GDM in any ethnicity **(Appendix Figure A.6)**.

### 3.5.3.2 Unhealthy diets

Thirteen studies reported on the relationship between unhealthy diets (Western diet, fried/fast food, unhealthy dietary score, sweet and seafood pattern) and GDM across 4 ethnicities (Asian, Iranian, Mediterranean and WE). When including all ethnicities, high adherence to an unhealthy dietary pattern associated with an increased odds of GDM by 44% (OR = 1.44, 95% CI 1.25 – 1.67, $I^2$ =41%) (**Figure 3.4).** This association was identified in WE (OR = 1.59, 95% CI 1.41 – 1.81, I2 =0), Mediterranean (OR = 1.69, 95% CI 1.21 -2 35, I2 =0), and Iranian (OR = 2.12, 95% CI 1.12 – 4.01) populations, although only one Iranian study was identified. Following stratification by unhealthy diet type, high adherence to a Western diet (n=6), fried/fast food (n=4) and an unhealthy diet score (n=1) was found to significantly associate with an increase

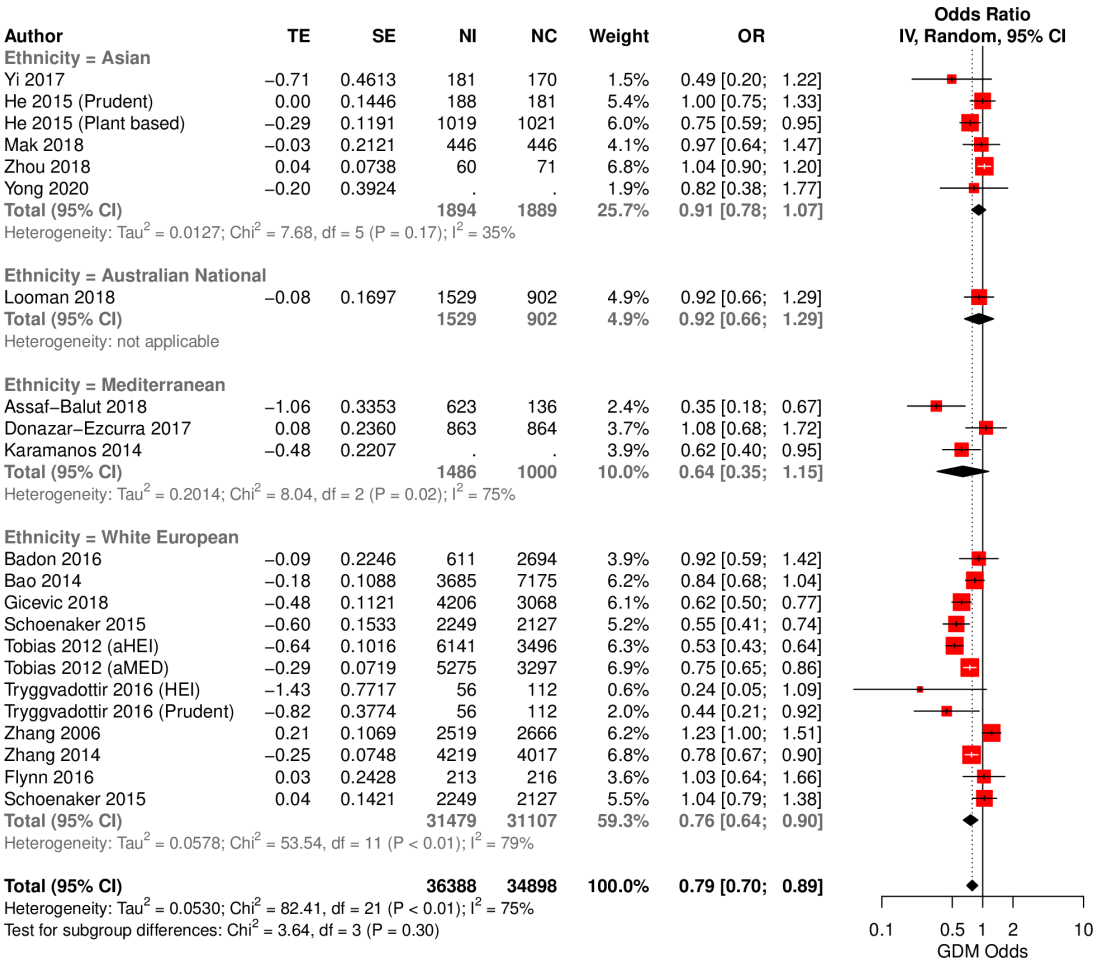



**Figure 3.4: Forest plot of observational studies for healthy diets.** Results from a DerSimonian and Lard (DL) approach. CI: confidence interval. IV: inverse variance. OR: odds ratio. SE: standard error. TE: treatment effect.

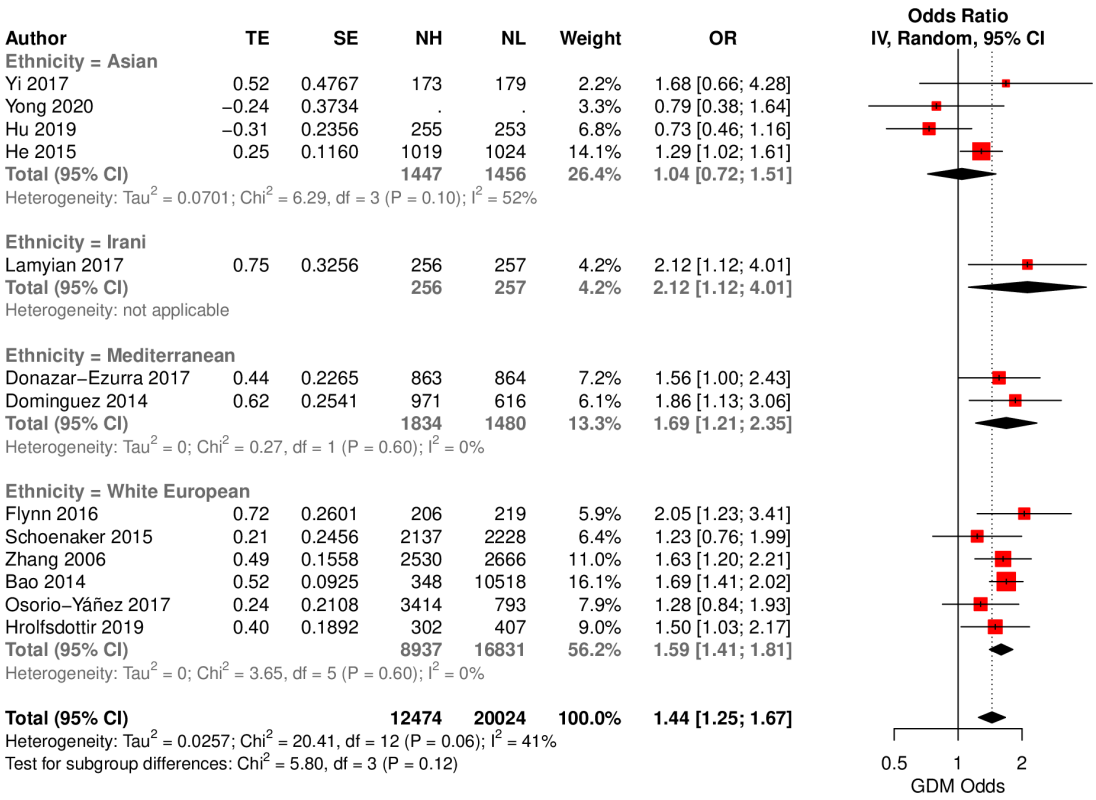



**Figure 3.5: Forest plot of unhealthy diets.** Results from a DerSimonian and Lard (DL) approach. CI: confidence interval. IV: inverse variance. OR: odds ratio. SE: standard error. TE: treatment effect.

in GDM odds by 51%, 66% and 50% respectively, the latter in a single WE population (**Appendix Figure A.7)**. Significant associations in the WE subtypes were also observed for the Western diet and fried food exposures. The only population for which unhealthy diets were not associated with an increase in GDM odds was the Asian population, where no association was identified in the collective analysis of all unhealthy diets, the Western diet and the sweet and seafood pattern.

### 3.5.3.3 Unclassified diets

Four dietary patterns (meat-based, high protein, traditional Asian and high-fish diets) could not be classified as healthy or unhealthy due to their constituent components. The meat-based pattern was associated with an increased odds of GDM when evaluating all individuals collectively (OR = 1.41, 95% CI 1.22 – 1.63, $I^2$= 23%), and when evaluating the WE subgroup

(OR = 1.41, 95% CI 1.18 – 1.68, $I^2$=0%) or the Mediterranean subgroup (OR =1.68, 95% CI 1.07 – 2.65) separately. Only one study was identified in a Mediterranean population (**Appendix Figure A.8)**. A high protein diet was also found to associate with an increased risk of GDM in the overall population (OR = 1.36, 95% CI 1.05 – 1.76, $I^2$=78%) and in WEs (OR=1.28, 95% CI 1.09 – 1.52, $I^2$ = 0%), but not in Asians. Likewise, no association between fish intake and GDM was identified in Asians, while a protective effect was identified in WEs (OR = 0.85, 95% CI 0.73 – 0.98, $I^2$ = 0%).

### 3.5.3.4 Macronutrients

Four dietary exposures identified in this review were classified based on the % of total energy intake of a specific macronutrient: animal protein, vegetable protein, fat and carbohydrate. Animal protein (OR=1.49, 95% CI 1.25 – 1.77, $I^2$=0), carbohydrates (OR=0.49, 95% CI 0.38 – 0.63, $I^2$ =0) and fat (OR = 1.50, 95% CI 1.22 – 1.83, $I^2$=0), were all associated with the risk of GDM in the combined analyses of all ethnic groups (**Appendix Figure A.9)**. These associations were also found to be significant in both the WE and Asian subgroups with comparable effect sizes being identified in both ethnicities. All associations remained significant following a Bonferroni correction and after the HKSJ approach with the exception of animal protein in Asians (**Appendix Figures A.10, A.11, Table A.2)**. Vegetable protein intake was not found to be associated with GDM incidence in any ethnicity.

### 3.5.3.5 Dose-response analyses

Healthy diets, HRs, a Mediterranean diet, prudent diet, plant-based diets, unhealthy diets , western diets, fried/ fast food, meat pattern and high protein diets showed evidence of a dose response following the visual inspection of ORs; however, no strong evidence of a dose response was identified for any exposure (**Appendix Figure A.11, Figure A.12**). Limited evidence of a dose-response relationship was identified for HRs and fried/ fast food intake in WEs ($\beta_{HR}$= -0.098, $\beta_{Fried}$ = 0.138) in the quadratic models however these associations were no longer significant following a Bonferroni correction (**Appendix Table A.3, Figure A.13**). A significant Wald ratio (P

value = 0.0001) was also obtained in the 3K cubic spline model of fried/fast food intake in WEs adding further evidence that any dose response related to fried/fast food intake in WEs may not be linear. However, the quadratic model was consistently found to be the best fitting model for all exposures (**Appendix Table A.3)**.

### 3.5.4 *Post-hoc* analyses: Combination of RCTs and observational studies

For exposures investigated in both RCTs and observational studies (i.e., overall healthy diet exposure, healthy dietary recommendations and the Mediterranean diet), the effect estimates obtained from the analyses of each study design were similar and often had overlapping confidence intervals as well as comparable $I^2$ statistics. As a result of this, a *post-hoc* analysis involving the combination of both study types were performed in order to increase power (**Appendix Figure A.13).** This analysis showed a novel association between healthy diets and GDM in Australian nationals (OR =0.92, 95% CI 0.88 – 0.97, $I^2$=0), while only negligible changes in the effect sizes of other associations were observed. The combination of RCTs and observational studies improved power for analyses in all ethnicities.

### 3.5.5 Sensitivity analyses

Following the combination of study types, sensitivity analyses were performed for analyses with an $I^2 \geq 40\%$ in an attempt to minimise heterogeneity. Sensitivity analyses were executed based upon: (i) assessment of diet during pregnancy, (ii) adjustment for obstetric risk factors (parity, gravidity or multiple pregnancy), (iii) pre-pregnancy BMI, classified with ethnic-specific cut-offs, (iv) maternal age, and (v) Asian subpopulation (East Asian: Chinese and Japanese, South/South-East Asian: Indian subcontinent and Malaysia). When only considering dietary intake during pregnancy, no association was found between healthy diets and GDM in any ethnicity. (**Appendix A, Tables 3-4)**. In addition, when considering the impact of unhealthy diets overall, or the Western diets, no association was found with GDM in overweight/obese WEs. Interestingly, a high protein diet increased odds of GDM in older women that was driven by WEs (OR=1·28;95%CI:1·09-

1·52; $I^2$=0). Almost all sensitivity analyses were well powered (0·80%) to detect an effect size ≥10% with the majority suitably powered to detect an effect size ≥5%. Two exceptions were the assessments of the plant-based diet in overweight women and healthy diets in overweight/obese Asian women.

### 3.5.6 Subgroup differences

Prior to the implementation of a Bonferroni correction, the effect size of the association between unhealthy diets and GDM were significantly different when comparing WEs to Asians (p = 0.03). Effect sizes were not statistically different between any pair of ethnic subgroups for all other exposures.

### 3.5.7 Risk of Bias

No study included in this review exceeded the Academy of Nutrition and Dietetics' exclusion threshold for high ROB (i.e., 6 negative scores) **Appendix Figure A.14)**. On average, RCTs scored positively for ROB 45% of the time, neutrally 34% of the time and negatively 21% of the time, indicating some risk of bias. Areas of the greatest concern were the comparability of study groups, blinding and the management of withdrawals. (**Appendix Figure A.15)**.

On average, observational studies had a lower risk of bias than RCTs, with 55% of studies having a high risk of bias, 22% a neutral risk of bias, and 24% having a high risk of bias (**Appendix Figure A.16)**. Dietary exposures with the highest risk of bias were the carbohydrate, fat and Mediterranean dietary patterns. On the contrary, the prudent diet and fast-food exposures had the lowest risk of bias. When considering RCTS, WE studies had a higher risk of bias than Asian or Mediterranean studies. In observational studies Asian studies had a higher risk of bias than WE or Mediterranean studies, although the results were similar across ethnicities. ROB was similar in all domains when comparing study types (**Appendix Figures A.17, A.18)**. Four analyses, healthy (observational), unhealthy (observational), healthy (combined), healthy recommendations (combined) and unhealthy had a sufficient number of studies (≥10) to assess for funnel plot asymmetry without the introduction of bias (Sterne et al., 2017). The tests for asymmetry of the

funnel plot (the Harbord modified test) were non-significant for both exposures and funnel plots visually looked symmetrical indicating an absence of publication bias. (**Appendix Figure A.19)**

### 3.5.8 Power analysis

All analyses with the exception of the single sensitivity analysis of the effect of consumption of a Western diet during pregnancy had inadequate power to detect a change in odds of 20%. (**Appendix Table A.5)**. In RCTs, power was lowest for the WE subgroup where a change in odds of 18% could be detected for the healthy dietary exposure. In observational studies, power was lowest for the Asian subgroups of Asian traditional and fish exposures where analyses were powered to detect an 18% change in odds.

## 3.6 Discussion

This review aimed to assess the role of diet in GDM prevention in distinct ethnic groups through the evaluation of evidence from both RCT and observational studies. Six RCTs were identified in ethnically distinct populations, 5 of which commenced during pregnancy. No evidence of an association between healthy dietary interventions or healthy dietary interventions which focused on providing healthy dietary recommendations and GDM incidence was obtained before or after stratifying results by ethnicity. This agrees with the most up-to-date Cochrane review that found dietary interventions to have an unknown benefit or harm regarding GDM (Griffith et al., 2020), although no ethnically stratified analyses were presented in this review. Despite these findings, a meta-analysis of 37 RCTs aimed at reducing gestational weight gain did find dietary interventions to be effective in preventing GDM, suggesting that for dietary interventions to be effective at preventing GDM they may need to place a greater emphasis on weight management and may not apply to all women (Bennett et al., 2018).

Through the inclusion of observational studies in the review the impact of pre-pregnancy diet on GDM could be determined. This is the first review to evaluate, by meta-analysis, the impact of diet and ethnicity on GDM as assessed in observational studies. A single systematic review has been

conducted in relation to both cohort and case-control studies, which found a high consumption of cholesterol, heme iron, and processed meat increased risk of GDM, while patterns rich in fruit, wholegrains and vegetables reduced risk of GDM (Schoenaker et al., 2016). However, a high heterogeneity was observed by authors, possibly due to ethnic specific-effects that were not accounted for.

To address this, this study involved the use of meta-analyses performed in ethnic-specific subgroups; thereby, minimising confounding in each ethnic analysis, while permitting a comparison of effect sizes between them. For this, identified dietary exposures were classified as either healthy, unhealthy, or unclassified depending on their constituent components. In observational studies healthy diets were found to associate with a decreased odds of GDM in WEs but not in Asians, while unhealthy diets were found to associate with an increased risk of GDM in WEs but not Asians. Likewise, a meat-based diet and high-protein diet were found to associate with an increased risk of GDM in WEs but not Asians. No additional associations were identified when stratifying the Asian population by their geographical location (i.e., East vs South). A prudent diet, plant-based diet, the sweet and seafood and a traditional Asian diet were not associated with GDM in any ethnicity.

Interestingly, almost all associations were unaffected by mother's age and BMI, suggesting that modified guidelines for WE women at high-risk of GDM due to age or BMI may not be required. The presence of an association in WEs in observational studies but not RCTs could be a result of increased power, or it could highlight the importance of a healthy diet prior to conception. However, future RCTs investigating dietary interventions during 'family planning' are required to test this hypothesis.

When considering macronutrients, an increased intake of animal protein and fat associated with an increased odds of GDM by up to 50% in both WEs and Asians, whereas a carbohydrate-rich diet associated with a reduced the odds of GDM by ≈ 50%. Unfortunately, because all exposures were quantified as % energy intake, it was not possible to tease apart whether the protective effect on GDM was driven by reductions in protein and fat or increased consumption of carbohydrates, or a combination thereof. Interestingly, while animal protein associated with GDM risk in WEs and

Asians, no association was observed with the meat-based dietary pattern in Asians. While the animal-protein diet may have been carbohydrate-rich and negated the effects of high-animal protein, an alternative explanation may be that ethnic-specific foods and cooking methods are difficult to capture with some dietary recall tools.

Due to the limited number of associations identified in non-WE studies and the similar effect sizes and risk of bias obtained in the analysis of RCTs and observational studies, an additional *post-hoc* sensitivity analysis of the study types combined was performed to increase power. Through this, an additional association between healthy diets and GDM in Australian nationals, (a heterogeneous group comprised of WEs and Asians) was identified, but no other additional associations were reported. Overall, this meta-analyses has a moderate risk of bias specifically in terms of blinding and comparability of study groups. Practically however these biases may not be less important in this context in a real-world setting. Blinding of a practitioner to diet is less important in a disease such as GDM with a stringent diagnostic criterion, while differences in the demographics of cases compared to non-cases is expected in relation to factors such as age and BMI.

Despite numerous significant associations between diet and GDM in WEs no consistent evidence was found in other ethnicities. The reason for this is unclear, but the inconsistent reporting (or limited use) of ethnically tailored and culturally informed assessment tools may have contributed to this, especially in RCTs where a third of studies did not report the cultural appropriateness of their interventions (**Appendix Table A.6).** However, in observational settings, the majority of studies did report on the cultural validity of the assessment tools utilised, with only 11% (1/9) of Asian studies failing to do this. This indicates that although the appropriateness of assessment tools may contribute to discrepancies in effect estimates between ethnicities, other factors, including biological factors, are likely to contribute to the observed discrepancies in effects. For example, one potential biological factor driving the disparity in the association between diet and GDM between ethnicities is the metabolome. Indeed, recent work in a the multi-ethnic NUTRIGEN consortium also found diet to have different impacts between ethnicities, with

a plant-based diet increasing birthweight in a SA while reducing birthweight in WEs (Zulyniak et al., 2017).

This study had numerous strengths. Firstly, this is the first study to meta-analyse the results of both RCTs and observational studies in relation to diet and GDM. It is also the first study to examine the role of ethnicity in the relationship between diet and GDM. Furthermore, both the standard meta-analysis method (DL) and the more conservative HKSJ approach were utilised when ≤ 5 studies were available for an exposure. Moreover, a single ROB assessment that is translatable for both RCTs and observational studies permitted comparison of bias between study design. Finally, power analyses confirmed adequate power throughout.

Nonetheless, this study also has several limitations. Firstly, only studies written in English were included in this review. Given that the aim of the review was to identify studies conducted in distinct ethnic populations, studies conducted in non-English speaking regions of the globe (including South Asia) may have been excluded from the review. This could mean that some studies exploring the association between diet and GDM conducted in high-risk ethnicities may have been missed which could have contributed to the absence of effect identified in Asian populations.

Secondly, although all observational studies are limited by confounding, the comparability of the effect estimates obtained from RCTs, and observational studies implies that this was accounted for reasonably well by study authors. It is however likely that some residual confounding does exist (for example as a result of varying population demographics) in the included observational studies which could explain the high heterogeneity for some exposures and can make comparisons between study populations more difficult. In an attempt to help minimise this confounding from known GDM risk factors (including increased maternal age and BMI) a range of sensitivity analyses were utilised in this review.

Thirdly, heterogeneity was present throughout this analysis, likely due to a range of methodological reasons. For example, numerous studies derived dietary patterns via principal component analyses (PCA) conducted on FFQ results. This can introduce heterogeneity due to the consumption of varied foodstuffs between populations that may result in differing components

defining dietary patterns in different populations, depending on how frequently these components were consumed and how their input was characterised by FFQs in individual studies. Furthermore, differing cut-offs of what was considered to be an 'important loading variable in PCA analyses conducted in different studies may also contribute to the observed heterogeneity.

Furthermore, through this method it is possible that the factors which characterise a dietary pattern in one study may not in another, adding further heterogeneity. In addition, the categorical nature used to classify individuals as ‘high consumers’ or ‘low consumers’ are population specific results in the possibility of highest consumers in one study being incompatible with the highest consumers from another study obtained from a different population or region. This use of the categorical scale also disadvantageously prevents the aggregation of multiple relevant effect estimates from a single study into a single estimate due to the inability to assess whether the exposed and unexposed groups were comprised of the same individuals. Specifically, this could result in a potential bias in the analyses of healthy diets in observational studies and in the combined analyses. In order to assess the impact of this bias, one exposure from each relevant study was randomly dropped. This resulted in limited deviations in effect sizes suggesting the impact of this bias to be minor. Similarly because numerous studies utilised the Nurses’ Health Study there is a risk of a type 2 error, as noted by a previously published systematic review of observational studies (Schoenaker et al., 2016). To account for this, studies including the Nurses’ Health Study were dropped at random, which was shown to have no impact on the results.

In conclusion, through the analysis of both RCTs and observational studies the impact of diet on GDM incidence has been assessed in distinct ethnic groups. This analysis has confirmed the presence of an association between healthy and unhealthy diets evaluated in observational studies with GDM in WEs but not in Asians, despite evidence of sufficient power and the wide-spread use of culturally sensitive dietary assessment tools.

## 3.7 Summary

- Healthy dietary interventions administered via RCTs were not found to be effective in reducing GDM incidence in any population.

- In observational studies healthy dietary patterns were found to reduce GDM incidence in WEs but not in Asians.
- Unhealthy dietary patterns assessed in an observational setting were found to increase the odds of GDM development in WEs but not Asians.
- Known GDM risk factors, including age and BMI were not found to influence the identified associations. Furthermore, power was sufficient to detect a ± 0.2 change in odds in the majority of analyses.
- The use of culturally sensitive dietary interventions was somewhat limited in RCTs, however the majority of observational studies utilised culturally sensitive assessment measures to record dietary intake.
- Differences in the associations between dietary patterns and GDM between ethnicities may be of a result of ethnic-specific differences in the biological drivers of GDM.

# Chapter 4: Metabolomic Analysis of the Born in Bradford Cohort: A Multivariate Analysis

Authors original report of the study submitted as Fuller H, et al. to the Journal of Nutrition. https://doi.org/10.1101/2022.04.11.22273658.

## 4.1 Abstract

Despite the high disease burden of GDM, the biological mechanisms driving GDM development are still largely unknown, however metabolism is thought to play a key role. Current evidence that utilises metabolomics to investigating the role of metabolism in GDM developed in South Asians (SAs) is however limited. To address this, this study aimed to (i) characterise the metabolic profiles of GDM in white Europeans (WEs) and SAs and to (ii) evaluate the association between metabolite values and pregnancy dysglycemia in an ethnic-specific manner.

146 metabolite values from fasting serum samples from 2,668 WE and 2,671 SA women from the Born in Bradford (BiB) cohort (mean gestational age 26.1 weeks). These values were analysed using partial least squares discriminatory analyses (PLSDA) to identify metabolites characteristic of GDM in each ethnicity. Linear associations between metabolites and pregnancy dysglycemia were also tested via linear regression.

Seven metabolites associated (VIP ≥1) with GDM in both ethnicities, with an additional 6 associated with GDM in WEs only. Unique metabolic profiles were observed in women of healthy weight who later developed GDM, with distinct metabolite patterns identified by ethnicity and BMI status. Furthermore, lactate, histidine, apolipoprotein A1, HDL cholesterol, HDL2 cholesterol, and DHA, as well as the diameter (nm) of very low-density lipoprotein particles (VLDL_D) were associated with dysglycemia in WEs. In SA women only albumin was associated with dysglycemia. These results suggest that pathways involving fatty acids, cholesterols, glycolysis and amino acids may be characteristic of GDM, in both ethnicities with fatty acids being the most frequently identified class. This is the first study to show that the metabolic patterns of GDM may differ between ethnicities, and highlights the potential need for ethnically appropriate GDM prevention strategies.

## 4.2 Background

During pregnancy, there is a natural shift towards increased catabolism to ensure that energy demands of the foetus are met (Schaefer-Graf et al., 2018; Chen et al., 2018). This process is governed by maternal hormones and starts as a mild reduction in insulin sensitivity, progressing through hyperinsulinemia to controlled insulin resistance by the start of the third trimester (Chen et al., 2018; Mao et al., 2017; Taylor et al., 2019). When this insulin resistance exceeds beyond the normal levels, an intermittent state of hyperglycaemia can ensue resulting in the development of GDM (Wright et al., 2013; Bird et al., 2019). GDM is the most common complication to occur during pregnancy worldwide; its prevalence is estimated to be 2-3-fold higher in SA populations compared to WE populations, with SAs being at increased risk of further health consequences following a GDM diagnosis (McIntyre et al., 2019; Vounzoulaki et al., 2020). This dysregulation in insulin resistance in GDM cases is thought to be a result of metabolomic dysregulation, as seen in T2D. As a consequence of this, the Diabetic Pregnancy Study Group called for increased research into the role of metabolism (through investigation of the metabolome) in the development of GDM in 2018 (Schaefer-Graf et al., 2018). However, to date, the metabolic drivers of GDM remain unclear. Results from different studies have been difficult to reproduce, likely due to varying methodologies (in regard to sample types, metabolite panels and quantification techniques), case-control designs (meaning temporality cannot be inferred) and small, ethnically heterogeneous cohorts (Mao et al., 2017). Indeed, only one study has conducted an analysis of individual metabolites and GDM in an ethnic-specific approach, despite the well-established increased disease risk in SAs. Taylor et al conducted univariate tests of association between numerous metabolites and GDM in the multi-ethnic BiB cohort and demonstrated evidence of ethnic-specific associations between fatty acids and GDM in a stratified sample of WE and SA women.

It is possible that discrepancies in the current evidence for the effectiveness of diet as a means of GDM prevention, particularly in high-risk ethnicities, may be a result of unique shifts in metabolism that predispose certain populations to elevated GDM risk (Chen et al., 2018; Wang et al., 2016; Law and Zhang, 2017; McCabe and Perng, 2017). For example, evidence

from BiB suggests that modified GDM assessment criteria for SAs are required because even at current “safe levels” of glucose thresholds, SAs are at increased risk of delivery complications and new born macrosomia compared to WEs (Farrar et al., 2015).

Currently, evidence highlighting potential ethnic differences in associations between metabolites and GDM has been obtained through the use of univariate statistical analyses. Univariate analyses test for independent associations between each individual metabolite value and a single outcome. However, metabolite profiles are heterogeneous mixtures of metabolites, many of which are strongly correlated and interact with other metabolites to exhibit an effect. Therefore, the use of multivariate approaches that consider all metabolites and their interactions may uncover novel associations that are more reflective of true metabolism. These multivariate techniques can reveal both combinations of metabolites that associate with GDM risk as well as cardinal metabolites that independently associate with GDM risk. Therefore, in this chapter I aim to build upon existing evidence by identifying underlying metabolite patterns that are correlated with GDM to identify ethnic-specific metabolic drivers of GDM risk.

## 4.3 Hypotheses

i. The metabolomic profiles of GDM cases and non-cases are distinct and ethnic-specific.
ii. The relationship between the metabolome and GDM may be moderated by confounders, especially BMI.

## 4.4 Methods

### 4.4.1 Population characteristics

BiB is an ethnically diverse (45% SA) prospective birth cohort recruited from the north of England to examine the determinants of maternal and offspring health during and after pregnancy, and their association with the future health of the mother and offspring (Wright et al., 2013). BiB recruited

12,453 women (26-28 weeks' gestation, mean maternal age 27.8) at the Bradford Royal Infirmary between 2007 and 2010, collecting baseline data on 13,776 pregnancies and 13,858 births (Wright et al., 2013; Bird et al., 2019). Of these, 11,480 (~80%) women provided fasting serum blood samples for metabolite analyses, taken at the time of the 75g oral glucose tolerance (OGTT) test for GDM assessment, offered to all women at the Bradford Royal Infirmary (Raynor and Born in Bradford Collaborative, 2008). Written consent was gained from all participants and ethical approval was granted by the Bradford Research Ethics Committee (ref07/H1302/112) (Wright et al., 2013).

### 4.4.2 Blood metabolite analysis

Overnight fasting serum blood samples were taken at the Bradford Royal Infirmary by trained phlebotomists, processed in 2.5 hours and stored at -80°C in the absence of freeze-thaw cycles (Taylor et al., 2019; Taylor et al., 2021). Samples were processed using a previously validated high-throughput automated NMR platform (Nightingale Health ©; Helsinki, Finland), which utilized 3 molecular windows in the same experimental set-up to measure each sample, two from the native sample (LIPO and LMWM, 37°C) and one from lipid serum extracts (LIPID, 22°C). These 3 platforms quantify distinct metabolites: LIPO includes spectra from a broad range of lipids found in a range of lipoproteins, LMWM represents spectra for a range of low-molecular-weight molecules while the LIPID window is acquired with a standard 1D spectrum involving 32 transients (Taylor et al., 2019; Wurtz et al., 2015). All spectra were quantified using a Bruker ADVANCE III spectrometer at 600 MHz (Taylor et al., 2019).

Of the 227 metabolites quantified, 146 were presented as absolute values and 81 were relative measures of percentages and ratios. To ease interpretation and minimize redundancy this analysis focused on the 146 metabolite values expressed as absolute measures. This panel comprised measures of 98 lipoproteins, 9 amino acids, 2 apolipoproteins, 9 cholesterols, 8 fatty acids, 8 glycerides and phospholipids, 4 glycolysis-related metabolites, 2 ketone bodies, 3 measures of fluid balance and inflammation, and 3 measures of the mean lipoprotein particle diameter **(Appendix Table B.1)**. In regard to fatty acids, all were assumed to be *cis* isomers because typical levels

of *trans* isomers are below the detection threshold (Santos Ferreira et al., 2017). For lipid data, high-performance liquid chromatography was used to calibrate the quantification process, with lipids individually being cross-validated against NMR-independent lipid data (Taylor et al., 2021). Spectra for low molecular weight molecules were calibrated against a panel of manually fitted metabolite values.

### 4.4.3 Participant selection

#### 4.4.3.1 Metabolite data and imputation

Of the 11,480 blood samples analyzed for metabolites, 54 samples were excluded because they failed one of five Nightingale Health © quality control measures (low glucose, high lactate, high pyruvate, low protein concentration and plasma samples). Glucose levels are well defined biologically, particularly for values at the low end of the distribution. Deviations from these expected ranges of values can therefore indicate a quantification error. Alongside high levels of pyruvate and lactate, low glucose levels can also indicate increased metabolism post-collection due to samples being kept at room temperature for extended periods before freezing. Furthermore, sample dilution can also be determined by a low protein concentration (signified by low albumin), also indicating a low-quality sample. Of the 11,426 remaining samples, ~3% were missing ≥1 metabolite value for unknown reasons. To test whether these values were missing at random, the structure of missing data was assessed via the *visualization and imputation of missing values* (*VIM*) package in R (Alexander Kowarik, 2016). This package examines the frequency of all possible combinations of missing data to test for non-random distributions of missingness. In addition, multiple correspondence analysis (MCA) was also implemented to assess the randomness of missing data. No combination of metabolites was identified to be frequently missing in the data via the use of the *VIM* package. Likewise, following the utilization of MCA analysis, having low levels of a metabolite (i.e., quartile 1 of the distribution) was found to correlate with having a low-level of another metabolite value, while having high levels of metabolites (i.e., quartile 4) was found to correlate with having high levels of other metabolites. Hence there was no evidence that missingness

occurred in a non-random pattern, suggesting that it was appropriate to impute missing values **(Appendix Figure B.1)**.

Optimised multiple imputation with iterative principal component analysis (PCA; 100 simulations, K-fold cross validation) based upon the minimisation of mean square error of prediction (MSEP) was performed using the *missMDA* package (Josse and Husson, 2016). Unlike non-multivariate methods of imputation (such as imputing missing values with the mean value) this technique considers the data structure during imputation, leading to more accurate imputation. The impact of mothers with ≥3% metabolite values missing on imputation quality was assessed through the exclusion of these individuals ($n_{excluded}$ = 88), resulting in no detectable difference in imputation quality. Therefore, all 11,426 samples were included for imputation.

### 4.4.3.2 Descriptive data

Imputed metabolite data were then combined with descriptive BiB reported characteristics, including participant's ethnicity, GDM status, gestational age at sample collection, history of diabetes, age, BMI, smoking status, parity and whether they were carrying a singleton/multiple pregnancy. Ethnicity was self-reported and based upon guidelines from the UK Office of National Statistics (ONS) (Taylor et al., 2019). Participants whose samples were collected after GDM diagnosis (28$^{th}$ week or later) were excluded from the analysis as well as mothers with a history of diabetes. Individuals who reported being of South Asian origin other than Pakistani (PK) were also excluded, due to the small sample size (therefore limited power) of other SA ancestry groups. Utilising the ONS guidelines, for a participant to be categorised as PK they had to be classified as belonging to the Asian/British Asian subcategory and be labelled as having PK ancestry. From here on in, the included SA group hence refers to the PK population in the BiB cohort. Additional information on the country of birth was obtained for SA women, along with the country of birth of their parents and partner (Lawlor et al., 2014). WEs were categorized as individuals of either WE or White British descent (Taylor et al., 2019). If a participant's ethnicity was not collected at recruitment the information was obtained from primary medical records which utilise a

similar classification for ethnicity as the ONS. Information on parity and the number of registerable births was also obtained from primary care records.

In total, 5,339 participants, 2,671 SA and 2,668 WE women were retained for analysis **(Appendix Figure B.2)**. All women were recruited prior to their scheduled GDM assessment (mean gestational age 26.1 weeks), and prior to the 28th week of pregnancy. GDM was diagnosed using a modified version of the World Health Organization criteria: a 75g OGTT of fasting glucose ≥6.1 mmol/L and/or a 2-hour post-load glucose level of ≥ 7.8mmol/L at 26-28 weeks of pregnancy (Taylor et al., 2019; Wright et al., 2013). Maternal age was recorded at pregnancy booking and BMI was calculated using height measured at recruitment and maternal weight recorded at the first antenatal visit. When examined as a categorical variable, ethnic-specific cut-offs were used to classify mothers into BMI groups (overweight: 25-29.9 kg/$m^2$ for WE or 23-27.4 kg/$m^2$ for SA women; obese: > 30kg/$m^2$ for WE or >27.5kg/$m^2$ for SA women) (WHO Expert Consultation, 2004). Smoking status was self-reported at baseline.

### 4.4.4 Unsupervised analyses

PCA is a dimensionality reduction technique that utilises the correlation structure in large scale datasets to reduce the number of components in a dataset while capturing the maximum proportion of the dataset's variation (Jolliffe and Cadima, 2016) During this process, the initial data are transformed into a new dataset of uncorrelated principal components (PCs), which contain linear combinations of the initial variables and progressively explain smaller proportions of the variation of the dataset. As PCA utilises variation in a dataset to perform dimensionality reduction it is an ideal tool to uncover underlying grouping in metabolite datasets as it will also separate uncorrelated groups from one another, while highly correlated data points will cluster together. PCA determines this grouping in an unsupervised way, meaning that the algorithm is not informed of the potentially desired expected groupings prior to analyses and will only separate these groupings if this separation arises during the PCA process from the underlining correlation structure of the data (Jolliffe and Cadima, 2016).

Prior to PCA, the Normality of metabolite values was assessed using histograms and Q-Q plots. Most metabolite values (136/146) required Normalisation. Initially, a log Normalisation was implemented, however, if metabolite distributions still appeared to deviate from Normal (87/146 metabolites), a square-root transformation was utilised. In scenarios where metabolite distributions still appeared to be non-Normal (42/146) a Normal score transformation (NST) was implemented via the *rcompanion* package (Mangiafico, 2021). Data were Pareto scaled and mean-centred prior to analyses. PCA was performed using the *prcomp* function. All analyses were conducted in R version 4.0.2 (R Development Core Team, 2020).

### 4.4.5 Metabolite discriminatory analysis

Partial least squares discriminatory analysis (PLSDA) is a supervised dimensionality reduction technique. PLSDA utilises all included variables to discriminate group data based upon predefined outcome groups by trying to maximise the covariance explained between the input and output variables, as opposed to PCA which aims to maximise the variance explained in the output variables. In PLSDA, included variables are then ranked by the degree to which they explain the variance between groups (i.e., GDM vs non-GDM). These are known as variable importance in the projection (VIPs), where VIPs ≥1 denote a variable with good discriminatory quality and predictive ability (Perreault et al., 2014; Badoud et al., 2014). Through PLSDA it is possible to assess the predictive capacity of metabolite values for GDM in models unadjusted and adjusted for other known GDM risk factors. Models were adjusted for BMI, maternal age, parity, multiple pregnancies, and smoking status (yes/no) initially and then additionally for ethnicity to identify metabolite values predictive of GDM in the whole population. Following this PLSDA models were fitted in each ethnic group to determine whether ethnic-specific effects were present. To assess bi-directionality, models predicting ethnicity were fitted in the overall population and GDM cases/ non-cases separately using the same criteria as above.

The optimal number of components to include in the model was selected based upon the component's ability to significantly predict group membership in the training ($pR^2Y$ =0.05) and validation ($pQ^2Y$= 0.05) datasets

using the NIPALS algorithm and 7-fold cross validation. When multiple components' were significantly predictive ($pR^2Y \leq 0.05$, $pQ^2Y \leq 0.05$), the component that best discriminated between groups (i.e., maximisation of outcome variance explained, $R^2Y$) with the minimal error (root mean squared error of estimation (RMSEE)) was selected. External validity was assessed via 7-fold cross-validation. PLSDA models were performed using the *ropls* package in R (Thevenot et al., 2015). When the size of the outcome groups differed by ≥ 1% the larger group was randomly sampled (n=20) to minimise error. The VIPs for each metabolite were mean averaged across all significant iterations ($pR^2Y = 0.05$, $pQ^2Y = 0.05$) and presented with their standard errors (SEs) following the removal of outlier VIPs, defined as 1.5 x interquartile range of VIP values. Differences in the distribution of VIPs between ethnicities and case status were assessed for significant iterations using a Mann-Whitney (MW) test; this was possible because all comparisons were tested against the same panel of metabolite measures.

## 4.5 Results

### 4.5.1 Population characteristics

In the overall study population, women's mean age was 27.3 years, had a mean BMI of 26.2, with 59% of the population being considered overweight or obese when using ethnic-specific BMI cut-offs. (**Table 4.1).** The included pregnancy was the first pregnancy for most women (43.2%) and almost all were singleton pregnancies (98.8%). GDM cases were significantly older and heavier than non-cases in the overall population and in both ethnic groups. In WEs, GDM cases were more likely to be smokers compared to non-cases. This association was not identified in SAs, possibly due to the small proportion of SA smokers (3.3%). SA women were significantly younger than WE women in the overall population and in GDM non-cases, but not in GDM cases and had a significantly higher BMI than WEs in all three groupings. No difference in the proportion of singleton pregnancies (>97%) was observed between WE women and SA women. Alcohol intake was not assessed because it was reported by only 1% of SA women. The mean gestational age at time of sample collection was 187 days.

### 4.5.2 Unsupervised analysis

No overt differences were seen between ethnic groups in the overall set of metabolites or in any one metabolite class via PCA. Likewise, no difference between ethnicities was seen when the analysis was applied only to first-time mothers or BMI group (examined as a categorical or examines a binary variable (above/below overweight BMI threshold)). When examined as a binary or categorical variable BMI was classified utilising ethnic-specific BMI cut-offs for SAs and WEs: 23 $kg/m^2$ and 27.5 $kg/m^2$ were classified as overweight or obese in SAs; 25$kg/m^2$ and 30$kg/m^2$ were classified as overweight or obese in WEs (WHO Expert Consultation, 2004). In the PCA analysis of the included study sample (n=5,339), PC1 explained 63.6% of the variation in metabolite values while PC2, PC3, PC4, PC5 explained 19%, 7.5%, 2.8% and 1.8% respectively. Cumulatively the first 10 PCs explained 98% of the variation in the dataset indicating that the metabolite values were highly correlated. The lack of separation (assessed via visual inspection of PCA plots) between groups in an unsupervised method highlighted the need for supervised multivariate statistical methods to determine metabolite values characteristic of GDM status.

## Table 4.1: Characteristics of individuals included in the multivariate analyses.

| Population Characteristics | Overall | White European (WE) | | | | South Asian (SA) | | | | WE vs SA P values | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Total (n=5,339) | Total (n=2,668) | GDM (n=128) | Non-GDM (n=2,540) | P value | Total (n=2,671) | GDM (n=286) | Non-GDM (n=2,385) | P value | Overall (n=5,339) | GDM (n=414) | Non-GDM (n=4,925) |
| **Age** | 27.3 (0.08) | 26.7 (0.1) | 30 (0.5) | 26.5 (0.1) | **< 0.001** | 25.7 (0.1) | 30.6 (0.3) | 27.6 (0.1) | **< 0.001** | **< 0.001** | 0.29 | **< 0.001** |
| **Mother's Booking Weight (Kg)** | 68.8 (0.2) | 72.0 (0.3) | 76.8 (1.6) | 71.8 (0.3) | **0.002** | 65.5 (0.3) | 72.3 (0.9) | 64.7 (0.3) | **< 0.001** | **< 0.001** | **0.03** | **< 0.001** |
| **Mother's height (cm)** | 162 (0.09) | 164.3 (0.1) | 163.8 (0.5) | 164.3 (0.1) | 0.47 | 159.7 (0.1) | 158.1 (0.3) | 159.9 (0.1) | **< 0.001** | **< 0.001** | **< 0.001** | **< 0.001** |
| **BMI** | | | | | | | | | | | | |
| ***mean (kg/m$^2$)*** | 26.2 (0.08) | 26.7 (0.1) | 28.5 (0.5) | 26.6 (0.1) | **< 0.001** | 25.7 (0.1) | 28.9 (0.4) | 25.3 (0.1) | **< 0.001** | **< 0.001** | 0.49 | **< 0.001** |
| ***underweight/normal*** | 2198 (41.2) | 1254 (47) | 47 (36.7%) | 1207 (47.5%) | **0.02** | 944 (35.3) | 46 (16.1) | 898 (37.7) | **< 0.001** | **< 0.001** | **< 0.001** | **< 0.001** |
| ***overweight or obese*** | 3141 (58.8) | 1414 (53) | 81 (63.3%) | 1333 (52.5%) | | 1727 (64.7) | 240 (83.9) | 1487 (62.3) | | | | |
| **Parity** | | | | | | | | | | | | |
| ***0*** | 2311 (43.2) | 1394 (52.2) | 73 (57) | 1321 (52) | | 917 (34.3) | 82 (28.7) | 835 (36.5) | | | | |
| ***1*** | 1508 (28.2) | 813 (30) | 39 (30.5) | 774 (30.5) | 0.47 | 695 (26) | 49 (17.1) | 646 (28.3) | **< 0.001** | **< 0.001** | **< 0.001** | **< 0.001** |
| ***2*** | 813 (15.2) | 293 (11) | 11 (8.6) | 282 (11.1) | | 520 (19.5) | 57 (19.9) | 463 (20.3) | | | | |
| ***≥ 3*** | 707 (13.2) | 168 (6.3) | 5 (3.9) | 163 (10.4) | | 539 (20.2) | 98 (34.3) | 441 (14.9) | | | | |
| **Singleton pregnancy (%)** | 5274 (98.8) | 2634 (98.7) | 123 (96.1) | 2511 (98.9) | **0.01** | 2640 (98.8) | 280 (97.9) | 2360 (99) | 0.12 | 0.70 | 0.29 | 0.75 |
| **Smoked during pregnancy (%)** | 958 (17.9) | 870 (32.6) | 25 (19.5) | 845 (33.3) | **0.001** | 88 (3.3) | 12 (4.2) | 76 (3.2) | 0.37 | **< 0.001** | **<0.001** | **< 0.001** |

Summary table of population characteristics. Continuous variables are expressed as a mean and standard error (SE). Categorical variables are expressed as counts (%). Differences between women with and without GDM for continuous variables were tested using a Mann-Whitney test. Differences for categorical variables were tested using Pearson's Chi-squared test. Significant p values are shown in bold.

### 4.5.3 Metabolite characterisation of GDM

In the analysis of the full study sample (n=5,339) PLSDA models including known GDM risk factors (i.e., BMI, age, parity, multiple pregnancy, smoking status, and ethnicity) explained 21.7% of the variation between GDM and non-GDM groups in optimised models (i.e., minimisation of RMSEE and maximisation of $R^2Y$). All included covariates had an average VIP ≥1 confirming their importance as GDM risk factors. The model including metabolite values and known GDM risk factors explained an additional 9.3% of the variance when compared to the model containing only known GDM risk factors and significantly characterised GDM status (p value $R^2 < 0.05$, p value $Q^2<0.05$). Models only including these risk factors (i.e., BMI, age, parity, multiple pregnancy, smoking status, and ethnicity) explained 12.4% of the variation between cases and non-cases and were also statistically significant, confirming their ability to characterise GDM cases from non-cases. On the contrary, although the model containing only metabolite values explained a comparable amount of the variance in GDM status (13.5%) this model could not significantly predict GDM status (p value $R^2 > 0.05$, p value $Q^2>0.05$).

In total 6 metabolite values were important in the prediction of GDM status in the overall population when models were adjusted for maternal age, BMI (continuous), smoking status, parity, and multiple pregnancy status. These were lactate, VLDL_D, total fatty acids, total monounsaturated fatty acids (MUFAs), 18:2 linoleic acid and total saturated fatty acids (SUFAs) (**Figure 4.1, Table 4.2**). Following the addition of ethnicity into the model, 26.6% of the variation in GDM case status was explained, an additional 3% compared to when ethnicity was not included in the model (i.e., model 1).

The addition of ethnicity into the model also resulted in an additional metabolite, esterified cholesterol, which was also found to be important in the characterisation of GDM status, resulting in a total of 7 metabolites being identified as important variables ( VIP ≥1). 4 of these metabolites were fatty acid measures, 1 was a measure of lipoprotein density, 1 was a glycolysis related metabolite and 1 cholesterol measure.

**Table 4.2: VIPs of variables that discriminate GDM from non-GDM.**

| Variable | Model 1 | Model 2 |
|---|---|---|
| **Age** | 6.4 (0.03) | 5.9 (0.03) |
| **BMI** | 5.4 (0.04) | 5.1 (0.02) |
| **Ethnicity** | - | 2.9 (0.02) |
| **Parity** | 2.4 (0.01) | 2.3 (0.01) |
| **Smoking Status** | 1.9 (0.02) | 1.7 (0.01) |
| **Multiple Pregnancy** | 1.5 (0.01) | 1.3 (0.009) |
| **Lactate** | 1.5 (0.01) | 1.2 (0.008) |
| **VLDL_D** | 1.3 (0.01) | 1.3 (0.01) |
| **Total FAs** | 1.2 (0.01) | 1.5 (0.01) |
| **Total MUFA** | 1.2 (0.001) | 1.2 (0.008) |
| **18:2 Linoleic Acid** | 1.1 (0.01) | 1.1 (0.004) |
| **Total SFA** | 1.1 (0.01) | 1.2 (0.007) |
| **Esterified Cholesterol** | - | 1.0 (0.008) |

Mean VIP scores of important variables (VIP ≥1) and standard errors (SE) across 20 model iterations. Model 1: Adjusted for maternal age, BMI (continuous), smoking status, parity, and multiple pregnancy status. Model 2: Model 1 adjustment + ethnicity. SFA: total saturated fatty acids. MUFA: total monounsaturated fatty acids. VLDL_D: diameter of very-low density lipoproteins.

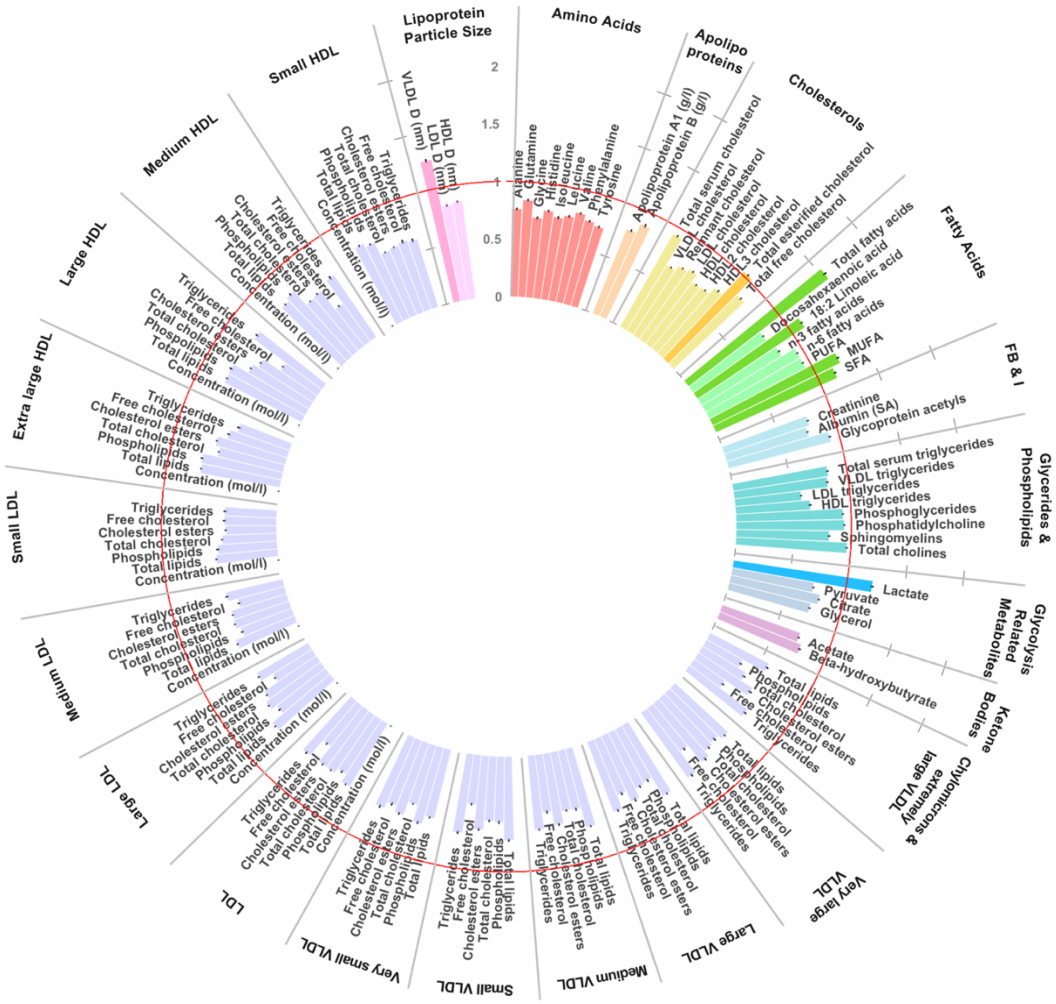



**Figure 4.1: VIPs for metabolites in the characterisation of GDM.** VIPs from PLSDA models adjusted for maternal age, BMI, smoking status, parity, multiple pregnancy status, and ethnicity. Red line denotes VIP cut-off of 1. Bolder bars indicate metabolites with a VIP ≥ 1. No lipoprotein had a VIP ≥1 so the lipoprotein class was not included in this plot in order to conserve space. GRM: Glycolysis Related Metabolites. LPS: Lipoprotein Particle Size. MUFA: total monounsaturated fatty acids. SFA: total saturated fatty acids. VLDL_D: mean diameter of very-low density lipoproteins. Units mmol/L unless stated.

### 4.5.4 Characterisation of GDM: Ethnically stratified analysis

In an ethnically stratified analysis (20 iterations), models that only included established risk factors for GDM significantly predicted GDM and explained a median of 3.3% of the variation in GDM case status in SAs and a median of 12.8% of the variation in WEs. Conversely, models with only metabolite values were not significant but explained a median of 6.5% of the variation in GDM status in SAs women and 5.8% of the variation in WE women. The combination of models resulted in significant prediction of GDM in both ethnicities, with models explaining 26% of the variance in GDM status in WEs and 20% of this variation in SAs (**Appendix Table B.2)**. In the WE population maternal age was the most important predictor of GDM status (VIP = 5.99) whereas in SAs, BMI was the most important predictor (VIP = 7.06). When comparing the VIPs obtained across model iterations in both ethnicities via a Mann-Whitney (MW) test, only the importance of BMI and smoking status on GDM status significantly differed between ethnicities, although all adjusted covariates were categorized as important in both ethnicities (**Appendix Table B.3)**.

After adjusting for known GDM risk factors (i.e., BMI, age, parity, multiple pregnancy, smoking status) 7 metabolite values characterized GDM status in both ethnicities: total fatty acids, total MUFA, total SFA, linoleic acid, glycoprotein acetyls, lactate, and mean diameter of VLDL. Aside from glycoprotein acetyls, these were the same metabolites that characterised GDM in the analysis of the overall cohort before stratification by ethnicity (**Table 4.2**). Three of these metabolites, lactate, glycoprotein acetyls and linoleic acid were found to characterise GDM comparatively well between the ethnicities (VIP≥1; MW p value>0.05), while total fatty acids, total MUFA, total SFA and VLDL_D were more robust predictors in WEs (VIP≥1 in both ethnicities; MW p value<0.05 between ethnicities) **(Figure 4.2, Appendix Table B.4)**. Additionally, alanine, glutamine, total cholesterol, total n-6 PUFA, total PUFA, and citrate were markers (VIP≥1) of GDM status in WE women only. No markers of GDM were specific to SA women.

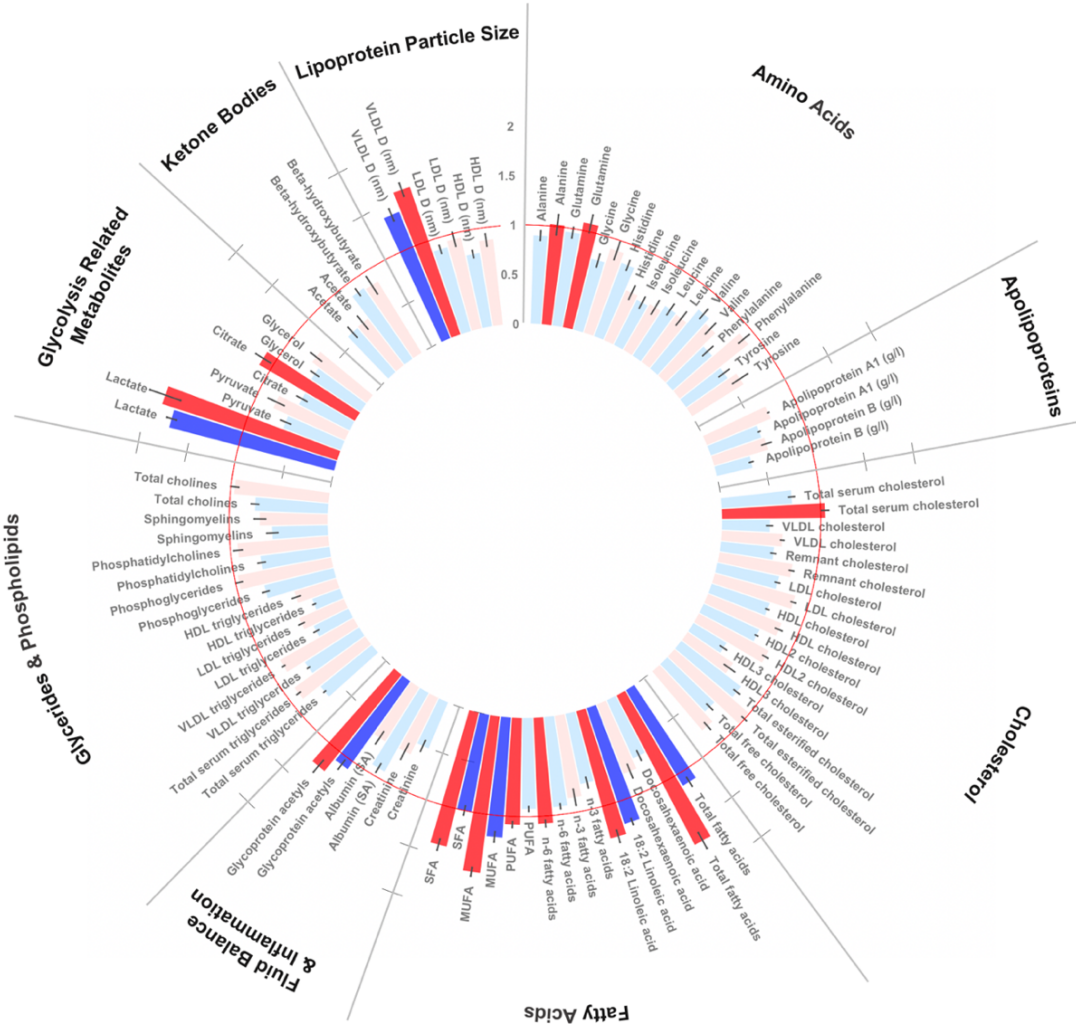



**Figure 4.2: VIPs for metabolites distinguishing GDM and non-GDM women in an ethnically stratified analyses.** VIPs for PLSDA models in South Asians (blue) and white Europeans (red). Models were adjusted for maternal age, BMI, smoking status, parity, and multiple pregnancy status. Red circular line denotes VIP cut-off of 1. Bolder bars indicate metabolites with a VIP ≥ 1. No lipoprotein demonstrated a VIP ≥1 and were not included in the figure to preserve space. GRM: Glycolysis Related Metabolites. LPS: Lipoprotein Particle Size. MUFA: total monounsaturated fatty acids. SFA: total saturated fatty acids. VLDL_D: diameter of very-low density lipoproteins.

### 4.5.5 Metabolites characteristic of ethnicity

In order to explore the underlying metabolic profiles in each ethnic group, PLSDA models with ethnicity as an outcome variable were conducted in the whole population and then in GDM cases and in GDM non-cases separately. In PLSDA (i.e., adjusted for maternal age, smoking status, parity,

BMI, and GDM status), 12 metabolic measures characterised ethnicity in the overall population: total fatty acids, total serum cholesterol, total saturated fatty acids (SFA), total MUFA, n-6 fatty acids (FAw6), total esterified cholesterol, 18:2 linoleic acid (LA), remnant cholesterol, phosphatidylcholine and total cholesterol **(Appendix Table B.5).**

Following analyses in GDM cases and controls separately 9 metabolite values were found to be predictive of ethnicity in both the case and non-case stratum (total serum cholesterol, LDL cholesterol, total esterified cholesterol, n-3 fatty acids (FAw3), total PUFAs, total MUFAs, total SFAs, phosphatidylcholine and total chlorines), 5 of which were identified in the analyses of the overall population. Additionally, 6 metabolite values (alanine, total fatty acids, LA, glycoprotein acetyls, lactate and mean diameter of VLDL (VLDL_D)) were found to characterise ethnicity in GDM cases but not GDM non-cases, indicating that these metabolites may be part of ethnic-specific pathways involved in the development of GDM. Finally, 1 metabolite (remnant cholesterol) was predictive of ethnicity only in non-cases, potentially indicating shared pathways of GDM development. **(Appendix Table B.3, Figure B.3)**.

### 4.5.6 *Post-hoc* analyses: Characterisation of smoking status

Smoking is a very well-established risk factor for many non-communicable diseases, including GDM (McIntyre et al., 2019). In addition to the suspected causal associations between smoking and a range of diseases, smoking is highly correlated to other demographic factors, including socioeconomic status and diet. Furthermore, smoking prevalence is known to vary between ethnicities, with a significantly higher proportion of WE in this study cohort smoking compared to SA women (**Table 4.2**).

To assess the impact of smoking on the results, women were stratified by their smoking status during pregnancy (yes/no) and their ethnicity and PLSDA models predicting GDM status were performed as before (i.e., 20 iterations averaged over components with a $pR^2Y \leq 0.05$ and $pQ^2Y \leq 0.05$). Models were adjusted for BMI, age, parity and multiple pregnancy.

Limited evidence was found to support differences in the VIP scores for the majority of metabolites in this analysis following stratification by ethnicity and smoking status ***(*Appendix Figure B.4)***. LDL cholesterol (LDL-C),

phosphatidylcholine, remnant cholesterol (Remnant-C), serum triglycerides (serum TG) and VLDL triglycerides (VLDL-TG) appeared to be more important in determining GDM in WE smokers, however, the evidence in SA was limited due to the low number of significant iterations, likely due to a restricted sample size of this group (12 cases, 76 non-cases).

### 4.5.7 *Post-hoc* analysis: Characterisation of GDM in low-risk women

An increased BMI is an established risk factor for GDM and it has also been shown that BMI has an impact on the metabolome. Therefore, it is possible that BMI is a mediator along the causal pathway that links metabolism and GDM (McIntyre et al., 2019). Moreover, SA women have been found to be at an increased risk of GDM at lower BMIs, indicating that it is possible that BMI has differential impacts on GDM risk between ethnicities (Read et al., 2021). Indeed, in this study, BMI was found to have a significantly higher VIP score in SA compared to WEs ($VIP_{SA}$ = 7.06 ± 0.22 vs. $VIP_{WE}$ = 4.33 ± 0.22; p value<0.001).

To explore this finding further, the ethnic-specific impact of BMI on the metabolome and subsequent GDM diagnoses was investigated *post-hoc* using sPLSDA. sPLSDA is a supervised multivariate technique with the ability to predict group membership in multiclass problems (i.e., stratification by ethnicity, bodyweight, and GDM status) by simultaneously performing variable selection (in turn reducing noise) with group discrimination (Lê Cao et al., 2011). Women were classified as 'healthy or 'overweight' based upon ethnic-specific cut-offs (BMI ≥ 25kg/$m^2$ for WE women and BMI≥ 23kg/$m^2$ for SA women). To examine the role of BMI and ethnicity in the absence of other GDM risk factors mothers who were either (i) carrying a multiple pregnancy (ii) had previous children, (iii) smoked during pregnancy, or (iiii) were ≥35 years of age were excluded from the analysis. This resulted in a sample of 1,736 low-risk women (872 WE, 864 SA) whose only major risk observable GDM risk factors were ethnicity or BMI status. Furthermore, as sPLSDA performs variable selection prior to group discrimination the inclusion of GDM risk factors in the model would result in the inclusion of fewer metabolites due to the fact that these risk factors are likely to be more important determinants

of GDM status than individual metabolites, meaning GDM risk factors would be included in the models in place of metabolite values.

Healthy weight SA women who developed GDM ($SA_{Healthy\text{-}GDM}$) presented the most distinct metabolic profile (Receiver Operator Curve; ROC = 0.783) and were most similar to healthy weight WE women who developed GDM ($WE_{Healthy\text{-}GDM}$; ROC = 0.691) **(Appendix Figure B.5**). The remaining groups of SA women and WE women, including all non-GDM and overweight women, were indistinguishable from each other. When the dataset was reduced to only the 4 GDM groups, healthy SA women ($SA_{Healthy\text{-}GDM}$) remained distinguishable. A sensitivity analysis involving the removal of underweight mothers ($n_{removed}$ = 93, BMI ≤18.5 kg/$m^2$) was conducted due to the higher proportion of underweight SA mothers, however, this did not affect the outcome.

Metabolites selected by sPLSDA in each comparison were included in PLSDA models (20 iterations) alongside highly correlated metabolites (Pearson's correlation coefficient ≥0.9) to identify key metabolic drivers of this separation. Alanine, glutamine, and glycerol were important to distinguish healthy weight SA women who developed GDM ($SA_{Healthy\text{-}GDM}$) from all others, while fatty acids were important to distinguish $SA_{Healthy\text{-}GDM}$ from other GDM cases. Interestingly, in healthy women, aromatic and BCAAs distinguished GDM and non-GDM women between (but not in) ethnic groups **(Appendix Figure B.6**). Glycerol distributions were significantly different in all comparisons (MW p value <0.05).

#### 4.5.7.1 Characterisation of GDM in low-risk women by BMI and ethnicity

In addition to sPLSDA, orthogonal partial least squares discriminant analysis (oPLSDA) was utilised as a sensitivity analysis. oPLSDA is a supervised multivariate technique that separates variation in each predictor variable based upon its linear (correlated) and orthogonal (uncorrelated) association with the outcome variable (Worley and Powers, 2013; Blasco et al., 2015). This can provide better separation along fewer components when a large proportion of variance in the dataset does not directly correlate with the outcome variable. Furthermore, through the creation of shared and unique

structure (SUS) plots it is possible to determine shared and unique factors of a main group of interest from the two most relevant comparisons. In this instance, the main group of interest is the healthy weight SA cases (SAC-N) because this group developed GDM in the absence of the other key GDM risk factors. By comparing this group to healthy weight SA non-cases (SANC-N) and healthy weight WE cases (WEC-N) it may be possible to determine unique metabolic features that (i) characterise GDM development in "apparently" low-risk SAs and (ii) highlight ethnic-specific metabolic features that associate with GDM in low-risk women.

In this analysis, no significant separation was identified via oPLSDA and the creation of SUS plots between these three groups. Following the inclusion of BMI and age in the models the SAC-N group separated from the other three groups. In this model, pyruvate, L-HDL and XL-HDL contributed toward the separation of the SAC-N group but with low reliability, as shown in SUS plots (**Appendix Figure B.7)**.

### 4.5.8 *Post-hoc* analysis: Correlation between identified metabolites and postprandial glucose measures

#### 4.5.8.1 Fasting glucose

In a *post-hoc* analysis correlation between metabolite values and the composite glucose measures utilised by BiB for GDM diagnosis (i.e., fasting glucose and 2-hour post glucose) were assessed via the assessment of Pearson correlation coefficient with each ethnic stratum. Both fasting glucose and 2-hour post glucose values were log normalised before analyses.

Nineteen metabolite values were found to be significantly correlated (1 positive, 18 negative) with fasting glucose in WE after stratification by ethnicity. The majority of these correlated metabolites were HDL measures, except for lactate, mean diameter of LDL and Apolipoprotein A1 **(Appendix Figure B.8).** Overall, significant correlations ranged from R=0.042 to 0.064 indicating that metabolites were associated with postprandial glucose measures during pregnancy and explained a reasonable amount of variation in these glucose levels. No metabolite values were significantly correlated with fasting glucose in SAs. Following a 2x2 stratification by ethnicity and case

status, significant correlations were only observed in WE non-cases, where 1 metabolite was positively correlated with fasting glucose levels (mean diameter LDL) and 20 were negatively correlated with fasting glucose levels.

#### 4.5.8.2 2-hour post glucose

No metabolite values were correlated with 2-hour post glucose in either ethnic group. When the sample was stratified by case status, 4 metabolite values (DHA, concentration of XL-HDL (XL-HDL-P), phospholipids in X-HDL (X-HDL-PL), and lipids in XL-HDL (XL-HDL-L)) were positively correlated ($R \approx 0.03$, p value <0.05) with 2-hour post glucose. No associations were observed in cases **(Appendix Figure B.9).** Following stratification by ethnicity and case status, DHA was significantly correlated with 2-hour post-glucose ($R \approx 0.05$, p value = 0.018) in WE non-cases. While in WE cases, lactate was negatively correlated ($R \approx -0.20$, p value = 0.03) and glycine and histidine were positively correlated to a similar degree ($R \approx 0.03$, p value <0.05) with 2-hour post-glucose.

### 4.5.9 *Post-hoc* analysis: Linear regression between identified metabolites and postprandial glucose measures

In light of the observed correlation between metabolites and postprandial glucose measures metabolite values identified in multivariate analyses were regression on the outcomes of fasting glucose and 2-hour post glucose (two key components of an OGTT) to test for the presence of linear associations. Linear regression models were adjusted for maternal age, gestational age, parity, and smoking status during pregnancy. As before, post-prandial glucose measures were log normalised prior to analysis. When metabolite values were identified as significant predictors in linear regression (p value ≤ 0.05), BMI was added to the model first as a continuous variable and then alternatively as a dichotomous variable of overweight status (assessed using ethnic-specific BMI cut-offs) to assess the role of BMI as a mediator of dysglycemia during pregnancy. In order to assess multicollinearity between model covariates variance inflation factors (VIFs) were calculated in R. In the overall population, all covariates had an average VIF below 1.35 indicating the absence of multicollinearity (Johnston et al., 2018).

#### 4.5.9.1 Fasting glucose

Three metabolite values, Lactate, Albumin and mean diameter of LDL particles (LDL_D), were associated with fasting glucose levels in the overall population in all models **(Appendix Table B.6)**. Following the inclusion of BMI (dichotomous) in the model, a 1 unit increase in lactate (mmol/L) was associated with a 0.01 mmol/l decrease in fasting glucose (SE 0.002) while a 1 unit increase of albumin signal area was associated with a decrease of 0.61 in fasting glucose (SE 0.23). A 1nm increase in the mean diameter of LDL particles was associated with a 0.05 increase in fasting glucose (SE 0.02). When stratified by ethnicity, Lactate and mean diameter of LDL particles were associated with fasting glucose levels WEs only, to a similar extent to that seen in the overall population. Meanwhile, Albumin was only associated with fasting glucose in SAs, to a greater extent than in the overall population (1 unit increase associated with a decrease of 0.86 of fasting glucose, SE = 0.35). In addition to these associations, increases in histidine, apolipoprotein A1, HDL cholesterol and HDL2 cholesterol were all associated with a decrease in fasting glucose. No additional associations were identified in SAs.

#### 4.5.9.2 2-hour post glucose

No metabolite value was associated with 2-hour post glucose level in the overall cohort **(Appendix Table B.7)**. Following stratification by ethnicity, DHA was positively associated with 2-hour post glucose (1 mmol/L increase in DHA was associated with a 0.2 mmol/L increase in fasting glucose (SE 0.01) in WEs). In SAs, Albumin was negatively associated with 2-hour post glucose (1 unit increase in Albumin was associated with 1.89 mmol/L decrease in 2-hour post glucose SE 0.81). No other metabolite values were associated with 2-hour post glucose. A summary of the identified associations can be seen in **Table 4.3**.

**Table 4.3: Associations between metabolites and postprandial glucose.**

| Postprandial glucose measure (mmol/L) | Overall (n=5,538) | South Asian (n=2,671) | White European (n=2,267) |
|---|---|---|---|
| **Fasting glucose** | Albumin (-)<br>Lactate (-)<br>LDL_D (+) | Albumin (-) | Lactate (-)<br>Histidine (-)<br>Apolipoprotein A1 (-)<br>HDL-C (-)<br>HDL2-C (-)<br>LDL_D (+) |
| **2-hour post glucose** | | Albumin (-) | DHA (+) |

Metabolites associated (P<0.05) with measures of fasting glucose or 2-hour post oral glucose tolerance test (OGTT) in multivariable linear regression in the overall population or in ethnic-specific analyses are presented. All models were adjusted for maternal age, gestational age, parity, BMI (continuous), and smoking status during pregnancy. For ease of interpretation, the direction of associations is presented in brackets, i.e., positive (+) or negative (-). DHA: docosahexaenoic acid. HDL-C: High-density lipoprotein cholesterol. HDL2-C: High-density lipoprotein-2 cholesterol. LDL_D: Mean diameter of low-density lipoprotein (nm).

## 4.6 Discussion

Using a cohort with an equal proportion of WE and SA women, ethnic-specific metabolite signatures of GDM were identified. Of the 146 metabolite values tested, 7 were important for stratifying GDM and non-GDM across both ethnic groups: lactate, VLDL_D, total FAs, total MUFAs, LA, total SFA and esterified cholesterol. Following stratification by ethnicity, 7 metabolite values were found to characterise GDM in both WE and SA women, 6 of which were identified in the collective analysis of the overall cohort. 4 metabolite values were identified to be characteristic of GDM status in both ethnicities and were significantly more predictive of GDM in WE women (total FAs, total MUFA, total SFA and VLDL_D) following a MW test. In addition, 6 metabolite values (alanine, glutamine, total serum cholesterol, n-6 FAs, PUFAs, and citrate) were found to be characteristic of GDM only in WEs. No metabolite value were identified as being uniquely predictive of GDM status in SA women.

Previous work, utilising the Omega cohort (78.5% non-Hispanic white; nested case-control; 46 cases, 47 controls), demonstrated a distinct metabolic profile during early pregnancy that was associated with subsequent diagnosis of GDM (Enquobahrie et al., 2015). The pattern, identified consisted of fatty acids, sugars, alcohols, amino acids and organic acids. This pattern, determined by penalized logistical regression models, shares many common features with a previous univariate analysis (Taylor et al., 2019) and the present multivariate analysis in BiB — namely, amino acids, glycolysis related metabolites, and FAs. Furthermore, this evidence agrees with a recent evidence review of evidence (2021) of GDM metabolic biomarkers which found amino acids, lipids, carbohydrates and purines to be the most commonly identified metabolite classes associated with GDM (Alesi et al., 2021).

Previous molecular analyses have also found evidence to suggest that fatty acids alter insulin resistance and insulin secretion during pregnancy, which could in turn contribute to GDM risk (Villafan-Bernal et al., 2019; Chen, X. et al., 2019). This association between fatty acids and dysglycemia has been previously observed in Asian populations, with PUFA-derived eicosanoids discriminating between type-2 diabetics and controls with good accuracy ($R^2X = 0.824$, $R^2Y = 0.995$, $Q^2 = 0.779$) in a Chinese population (Xia et al., 2020). Regarding SAs, earlier work by Taylor *et al* (Taylor et al., 2019) that utilised univariate statistical techniques in the BiB cohort identified some evidence of ethnic-specific associations between fatty acids and GDM and agrees with molecular analyses that demonstrate that fatty acids alter insulin resistance and insulin secretion during pregnancy (Villafan-Bernal et al., 2019; Chen, X. et al., 2019). It is also known that differences in PUFA metabolism exist between ethnicities meaning that it is plausible for fatty acid metabolism to have differential impacts on GDM risk in different ethnic groups (Gray et al., 2013; Benedetti et al., 2019; Ralston et al., 2013).

Through the multivariate analysis of the BiB cohort, this study identified fatty acids to be the most represented class of metabolites to be deemed as important (VIP≥1) in characterising GDM status in the overall BiB cohort, agreeing with previous work that has highlighted the importance of the fatty acid class in modulating pregnancy dysglycemia. Furthermore, in the separate

analysis of each ethnic stratum, fatty acids were commonly identified as important metabolites in WEs and SAs, with 75% and 50% of included fatty acids measures being 'important' to characterise GDM in WE women, while in SA women fatty acids constituted more than half of all important metabolites. Evidence of ethnic-specific associations between fatty acid metabolites and GDM were also discovered, with a larger number of fatty acids metabolites being classified as important in WEs, and total fatty acids, total MUFA, total SFA being significantly more important in determining GDM status in WEs. Furthermore, of the numerous fatty acid measures that were associated with GDM, only DHA was associated with an increase in 2-hour post glucose levels in WE women.

Overall, DHA is considered a protective metabolite against insulin resistance (HOMA-IR); however, recent evidence suggests high heterogeneity in the direction and magnitude of effect (Chen, X. et al., 2019; Zhu et al., 2019; Brown et al., 2019). Similarly, researchers investigating the Camden pregnancy cohort (n=1,368) reported a significant positive linear association between DHA and HOMA-IR (0.303 ± 0.152 per unit DHA %; p value<0.05) (Chen, X. et al., 2019), while, conversely, the DOMINO trial (n=1990 pregnant women) reported no difference in 1-hr post-OGTT glucose levels between DHA supplemented mothers and controls (Zhou et al., 2012). The reason for such discrepancies is unclear but may be because n-3 PUFAs (such as DHA) require interactions with other metabolites (e.g., Vitamin D) (Jamilian et al., 2017) to impart an effect, levels of which vary considerably between populations, seasons, and geographic region (Darling et al., 2013; Sedhain et al., 2020; Lagunova et al., 2009).

Outside of the fatty acid class, analine, glutamine and citrate were all identified to be important in the characterisation of GDM status in WEs. Analine, glutamine, and citrate are biologically connected and could moderate dysglycemia through their interaction with the tricarboxylic acid cycle (TCA) to promote the formation of TCA intermediates, fatty acid synthesis, and modulate glucagon and insulin secretion (Newsholme et al., 2005; Haber et al., 2006). Taken together, it may be that alanine and glutamine are more robust markers of dysglycemia, while citrate is a marker of metabolic or physiologic stress, such as pregnancy, in diabetic individuals — such as

pregnancy. The association between metabolites that interact with the TCA cycle and dysglycemia has also been reported elsewhere: a small case-control study (26 T2Ds vs 7 controls) that found alanine, glutamine, and citrate to characterize T2D, with citrate being a key marker of diabetics with underlying complications (e.g., CVD) (Del Coco et al., 2019), and in a cohort study of 431 pregnant Chinese women (12-16 weeks' gestation), where alanine and glutamine associated with GDM (Jiang et al., 2020).

An additional metabolite value that has previously been associated with dysglycaemia is the mean diameter of VLDL particles, with a recent hypothesis linking increased VLDL diameter to increased insulin resistance and triglyceride synthesis (Zhao et al., 2019; Krauss, 2004). In this Chapter, VLDL diameter was found to be predictive of GDM in the overall population (VIP = 1.30) and in both ethnic strata (albeit a significantly stronger predictor in WEs), agreeing with the previous findings linking VLDL diameter to increased dysglycemia.

In addition to VLDL diameter, lactate has also been proposed as a regulator of insulin resistance and metabolic syndrome severity (Wu et al., 2016; Jones et al., 2019) and identified to be associated with GDM in a small Chinese case-control study (n=12 GDM; n=10 controls) (Liu et al., 2016). Here, lactate was identified to be characteristic of GDM status in PLSDA models, where it was one of the strongest predictors of GDM in both ethnicities. Furthermore, lactate was also found to be negatively associated with fasting glucose in WEs but not in SAs via linear regression models. The multi-ethnic HAPO cohort demonstrated a similar ethnic-specific association between lactate and fasting glucose in individuals of Northern European ancestry but not minority ethnic groups (Enquobahrie et al., 2015; Jacob et al., 2017; Chen et al., 1993).

Additional metabolites identified as associated with postprandial glucose measures in this study were HDLC and HDL2C, both cholesterol metabolites. Despite these associations, the majority of cholesterol metabolite values in this study were not found to be predictive of GDM, with only total cholesterol identified as an important predictor (VIP≥1) of GDM and only in WEs. Despite this, remnant cholesterol was identified to be an important metabolite (VIP≥1) in the characterisation of ethnicity only in non-cases,

indicating more similar levels of remnant cholesterol in cases compared to non-cases between the WEs and SAs. When also considering that esterified cholesterol was identified to be important in the overall cohort following adjustment by ethnicity, this could indicate a shared role for cholesterol metabolism in the development of GDM between the ethnicities. The lack of consistent associations between cholesterol metabolites and GDM in this study aligns with the current evidence base which suggests that total cholesterol is not convincingly associated with dysglycemia (a meta-analysis of 73 observational studies found no association) (Ryckman et al., 2015), suggesting that associations between total cholesterol and GDM are complex and/or subject to confounding. Future studies should be conducted that aim to minimise this confounding to better examine the relationship between cholesterol metabolites and GDM development.

Overall, metabolite values identified in both ethnicities points to the importance of fatty acid metabolism in characterising GDM. Related pathways of amino acids glycolysis, cholesterol metabolism were also identified to be important classes of metabolites in WEs in regards to characterising GDM. All metabolites were found to be more important in determining GDM in WEs, fatty acids were the most frequent class of metabolite identified in both ethnicities highlighting their importance in characterising GDM.

This study has several strengths. It is the first study to perform multivariate statistical analyses to characterise GDM in a multi-ethnic population through the utilisation of a range of statistical techniques. Nonetheless, this study also has several limitations. These results may not be generalisable across other ethnic groups or geographic regions, including other regions of South Asia. Samples were also taken at a single time point before 28 weeks gestation; therefore (i) we were unable to account for differences in fasting duration and diurnal variation; and (ii) our results are not generalisable across the full-term of pregnancy. In addition, although samples were taken at the time of GDM diagnoses and all before the 28$^{th}$ week (i.e., before or at the time of a typical GDM diagnosis) it is not possible to know if the metabolomic profiles identified in this study were as a result of the diagnosis itself, meaning that reverse causality cannot be ruled out. Beneficially however, no included sample was taken after the 28 week and

variation in time of was minimal meaning variation in time of sample collection did not need to be accounted for in the analysis

Secondly, as with all observational studies, the effect of confounding cannot be disregarded, and causality cannot be inferred. Although confounding cannot be eliminated, all models were adjusted for known GDM confounders present in the BiB dataset (i.e, age, BMI, parity, multiple pregnancy and smoking status). However, other factors that may influence GDM development, including dietary information, hormonal measures or information on polycystic ovary syndrome status could not be including within the models as these variables were not collected by BiB.

Despite this, this is the first study to use a panel of multivariate statistical techniques to characterise GDM in a large prospective cohort with an equal representation of WE women and women from a minority ethnic population from a single geographical area. Additionally, the biological validity of the identified metabolites was tested, with many metabolites being found to be correlated with postprandial glucose measures. In additions, models characterising the overall metabolic differences between ethnicities were also performed to test the presence of reverse associations.

The findings from this study contribute to a greater understanding of the metabolites (and biological processes) that characterise GDM in SAs and WEs. Aetiologically, this study has shown the importance of fatty acid metabolism in GDM development in both WEs and SAs and highlights the need of future research to explore pathways related to fatty acid metabolism (including glycolysis related pathways) in order to gain a better understanding of the biological causes of GDM. Regarding public health, this study illustrates that metabolites commonly associated with diet (including fatty acid and cholesterol metabolites) may be characteristic of GDM. Future work should aim to explore these associations further and confirm these findings in additional cohorts to help inform ethnic-specific GDM prevention strategies.

In conclusion, a range of metabolite values characteristic of GDM status have been identified, including fatty acids, glycolysis related metabolites and measures of lipoprotein size. Evidence of ethnically distinct metabolic profiles in relation to GDM have also been identified.

## 4.7 Summary

- Numerous metabolites were found to be associated GDM in WEs, while no ethnic-specific associations were found in SA.
- Numerous metabolites that characterised GDM in WEs were associated with postprandial glucose measures, while in SAs only albumin was associated with postprandial glucose measures.
- Metabolites appear to be important for characterising GDM in WEs in general while in SAs, metabolites appear most strongly associated with GDM in non-overweight/obese women.
- Differences in metabolic profiles characteristic of GDM in WEs and SAs may be due to underlining genetic differences between the populations that influence how metabolism impacts GDM development.

# Chapter 5: Metabolites and Postprandial Glucose Measures During Pregnancy: a Mendelian Randomisation Analysis

## 5.1 Abstract

South Asians (SAs) are at three times greater risk of GDM compared to white European (WEs), which may be partially attributable to differences in metabolism. To assess the presence of ethnic-specific causal associations between metabolites and pregnancy dysglycemia one-sample Mendelian Randomisation (MR) was performed utilising genetic and metabolomic data for 146 metabolites from 3,668 SA and 3,354 WE women from the BiB cohort.

Genome wide association analysis (GWAS) was performed for all metabolites in both ethnicities. For each metabolite, identified SNPs (p value $\leq 1 \times 10^{-5}$) were included within MR analyses against log-Normalised fasting glucose and a 2-hour post oral glucose tolerance test (OGTT) levels.

Robust genetic instruments were identified for most metabolites (93% instruments of F-statistic ≥10). Fourteen metabolites were associated with postprandial glucose measures in WEs and 11 in SAs. In WEs, HDL cholesterols were the most prominent metabolite class identified, although the direction of the identified associations were mixed. In addition, a 1 mmol/L increase in leucine was also found to decrease fasting glucose and 2-hour post glucose by 0.19 and 0.44 mmol/L respectively in WEs. In SAs fatty acids were the most commonly identified metabolite class, with an increase of 1 mmol/L of FAw6 and 1 mmol/L of linoleic acid (mmol/L) both resulting in an increase of 2-hour post glucose levels by 0.4 mmol/L. No metabolite values were associated with pregnancy dysglycemia in both ethnic groups. Leave-one out analyses highlighted limited bias in the identified associations.

A range of metabolites were found to be associated with dysglycemia during pregnancy, with cholesterols and fatty acids being the most important classes in WEs and SAs, respectively. Metabolites associated with postprandial glucose in SAs were fewer in number and distinct from those identified in WEs, suggesting that metabolism may contribute differently to GDM in SAs and WEs.

## 5.2 Background

Current evidence, including that from earlier work (Chapter 4) demonstrated that the metabolic profiles of GDM and non-GDM women differ and suggests that some of this discrepancy may be driven by ethnicity (Taylor et al., 2019). Collectively this confirms that metabolism contributes to GDM development and also suggests that ethnic disparities in GDM risk may be partly explained by differences in metabolism between pregnant SA and WE women. This implies that underlying differences in metabolism may contribute to the disparity in GDM risk that is observed between ethnicities.

Unfortunately, metabolite profiles and analyses of GDM from previous studies are largely heterogeneous as a result of varying quantification methods, sample types, and GDM diagnostic criteria (de Souza et al., 2020; Wang et al., 2021). This makes it difficult make comparisons between cohorts and to clearly evaluate the role of individual metabolites on GDM risk as it's possible that metabolites have temporal variations in their associations with GDM and that some metabolites are only associated when assessed via certain quantification techniques and in certain sample types. Furthermore, differences in the classification of metabolites (i.e., using cumulative measures to measure associations with types of metabolites rather than using specific metabolite measures) also makes it difficult to compare results across cohorts.

Additionally, as well as genetic and heritable influences on metabolism, the metabolome is influenced by a range of environmental factors (including diet, physical activity, medication, and environmental pollutants) which all need to be accounted for in an observational setting to accurately compare results between studies (Zulyniak and Mutch, 2011). Collectively, this makes it difficult to confidently infer the causal role of inherited metabolic traits on GDM and their contribution to GDM disparity between ethnic groups in an observational setting. However, if the role of individual dietary related metabolites on GDM were to be accurately ascertained in an ethnic-specific manner it may be possible to develop ethnic-specific dietary guidelines that would be more effective in preventing GDM. Furthermore, if causative associations were identified between metabolite levels during pregnancy and GDM, it may be possible to develop dietary interventions aimed at favourably

moderating the levels of these metabolites to reduce GDM risk or dysglyceamia. Indeed, current evidence suggests that dietary strategies are not equally effective across ethnic groups at reducing GDM risk (Chapter 2) (Griffith et al., 2020). Finally, a better understanding of how metabolites impact dysglycemia during pregnancy will provide a clearer understanding of the pathology of the disease which is not fully understood (Chen et al., 2018; Schaefer-Graf et al., 2018).

One epidemiological method that can allow for the causal role of exposures on outcomes to be inferred with limited confounding is Mendelian Randomisation (MR). MR is a tool that uses a set of genetic variants that are associated with an exposure as instrumental variables (IVs) that acts as a proxy for the exposure.( **Figure 5.1)**. This is possible because genetic variants are randomly assigned during fertilisation and, therefore, the association between these variants and the outcome should be largely free of confounding (Sheehan et al., 2011; Liu et al., 2017; Carreras-Torres et al., 2017). This can be inferred if the 3 core assumptions of MR are satisfied: that (i) genetic variants are associated with the exposure (ii) while not being associated with confounding factors and (iii) genetic variants they must influence the outcome only via the exposure and not through an alternative pathway (i.e., horizontal pleiotropy) (Sheehan et al., 2011; Liu et al., 2017; Carreras-Torres et al., 2017). In practice, the only MR assumption that can be empirically tested is the first assumption. The second and third assumptions have to be evaluated through the implementation of sensitivity analyses, depending on the MR approach utilised. For one-sample MR, variants are typically dropped successively in order to assess the impact of individual variants on the outcome to assess whether any appear to have a more substantial impact on the outcome, a potential indicator of pleiotropy. A wider range of sensitivity analyses can be implemented in a two-sample setting, however in practice the choice utilising a one or two sample approach is based upon the data available, with a one-sample approach requiring individual level data whereas a two-sample approach requiring summary level data,

Despite the rising popularity of MR, no MR study has been conducted on post-prandial glucose measures during pregnancy (Diemer et al., 2021). Through the conduction of a MR analysis examining the association between

metabolite levels during pregnancy and measures of pregnancy dysglycemia it may be possible to better understand how metabolism impacts GDM development due to the ability of MR to identify causal associations. This chapter aims to determine whether there are ethnic-specific causal associations between metabolites and continuous glucose measures assessed at the 28th week of pregnancy that may contribute to GDM.

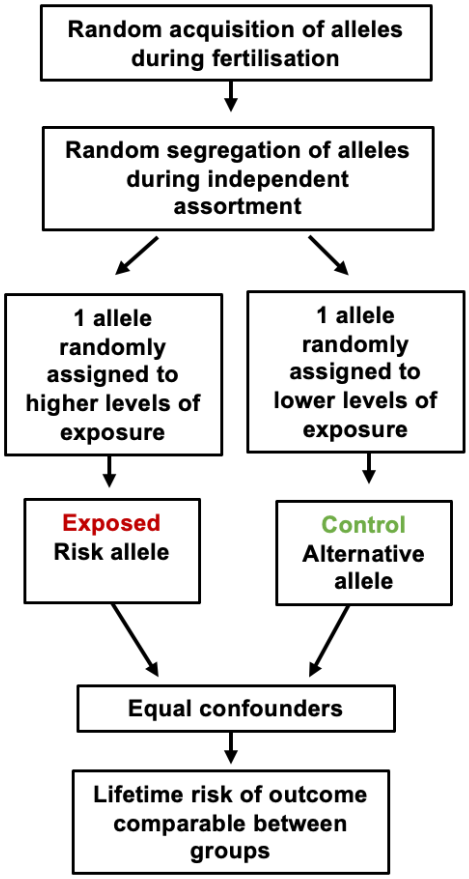



**Figure 5.1 Schematic highlighting how MR mimics an RCT.**

## 5.3 Hypotheses

i. Due to the metabolic dysregulation that characterises GDM, the same metabolite value may be causally associated with postprandial glucose measures in both ethnicities. However, additional ethnic-specific causal associations between metabolites and postprandial glucose measures may also exist which could explain some of the disparities in GDM risk.

ii. Some metabolite levels in WE and SA populations are associated with different genetic variants. This could either be due to differences selection as a result of historically varying diets that results in different allele

frequencies in both populations, or due to different gene-environment interactions in SAs compared to WEs as a result of varied lifestyle.

iii. Metabolites on common biological pathways are associated with similar genetic signals in both ethnicities.

## 5.4 Methods

### 5.4.1 Exposure data

Metabolite data from 11,480 fasting serum samples were obtained from BiB as described in Chapter 4 of this thesis. In brief, a metabolite panel from Nightingale Health © were quantified by high throughput NMR in a previously validated workflow (Taylor et al., 2019). Redundancy was minimised in the panel through the removal of metabolite values expressed as a percentage or ratio resulting in 146 metabolites being retained for analysis. This panel of 146 metabolites included metabolites from a range of biochemical classes, including lipoproteins, fatty acids, cholesterols, glycerides and phospholipids, triglycerides, amino acids, apolipoproteins, ketone bodes, glycolysis-related metabolites, measures of fluid balance and inflammation and measures of the mean diameter of lipoproteins (**Appendix Table B.1**). Samples failing quality control as defined by Nightingale Health © were removed from the dataset and missing data were imputed using multiple imputation.

### 5.4.2 Outcome data

In BiB, GDM was defined with a 75g OGTT at 26-28weeks based upon modified WHO diagnostic criteria utilising 2 postprandial glucose measures (fasting glucose and 2-hour post glucose). Using these criteria, individuals were diagnosed with GDM if either their fasting glucose concentration exceed 6.1 mmol/L or if 2-hour post-load glucose concentrations exceeded7.8 mmol/L. The OGTT was completed in the morning following an overnight fast and involved the consumption of a standard 75g of anhydrous glucose solution over a 5-minute period (Lawlor et al., 2014). To maximise analytical power, MR analyses were performed to detect causal associations between metabolite values and fasting glucose and 2-hour post glucose rather than

GDM. Fasting glucose and 2-hour post glucose values were log Normalised prior to analysis.

### 5.4.3 Participant data

Exposure and outcome data were combined with covariate data from BiB including ethnicity (self-reported), ancestry information (i.e., the country of birth of the mothers' parents, paternal grandparents and maternal grandparents), parity, age and BMI. Mothers with a history of T2D were excluded from analysis, as were those with fasting serum samples taken after the 28$^{th}$ week.

### 5.4.4 Genetic data

Imputed genetic data was obtained from BiB. BiB 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) (Arciero et al., 2021). Genetic data from the Illumina Global Sequencing Array (GSA) and Illumina CoreExome SNPs were combined and SNPs missing in >5% of individuals were excluded (Arciero et al., 2021). When evaluating imputed data the $R^2$ value can be a measure of quality control as it refers to the proportion of genetic variation maintained in the imputed data. As a result SNPs with an $R^2$ <0.9 were excluded prior to analysis.

### 5.4.5 GWAS

Conventionally a GWAS assumes individuals are unrelated; the inclusion of related individuals can potentially lead to spurious associations (Marees et al., 2018; Uffelmann et al., 2021). However, the removal of related individuals from the BiB sample would substantially reduce the sample size. In addition, high rates of consanguinity also exist in the SA strata of the cohort which can make relatedness difficult to assess when self-reported (Sheridan et al., 2013). As such, a GWAS mixed linear model association (MLMA) analysis was conducted that allowed for the inclusion of related individuals. 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. The inclusion of this random effects term avoids the inflation of the test statistic from related individuals, thereby maintaining the correct type 1 error rate for associations between the SNP and the outcome. (Yang et al., 2014; Uffelmann et al., 2021). GWAS MLMA models were implemented using the *GCTA* (Yang et al., 2011) on ARC4, part of the University of Leeds' High Performance Computing facility. MLMA GWAS modelling was conducted for each metabolite in both ethnicities (Yang et al., 2011). To increase power, MLMA-loco (leave-one-out) analysis was utilised, preventing a SNP from being included in both the fixed and random effects concurrently, thereby double fitting the SNP's effect (Yang et al., 2014).

Population stratification in a sample of genetic data can result in spurious associations as a result of variations in allele frequencies between sub-populations in the genetic data which should be accounted for in a GWAS (Uffelmann et al., 2021). To account for this potential stratification, PCA can be conducted on genetic data which will result in the grouping of sub-populations that can then be adjusted for in a GWAS model by including Principal Components (PCs) as covariates in the GWAS. Steps taken to perform PCA analyses are highlighted in the appendix **(Appendix Figures C.1, C.2)**. In brief, SNPs missing in ≥ 10% of individuals were removed along with SNPs in known regions of strong LD, such as regions with large inversions that restrict recombination, and rare SNPs (minor allele frequency (MAF) ≤ 0.005) **(Appendix Table C.1)**. Remaining SNPs were pruned by LD ($R^2$ > 0.3) and non-autosomal SNPs were removed. Following PCA, clear stratification was seen across PC1 and PC2 meaning GWAS models were included as covariates in the model. Alongside PC1 and PC2 to account for population stratification, MLMA models were also adjusted for parity. Those covariates to be adjusted for were selected based upon covariates adjusted for in previously published GDM GWAS studies, while also ensuring the maximum sample size; adjusting for parity alongside excluding mothers with a history of T2D (as mentioned above) resulted in a total of loss of 7.1% and 6.6% of SA and WE women respectively (Kwak et al., 2012; Wu et al., 2021). Further adjusting for maternal age would have resulted in a loss of 10.1% and

9.6% of SA and WEs respectively while adjusting for smoking would have resulted in a loss of 19.2% of SA women and 13.7% of WE women for BMI 22.1% of SA and 16.6% of WEs. Gestational age showed little variation so was not adjusted for (median gestational age SA = 184 days, IQR= 182-186.7, median gestational age WE = 184 days, IQR= 182-187). After combining with postprandial glucose data, 3,693 SA and 3,377 individuals whose samples were taken before the 28$^{th}$ week of pregnancy were retained before outlier removal.

Outliers were removed (those outside of 1.5 x IQR) were removed for each metabolite in each ethnicity separately and metabolite values were Normalised by taking the log, square root or NST as appropriate following the visual inspection of histograms and QQ plots. The final sample size used for all metabolite values following the removal of outliers can be found in the appendices. **(Appendix Tables C.2, C.3)**.

### 5.4.6 Assessment of genomic inflation

To minimise false-positives and evaluate residual population stratification, genomic inflation factors (λ) were calculated, with a $\lambda \geq 1.1$ considered indicative of genomic inflation (van den Berg et al., 2019; Rivadeneira and Uitterlinden, 2021). λ was calculated for all models after the utilisation minor allele frequency (MAF) cut offs of 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. A MAF cut-off of <0.05 was found to reduce λ to ~1 meaning this cut off was utilised in the analysis **(Appendix Table C.4, Figure C.3)**.

### 5.4.7 Meta-analysis of GWAS results

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 in METAL, a command-line tool that meta-analyses large scale GWAS data by weighting each estimated effect size (β) value by the reciprocal of its SE (Willer et al., 2010). This allowed for the heterogeneity of the identified associations in SAs and WEs to be determined as well as an overall effect size for each SNP in a larger multi-ethnic sample. Similarities in the

GWAS results between ethnicities were visualised using the *PheGWAS* package in Rstudio (version 4.0.2) (George et al., 2020; R Development Core Team, 2020).

### 5.4.8 Genetic Instruments

One-Sample MR was conducted for all 146 metabolite values in both ethnic populations using SNPs identified as significant beyond a genome-wide suggestive level (p value ≤ 1 x $10^{-5}$). Metabolites were grouped into their overall classes and SNPs in each class were thinned by LD ($R^2$<0.2) via the NIH LDlink online tool, reducing the overlap of instruments in each class (Machiela and Chanock, 2015; Myers et al., 2020). Following clumping by LD, the most significant SNP was selected from each LD block. For individuals of white European (WE) ancestry, all European (EUR) populations (Utah Residents from North and West Europe, the Toscani in Italia, Finnish in Finland, British in England and Scotland and Iberian in Spain) populations were used in LDlink to estimate LD all 1000G SA populations (Gujarati Indian from Houston Texas, Punjabi from Lahore Pakistan, Bengali from Bangladesh, Sri Lankan Tamil from the UK and Indian Telugu from the UK) were selected to estimate LD in BiB SAs. The decision to use all SA populations was due to the expected similarity in their LD structure allowing for an increase sample size and resultant improvement in the accuracy of LD estimates (Genomes Project Consortium, 2015).

As LDlink utilised 1000 Genome (1000G) data to determine LD estimates and SAs in BiB predominantly originated from a distinct region of Pakistan (Mirpur) the similarity between 1000 Genome and BiB SA samples was evaluated by performing PCA on genetic variants found in both datasets to assess whether LD structure was likely to be similar in the two datasets. 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 (Arciero et al., 2021). The same methodology utilised in obtaining the PC1 and PC2 covariates were

utilised along with an additional SNP of checking allele pairings in BiB and 1000G to check that they matched.

### 5.4.9 MR analyses

Genetic Risk Scores (GRS) summate the effect of multiple SNPs into a single score which can then be used as an instrument variable to assess the impact of a set of genetic variants on an outcome (Burgess and Thompson, 2013). Weighted GRSs 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 (Chang et al., 2015). Scores for each individual are then summated across all included SNPs based on the number of risk alleles an individual has for each SNP (0,1,2), with this number being multiplied by the weight of each SNP.

One-sample MR was then performed by Two-Stage Least Squares (TSLS) regression to obtain a causal estimate for the effect of each metabolite value on the log-Normalised continuous measures of dysglycemia following a 75g OGTT, fasting glucose, and 2-hour post glucose (Budu-Aggrey et al., 2019). Here, the level of a metabolite is regressed on its respective GRS score and, subsequently, the outcome is regressed onto these fitted values in the second stage. TSLS was performed using the *ivpack* and *AER* packages in R version 4.0.2. All MR results have been reported according to STROBE-MR guidelines (Skrivankova et al., 2021).

For significant associations leave-one-out analysis was performed; SNPs were removed in turn from the instrument and the impact on the effect estimate and F statistic was assessed. If the exclusion of a SNP was found to drastically 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 2$^{nd}$ or 3$^{rd}$ MR assumption. To further test for violations of these assumptions, SNPs included in significant instruments were searched for in both the Phenoscanner and GWAS Catalog databases in case of any previously identified associations (Buniello et al., 2019; Staley et al., 2016; Kamat et al., 2019). Specifically, in Phenoscanner the ‘disease and traits’ database was searched, and SNPs were recorded to be associated with a

different outcome if the p value for this association was below 1 x $10^{-5}$ (the suggestive level utilised in GWAS). 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 (Davies, N.M. et al., 2018). Deviations in these two measures can indicate confounding in the OLS estimate, violations of the MR assumptions or could be due to the fact that MR represents a lifelong estimate for the exposure-outcome association.

### 5.4.10 *Post-hoc* power analyses

*Post-hoc* power analyses were performed via the mRnd CNS genomics tool (https://shiny.cnsgenomics.com/mRnd/) in both ethnicities for all metabolites identified as associated with either postprandial glucose measure in only one ethnic group to assess whether the absence of an association in the alternate ethnicity was due to limited power (Brion et al., 2012). This tool estimates power in a given size by utilising estimates of the proportion of exposure variance explained by the instrument and the true causal exposure-outcome association by calculating the asymptotic mean of the instrumental variable along with its variance (Brion et al., 2012)

Observational and 'true' associations required by the tool were initially 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 as advised by the tool. Due to the *post-hoc* nature of this analysis, additional power analyses could be conducted assuming the MR to be the true causal effect. This analysis was performed in the non-significant population for each metabolite associated in only one ethnicity. If power was found to be significant at the 5% level ($\alpha = 0.05$) level power was also assessed at the 1% level ($\alpha = 0.01$).

## 5.5 Results

### 5.5.1 GWAS results

In SAs genome-wide significant SNPs were identified for two metabolite values: tyrosine (15 SNPs) and acetate (1 SNP, rs10945476). No previous associations have been identified for SNP rs10945476, with this study being the first to report on its association with acetate. All 15 SNPs found to be associated with tyrosine are located on the p arm of Chromosome 17 and were in LD with one another (all $R^2 \geq 0.38$. 3/15 of the SNPs are found in the gene *SLC13A2* involved in the transmembrane transport, 10/15 SNPs are found in the transcription factor gene, *FOXN1* (Ensembl, 2021b; Ensembl, 2021a). To date, neither gene has been associated with tyrosine levels. The remaining two SNPs are currently not in any known gene. No previous associations have been identified for SNP rs10945476, with this study being the first to report on its association with acetate.

Following meta-analysis of effect estimates across ethnicities, genome-wide significant associations were observed for 4 SNPs in relation to the alanine, even though these SNPs were only initially identified to be associated with alanine in the SA population at the genome wide suggestive level. These SNPs (rs12256633, rs17121228, rs7096521, rs12240368) are all found on chromosome 10 in the receptor gene *SORCS1*, and are in strong LD with each other ($R^2=1$) and have not been associated with alanine levels (Howe et al., 2020; Ensembl, 2021c). For 90 metabolite values, no associations were found to exceed the genome-wide suggestive level following meta-analysis between ethnicities, a further indication of distinct signals in both ethnicities. SNPs were found to be associated at the suggestive level despite the differing direction of effects in both SAs and WEs for four metabolite measures: concentration of XL-HDL, total lipids in M-VLDL, mean density of VLDL and citrate.

### 5.5.2 Genetic instruments

To ensure that SNPs included in each genetic instrument are independent it is necessary to thin SNPS by LD prior to performing MR analyses. The online LD link tool provides LD estimates in distinct population

groups, however this tool utilised 1000G samples to obtain LD estimates and 1000G SA samples are from a different SA populations (Gujarati Indian from Houston Texas, Punjabi from Lahore Pakistan, Bengali from Bangladesh, Sri Lankan Tamil from the UK and Indian Telugu from the UK) than BiB SA samples. In order to assess whether these SA samples appear similar to BiB SA samples PCA was utilised to compare the genomic data from BiB and 1000G.

PCA analysis of WE and SA BiB samples and SA samples from 1000G showed some separation between BiB SA samples and SA 1000G samples, with PC1 explaining 14.3% of the variation and PC2 explaining 7.9%. However, when performing PCA analyses on all ancestries from 1000G BiB SAs were found to be in close proximity to SA samples The greatest separation of the overall SA population from all other groups was seen when plotting PC2 against PC3. The BIB SA population was found to overlap with the 1000G SA populations from Lahore, Pakistan and Gujarati Indians. This is not surprising given that these two 1000G populations are geographically closer to Mirpur, Pakistan, where the BiB population predominantly originates from. **(Appendix Figures C.4, C.5, C.6)**. This same overlap was observed when restricting the analysis to BiB SAs and 1000G SAs indicating that BiB and 1000G SA samples were similar enough to allow for the utilisation of 1000G data and the LD link tool in order to assess LD structure in the SA stratum of BiB.

The number of SNPs remaining in each instrument following thinning by LD can be seen in the appendix **(Appendix Table C.5, Figures C.6, C.7).** 2.7% and 11.6% of genetic instruments in WEs and SA respectively had an F statistic ≥ 10, indicating the majority of instruments were not subjected to weak instrument bias. The average F statistic for WEs instruments was 72.4%, while in SAs it was considerably lower at 26.7%.

### 5.5.3 MR analysis: White Europeans

Two metabolite values, leucine and mean density of HDL (HDL_D) lipoproteins were found to be associated with both fasting glucose and 2-hour post glucose. A 1mmol/L increase in leucine was associated with a decrease of 0.193 mmol/L of fasting glucose and a 0.443 mmol/L decrease in 2-hour

post glucose while a 1nm increase in the mean diameter of HDL was associated with a 0.082 mmol/L decrease in fasting glucose and a 0.191 mmol/L decrease in 2-hour post glucose.

An additional two metabolite values, an increase in 1mmol/L total cholesterol in M-HDL (M-HDL-C) and cholesterol esters in M-HDL (M-HDL-CE) were also found to be associated with a decrease in fasting glucose measures 0.189 mmol/L and -0.327 mmol/L respectively meaning that a total of 4 metabolite values were associated with fasting glucose in WEs **(Table 5.1)**. In addition to Leucine and HDL_D, 9 additional metabolites were also identified as associated with 2-hour post glucose. Eight of these metabolite values were positively associated (HDLC, HDL2C, HDL3C, triglycerides in XSVLDL, 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. Cholesterol metabolites, measures of total cholesterols in lipoproteins and total cholesterols in lipoproteins were the most common types of metabolite to be associated with postprandial glucose in WEs, with leucine being the only metabolite identified in WEs which did not belong to a cholesterol or lipoprotein related metabolite class. All cholesterol metabolites associated with 2-hour post glucose resulted in an increase in glucose while all cholesterol metabolites associated with fasting glucose resulted in a decrease in glucose. Wu-Hausman p values $< 0.05$ indicate deviations of the instrumental variable estimate from the OLS estimate **(Table 5.1)**.

### 5.5.3.1 White European: Sensitivity analyses

All leave-out one analyses maintained significance ($P \leq 0.05$) indicating that no individual SNP was driving the identified associations for 6 of the identified metabolite values in WEs: leucine, mean diameter of HDL, total lipids in L-HDL, cholesterol esters in S-HDL and cholesterol esters in M-HDL **(Appendix Figure C.8)**. For the remaining metabolites, $\beta$ values were consistent across leave-one-out analyses although not all associations remained significant following successive removal of SNPs. Furthermore, F statistics did not substantially differ through the exclusion of individual SNPs from the instruments of most identified metabolites, suggesting that

instruments were not substantially impacted by an individual SNP, a potential indicator of pleiotropy **(Appendix Figure C.9)**. The exception to this was for the metabolite value of cholesterol esters in M-HDL, where the exclusion of rs2138011 or rs739018 increased the F statistic, suggesting that these SNPs contributed less to the instrument strength than the other SNPs comprising in the instrument.

Three of the metabolites (Leucine, L-HDL_L, L-HDL-C) identified as associated with postprandial glucose in WEs included a SNP in their instrument have previously been associated ($p \leq 1 \times 10^{-5}$) with at least one potential confounder (BMI, hypertension or waist circumference) **(Appendix Table C.6).** The removal of these SNPs from the instrument was not found to impact the significance of the associations identified between Leucine and either glucose measure or L-HDL_L and 2-hour post glucose **(Table 5.2)**. In the L-HDL-C instrument, two SNPS (rs5576825 and rs6811162) were identified as associated with a potential confounder (waist circumference and hypertension respectively). The exclusion of either or both SNPs resulted in non-significant associations between L-HDL-C and 2-hour post glucose. Importantly, for both of these SNPs it is conceivable that the identified confounders could reside on their causal pathway (i.e., vertically pleiotropic) rather than be in horizontal pleiotropy; therefore, these SNPs may not violate the 2$^{nd}$ MR assumption.

**Table 5.1 Significant MR results in White Europeans.**

| | Class | Metabolite | Outcome | F statistic | β estimate (95% CI) | WuH |
|---|---|---|---|---|---|---|
| Reduce 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.

**Table 5.2 Removal of potentially pleiotropic SNPs**

| 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<br>-0.443 (-0.774, -0.113) 2H | 0.004<br>0.024 | -0.203 (-0.339, -0.068) FG<br>-0.547 (-0.909, -0.185) 2H | 0.006<br>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 |

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.

### 5.5.4 MR analysis: South Asians

In SAs, 2 metabolites were associated with fasting glucose (FAw3 and S-HDL-C) and 9 associated with 2-hour post glucose : LA, FAw6, total lipids in M-VLDL (M-VLDL-L), total cholesterols in IDL (IDL-C), cholesterol esters in IDL (IDL-CE) concentration of L-LDL (L-LDL-P), total phospholipids in S-LDL (S-LDL-PL,), total phospholipids in L-HDL (L-HDL-PL), total lipids in small S-HDL (S-HDL-L), total cholesterols in S-HDL (S-HDL-C) **(Table 5.3)**.

Fatty Acids were the class of metabolites most frequently associated with postprandial glucose in SAs. All three fatty acids (LA, FAw3 and FAw6) in the dataset were associated with an increase in either postprandial glucose measure to a similar magnitude – i.e., a 1 mmol/l increase associated with a +0.4 mmol/l increase in fasting glucose (FAw3) or 2-hour post glucose (FAw6 and LA). In addition to fatty acid metabolites, an increase in 1mmol/L of total cholesterols in IDL (IDL-C) and increase in 1mmol/L cholesterol esters in IDL (IDL-CE) in SAs were associated with a decrease of 1.19 mmol/L of and 1.34 mmol/L of 2-hour post glucose respectively.

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 class were found to be associated with postprandial glucose: in SAs S-HDL-L was associated with a decrease in 2-hour post glucose (-1.23 mmol/L) and S-HDL-C was associated with an increase in fasting glucose (+ 1 mmol/L) while in WEs S-HDL-CE was associated with an increase in fasting glucose (+ 1.78 mmol/L). In addition to the S-HDL class, members of the L-HDL class were found to be associated with postprandial glucose measures in both populations. Total lipids in L-HDL and total cholesterol in L-HDL were both associated with a ~0.3 mmol/L increase in 2-hour post glucose in WEs and a 1 mol/L increase in the concentration of L-HDL particles associated with a 200 mmol/L increase in 2-hour post glucose also. In SAs, an increase of 1 mmol/L in phospholipids in L-HDL was associated with an increase of 0.682 mmol/L in 2-hour post glucose.

**Table 5.3 Significant MR results in South Asians.**

| | Class | Metabolite | Outcome | F statistic | β estimate (95% CI) | WuH |
|---|---|---|---|---|---|---|
| Reduce Glucose | L-LDL | LDL_P | 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 |

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). WuH: Wu-Hausman p value.

### 5.5.4.1 South Asians: Sensitivity analyses

Seven instruments in SAs were comprised of a single SNP meaning it was not possible to perform a leave-one-out analysis for these metabolites to assess the impact of pleiotropy on these identified associations. For the remaining 5 metabolites, no metabolite was consistently significant in each leave-one-out analyses. Nonetheless, associations were consistent across each leave-one-out analyses, particularly for the metabolite exposures of total lipids in M-VLDL and phospholipids in L-HDL, indicating that a single SNP was unlikely to be driving the associations **(Appendix Figure C.10)**. Likewise, no large differences in F statistics following the removal of individual SNPs were identified in any instrument **(Appendix Figure C.11)**.

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 **(Appendix Table C.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, 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 effect. 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.

### 5.5.5 *Post-hoc* analysis: Analysis of metabolite classes

Numerous SNPs were found to be associated with more than one metabolite value, particularly for metabolites in the same metabolite class **(Appendix Figures C.12, C.13)**. This is to be expected as metabolomic pathways are intrinsically intertwined, particularly when considering metabolites in the same class. To minimise the risk of violation of the 3rd MR assumption (i.e., the genetic instrument must only influence the outcome via the exposure and not via an alternative biological pathway) (Sheehan et al., 2011) the collective effect of each metabolite class on postprandial glucose measures was examined in each metabolite class through the creation of a composite score from PC1 values obtained from PCA for each metabolite class. Metabolite classes with > 2 metabolites in WEs and SAs respectively had ≥70% of their variation explained by PC1, indicating that PC1 accounted for the majority of the variation in the metabolite class, suggesting that it is a strong composite score for the class. To assess the impact of outliers on PCA results, outliers were defined and removed based on two cut-offs: a standard 1.5 X IQR difference from the median and a less stringent cut off of 3 x IQR. Values for PC1 and PC2 were found to be similar in both ethnicities following the removal of an outlier (defined as 1.5 x IQR) when compared to PC1 and PC2 values following the removal of 'extreme' outliers (defined as 3 x IQR). As a result, only extreme outliers were removed prior to analyses in order to maximise sample size **(Appendix Tables C.7, C.8)**.

4 classes were associated with a glucose measure in WEs: 2 with fasting glucose (M-LDL and all LDLS, i.e., the collective grouping of HDLs, MDLs and SDLs), 1 with 2-hour post glucose (XL-HDLs) and 1 with both measures (S-LDLs) **(Table 5.4).** Fatty acids were the only metabolite class to be associated with either glucose measure in SAs, where they were found to be associated with an increase in 2-hour post glucose levels. Compared to the analysis of individual metabolites, instruments tended to be weaker when analysing each metabolite class collectively **(Appendix Figure C.14, Table C.9)**.

The removal of an individual SNP was not found to greatly impact the F statistic in any instrument. Associations between all LDLs and M-LDL classes and fasting glucose were found to be robust throughout leave-one-out analyses **(Appendix Figure C.15, C.16)**. As with the individual analyses of LA and FAw6, the exclusion of SNP rs12720820 as well as the removal of rs7159441 (found within the *PELI2* gene) (GeneCards Suite., 2022a) resulted in non-significant associations, indicating that SNPs rs12720820 and rs7159441 had the largest impact on postprandial glucose levels. For the XL-HDL exposure, the removal of SNP rs55768285 also resulted in a non-significant association **(Table 5.4, Appendix Table C.10).**

**Table 5.4 Significant MR results from the analysis of metabolite classes**

| Ethnicity | Metabolite class | Outcome | F statistic | β estimate (95% CI) | WUH |
|---|---|---|---|---|---|
| **White European** | XL-HDL | 2-hour post | 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 | 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 | 32.8 | 0.172 (0.018, 0.327) | 0.0184 |

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

### 5.5.6 Power analyses

In order to determine whether the absence of an association in one ethnic stratum when a causal association was identified in the other ethnic stratum was potentially due to an ethnic-specific effect rather than a lack of power in the other ethnicity, *post-hoc* power analyses were performed.

When using β estimates from unadjusted linear regression models as an estimate of the true causal effect, power was low (~5%) and comparable between ethnicities in the majority of analyses, with the exception of S-LDL-P and L-HDL-P in WE, LL-LDL-P in SAs and the overall fatty acid class in SAs which were all adequately powered (≥80% power) at both 0.05 level **(Appendix Table C.11).** When using β estimates from unadjusted linear regression models as an estimate of the true causal effect power was low in the majority of analyses, with the exception of S-LDL-P and L-HDL-P in WE, LL-LDL-P in SAs and the overall fatty acid class in SAs which were all adequately powered (≥80% power) at both 0.05 level. However, 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, despite this association being non-significant, suggesting that the absence of an effect of FAw3 in WEs in these analyses may not be due to inadequate power. Likewise, the analysis of HDL3C in SAs was adequately powered at the 80% level to detect the identified MR association in SAs (power with $SA_{estimate}$ = 0.92) as well as the effect identified in WEs (power with $WE_{estimate}$= 0.71), despite the fact that HDL3C was not found to be associated with postprandial glucose in SAs. The analysis of HDL2C in SAs was also sufficiently powered to detect the observed MR effect in WEs. $R^2$ values were consistently lower in the ethnic group where an effect was not detected, as expected. Despite this, all genetic instruments for the identified metabolite values had an F statistic ≥ 10 in both ethnicities indicating that weak instrument bias was not responsible for the absence of effects identified in the other ethnicity.

## 5.6 Discussion

This analysis has identified ethnically distinct associations between a range of metabolites and postprandial glucose measures during pregnancy in SAs and WEs, although no shared associations were identified. These findings support the idea that metabolite levels are heritable and that individual risk of GDM is somewhat genetically predisposed (Kettunen et al., 2012). Fourteen metabolites were found to be associated with postprandial glucose measures in WEs while a distinct set of 11 metabolites were identified in SAs. In WEs cholesterols and lipoproteins were the most commonly identified class, while in SAs fatty acids were found to have the greatest influence on postprandial glucose levels. Through conducting GWAS analyses of metabolites, novel genome-wide significant associations have also been identified in relation to acetate (1 SNP) and tyrosine (15 SNPs) in SAs. 10/15 newly identified SNPs associated with tyrosine were found in the *FOXN* gene, a transcription factor that has previously been associated with ceramide levels (a lipid metabolite) in a Chinese cohort (Chai et al., 2020). 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 (Gulbins et al., 1997).

### 5.6.1 Identified associations in WEs

#### 5.6.1.1 Leucine

Branched-chain amino acids (BCAAs), including leucine, valine and isoleucine are predominantly metabolised in skeletal muscles (as opposed to other types of amino acids which are predominantly oxidised in the liver) where they are involved in signal transduction in the regulation of protein synthesis and mitochondrial function (Rondanelli et al., 2021). In addition to their roles in skeletal muscles, BCAAs are hormonal signalling regulators and are expected to module insulin resistance (IR). Although mixed results regarding the direction of the association between leucine and Insulin Resistance (IR) have been identified, likely as a result of heterogenous cohorts with differences in sex and obesity status, leucine is suspected to increase insulin secretion (Nie et al., 2018; Leenders and van Loon, 2011).

Leucine is an essential amino acid and, therefore, must be obtained from dietary sources (Rondanelli et al., 2021). Common dietary sources of leucine include meat products and cheese, with smaller amounts of leucine 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 (Rondanelli et al., 2021).

Few studies have looked at the role of leucine in influencing continuous postprandial glucose measures during pregnancy. However, the ratio of Leucine/ Isoleucine has been associated with a decrease in fasting glucose in the HAPO study: a multi-ethnic cohort of pregnant women of Afro-Caribbean, Mexican American, Northern European, and Thai ancestry with serum samples taken at ~28 weeks' gestation (Liu et al., 2020). Outside of pregnancy, an RCT of a water-based supplement containing 2.6g of an amino acid mixture (including L-Leucine, L-Threonine, L-Monohydrochloride, L-Isoleucine and L-Valine) and chromium picolate was found to decrease venous blood glucose (measures as an incremental area under the curve, $iAUC_{0\text{-}120min}$)) when consumed alongside a carbohydrate-rich meal in a cohort of overweight men and women (Östman et al., 2020). Furthermore, Leucine has also been found to improve glucose tolerance and insulin resistance in mice. Mice fed a high-fat diet alongside Leucine supplementation were found to have significantly lower glucose levels at all time points (15 minutes – 2 hours) following a GTT compared to mice who had only been fed a high-fat diet (Macotela et al., 2011) It has also been previously suggested that the increased consumption of leucine may stimulate insulin production resulting in a reduction in postprandial glucose levels in type 2 diabetics, particularly when combined with dietary interventions that improve the blood lipid profile (Leenders and van Loon, 2011). This 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 decrease of 0.193 mmol/L of fasting glucose and 0.327 mmol/L 2-hour post glucose respectively. 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 prevent GDM in WE populations.

### 5.6.1.2 Cholesterols

HDL cholesterol is seen as 'good cholesterol' due to its role in the removal of cholesterols from atherosclerotic plaque, in turn reducing an individual's risk of CVD (CDC, 2020). On the other hand, LDL cholesterol is seen as 'bad cholesterol' due to its ability to form atherosclerotic plaque deposits in blood vessels, increasing an individual's risk of CVD. Modifying dietary consumption of fats is widely utilised to control the ratio of HDL to LDL in the blood, with reductions in trans-fat consumption reducing LDL levels (Mayo Clinic, 2022). In addition to contributing to an increase in atherosclerotic plaque formation, low HDL levels have also commonly been associated with diabetes, with HDL being thought to increase insulin secretion and β-cell survival (Wong et al., 2018). Despite this, discrepancies in the direct effect of HDL cholesterol and dysglycemia have also been identified in the MR literature (Wong et al., 2018), with studies reporting heterogonous effects of HDL in relation to diabetes, with some finding a protective effect and others reporting no effect. To our knowledge, no previous study has conducted an MR of metabolites and GDM; a 2015 review did however find HDL levels to be higher in GDM cases compared to controls during the second and third trimesters, however heterogeneity was high in all analyses ($I^2 \geq 54\%$) (Ryckman et al., 2015).

Here, through the analysis of the BiB cohort, numerous associations between HDL cholesterol measures and postprandial glucose measures were identified in WEs; however, the directions of effects of these metabolites were mixed. Total cholesterol in HDL, total cholesterol in HDL2C, total cholesterol in HDL3C, total cholesterol in L-HDL, cholesterol esters in S-HDL, cholesterol esters in XL-HDL, the concentration of L-HDL and total lipids in L-HDL were all positively associated with 2-hour post glucose. Conversely, total cholesterols in M-HDL and cholesterol esters in M-HDL were negatively associated with fasting glucose. In addition to these identified associations for individual metabolites, the XL-HDL class was also associated with 2-hour post glucose in WEs. The majority of these associations were consistent across each leave-one out analysis; however, the removal of SNP rs6922 from the instrument for the phospholipids in L-HDL resulted in a large change in effect estimates. rs6922 is found in the *RGL1* gene and its protein product Ral

guanine nucleotide dissociation stimulator like 1 is classified as a metabolic protein under the HPA classification system due to its known role in the regulation of lipid metabolism (Reactome, 2022; Gillespie et al., 2021; GeneCards Suite., 2022b).

Although increased HDL levels have been associated with a decrease in GDM risk (Ryckman et al., 2015; Wang et al., 2019), this study identified numerous HDL metabolites to be associated with an increase in 2-hour post glucose. Indeed, some of these associations have been identified previously, with a Finnish sample of overweight and obese women finding cholesterol esters in S-HDL to be higher in the serum samples of GDM cases at ~14 weeks gestation (Mokkala et al., 2020).

When considering LDL cholesterols, the only LDL metabolite measure identified to be associated with pregnancy dysglycemia was the concentration of S-LDLs, which was associated with a decrease in fasting glucose of 1000 mol/L in WEs. However, when evaluating the collective effect of lipid classes, M-LDLs and 'all LDLs' (a combined measure of S-LDLs, M-LDLs and HDLs), were associated with fasting glucose. Additionally, the S-LDL class was found to be associated with both fasting glucose and 2-hour post glucose in WEs. Unfortunately, because composite scores were comprised of PC1 coordinates the direction of effect of these associations could not be evaluated. However, these results support the presence of a causal association between a range of cholesterol metabolites and pregnancy dysglycemia in WE populations.

The heterogeneous nature of the direction of associations obtained in this MR study highlight the complexity of biological mechanisms linking metabolites and disease in both an observational and MR setting, with even well-established biological mechanisms potentially being subjected to confounding from numerous sources, including diet, alternative biological molecules, and demographic factors such as BMI. In addition, the timing of sample collection is known to impact cholesterol levels during pregnancy, with total cholesterol, HDL cholesterol and LDL cholesterol have all been shown to be higher in third trimester compared to the first trimester (Emet et al., 2013). This temporal variation in cholesterol levels throughout pregnancy may result in different associations between cholesterol metabolites and pregnancy outcome measures being identified in studies that have taken metabolite

samples at different pregnancy time-points (Hu et al., 2021). OGTT samples in BiB were taken on the cusp of the second and third trimester, with the third trimester starting at the 27$^{th}$ week of pregnancy which means cholesterol levels are likely to be higher in these samples compared to ones taken during early pregnancy. Future work evaluating the role of cholesterol metabolites in pregnancy health outcomes should aim for large enough sample sizes to stratify by BMI with adequate power, while considering the timing of sample collection, potentially by comparing cholesterol levels in the sample with a non-pregnant control population to assess the extent of elevated cholesterol levels during pregnancy. Additionally, performing multiple MR analyses on samples collected at varied time points may also help shed light on the dynamic nature of metabolites during pregnancy and also the timing of GDM on set at a molecular level.

#### 5.6.1.3 Triglycerides

Triglycerides are an abundant class of lipid particles found in the blood. Like cholesterols, triglycerides have been associated with an increase in CVD risk and an increase in atherosclerotic plaque formation, although it is not known whether these associations are causal or a result of biological interactions between triglycerides and cholesterols (Wiesner and Watson, 2017). Triglycerides can enter the bloodstream either through intestinal absorption (the route of entry for dietary triglycerides) or as a result of hepatic synthesis (Alexopoulos et al., 2019). Once in the blood, triglycerides can be incorporated into HDL and LDL cholesterol particles where they contribute to cholesterol metabolism. Dietary sources of triglycerides include foods high in saturated fats, and it has been suggested that the best way to manage triglyceride levels is through a reduction in overall calorific intake by limiting the consumption of processed meat, oils and confectionery products (Laufs et al., 2020). In addition to triglyceride consumption, dietary fatty acids are also commonly converted into triglycerides before they enter circulation, highlighting the complex relationship between triglyceride, cholesterol, and fatty acid levels (Alexopoulos et al., 2019).

Increased triglyceride levels have been associated with both diabetes (Alexopoulos et al., 2019) and GDM. A 2021 meta-analysis reported that

triglyceride levels were higher in GDM cases than GDM non-cases in every trimester, although considerable heterogeneity was detected in all analyses ($I^2$ ≥ 84%) (Hu et al., 2021). Interesting, when stratifying studies by geographical location no difference in triglyceride levels in GDM cases and controls was identified in South Asian studies (Hu et al., 2021). 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 it is in WEs (Makshood et al., 2019).

Similar, to the above meta-analysis, no triglyceride metabolite values were associated with postprandial glucose measures in the SA stratum of the BiB cohort. However, triglycerides in XSVLDL (XS-VLDL-TG) were associated with an 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 (Mokkala et al., 2020). Despite this finding, no other triglyceride values were found to be associated with pregnancy dysglycemia in the WE strata of BiB in this analysis. One explanation regarding why no additional associations were detected could due to the average BMI of the WEs in BiB, with an additional analysis of a prospective Irish cohort (~94% WE) finding that triglyceride levels were only associated with GDM in obese individuals, a higher average BMI than that observed in the BiB cohort (O'Malley et al., 2020). If these findings of increased triglycerides in XSVLDLs are confirmed it is possible that this association is partially responsible for the identified associations between diets high in fats (such as the western diet) and increased GDM in WEs.

## 5.6.2 Identified associations in SAs

### 5.6.2.1 Fatty Acids

Polyunsaturated acids FAw3 and FAw6 are commonly consumed via diet, two of which are essential and cannot be produced by the body (i.e., alpha-LA and LA). (Koletzko et al., 2019). Once consumed, the body then

converts PUFAs to long-chain PUFAs (lc-PUFAs), from precursors such as LA, in order to produce membrane lipids, eicosanoids and docosanoids.

Changes in dietary patterns can have a large impact on fatty acid composition in the body, with a Western diet increasing n-6 fatty acid levels that could in turn impact GDM risk (Hosseinkhani et al., 2021). Increased saturated fat consumption (characteristic of a western diet) during pregnancy is also known to increase lipolysis in maternal adipose tissue, elevating the levels of free fatty acids and with it increasing insulin resistance (Hernandez et al., 2018). Broadly, dietary sources of n-3 fatty acids include plant oils, walnuts, flaxseeds, and oily fish, although the exact sources are dependent on the type of n-3 fatty acid (Gebauer et al., 2006). Furthermore, some PUFAs have been found to be present at more consistent levels throughout pregnancy compared to other types of fatty acids, potentially making them a more reliable biomarker for GDM if an association does exist (Hosseinkhani et al., 2021). This makes fatty acids an ideal target for dietary interventions aiming to prevent or manage GDM if the causal association between fatty acids and GDM is confirmed.

The association between fatty acids and dysglycemia in Asian populations has been previously reported. In a cohort of Chinese adults, total n-6 fatty acids and 18:2 n-6 levels at baseline were both found to associate with an increased risk of T2D after ~8 years of follow up, while n-3 fatty acids were protective against T2D (Bragg et al., 2021). However, a two-sample MR could only confirm a negligible effect of n-6 PUFA synthesis on T2D in a predominantly WE cohort (Zulyniak et al., 2020). In regard to GDM, a recent systematic review (2021) exploring the relationship between n-3, n-6 and n-9 fatty acids and GDM was inconclusive, and none of the identified studies (n=15) were conducted in a SA population (Hosseinkhani et al., 2021). A more recent observational study investigating the role of fatty acids in GDM development in 2 cohorts of Chinese individuals also found mixed results regarding the associations between a range of fatty acids (measured at approximately 10-14 weeks) and GDM incidence. Together these studies highlight the need for more studies exploring the role of fatty acids in GDM developed in Asian populations (Pan et al., 2021).

Through the utilisation of one-sample MR this study has provided evidence for the presence of an association between FAw6 and LA levels and an increase in 2-hour post glucose levels in SAs. In addition, the fatty acid metabolite class was also found to be associated with 2-hour post glucose in SAs. In addition, FAw3 was found to be associated with an in increase 2-hour post glucose, however only one SNP was included in the instrument which meant potential for pleiotropy could not be fully explored. Through the use of a leave-out-one sensitivity analyses for the FAw6 and LA instruments, the removal of the SNP rs12720820 (found within the *APOB* gene) was found to result in non-significant associations for both exposures, indicating that this SNP was the largest contributor to the identified associations. *APOB* is responsible for encoding two versions of Apolipoprotein B, a protein involved in fat and cholesterol transport; indeed, rs12720820 has previously been associated with cholesterol levels and the use of cholesterol lowering drugs (MedlinePlus, 2021). It is well established that fatty acid profiles can impact blood cholesterol levels (Mensink et al., 2003; Jacobson et al., 2012; Ooi et al., 2013). In addition, increased dietary cholesterol has previously been associated with an increased risk of GDM in a systematic review of observational studies (Schoenaker et al., 2016). Taken together, this evidence suggest that it is possible that fatty acids and cholesterol metabolites may be vertically pleotropic, impacting postprandial glucose and measures and GDM risk via the same causal pathway, with n-6 fatty acids fatty acids impacting GDM risk via their role in cholesterol metabolism. Unlike horizontal pleiotropy, vertical pleiotropy does not result in a violation of the 2$^{nd}$ MR assumption as cholesterol is not acting as a confounder, meaning MR estimates are still valid (**Figure 5.2**). 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, despite numerous associations being identified between cholesterol measures and postprandial glucose measures in WEs. In addition to possible variations in cholesterol metabolism, it is also plausible that variations in fatty acid synthesis are partially responsible for the increased GDM risk experienced by SAs. For example, variants within the *FADS* genes have been shown to impact lc-PUFA conversion, with haplotype D resulting in

a higher blood lipid level of lc-PUFAs due to higher rates of lc-PUFA conversion, while haplotype A is associated with lower conversion (Koletzko et al., 2019; Ameur et al., 2012). Currently, evidence suggests that in SAs haplotype D is slightly more common than it is in WEs, highlighting potential underlying differences in lipid metabolism which could differentially impact disease risk in both populations (Koletzko et al., 2019; Ameur et al., 2012).

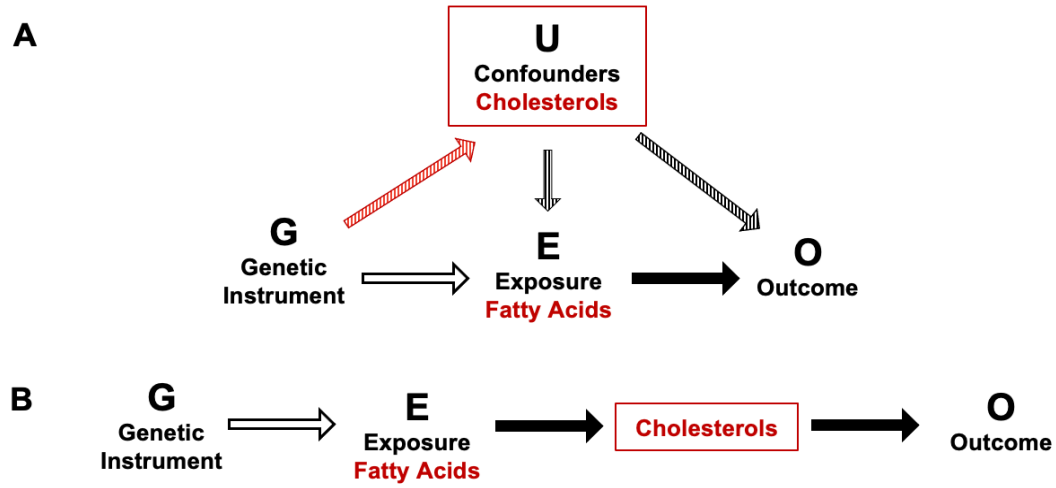



**Figure 5.2 Schematic of horizontal and vertical pleiotropy. A:** Illustration of horizontal pleiotropy. **B:** Illustration of vertical pleiotropy. Vertical pleiotropy does not result in a violation of the 2nd MR assumption.

Unfortunately, due to the data cleaning steps prior to MR no SNPs within the *FADs* genes were included in the GWAS of metabolites in SAs meaning it was not possible to determine whether i) these SNPs were associated with fatty acids in this study and, if so, ii) these SNPs impacted the MR result. Despite this, shared and distinct patterns of genetic associations were identified when comparing GWAS results in both ethnicities, indicating that throughout the genome there may be ethnic specific genes associated with fatty acid metabolism **(Appendix Figure C.17).** If these ethnic differences in fatty acid metabolism are confirmed, this finding could aid in the development of ethnically tailored GDM prevention strategies that focus on modifying fatty acid profiles in an ethnic-specific manner.

### 5.6.3 Strengths and limitations

This analysis has several strengths. Firstly, this study involved a large and comprehensive panel of metabolites spanning a range of metabolite classes allowing for the role of the metabolites in 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 2$^{nd}$ and 3$^{rd}$ 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 associated with a postprandial outcome measure as metabolites may be highly correlated or in the same biological pathway to the true causal metabolite (vertical pleiotropy) (Burgess and Harshfield, 2016). In this scenario, it is only possible to determine the role of the causal association of the overall pathway rather than an individual metabolite. In an attempt 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 exposure from conception to the time of measurement is constant, which is unlikely when studying metabolites. The limited sample size of this study also means that some associations may have been underpowered to detect an effect; however, a post hoc-power analyses did find 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 on SAs a two-sample MR could not be conducted as sensitivity analyses meaning it was not possible to assess the generalisability of the findings identified in this study.

### 5.6.4 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 have been established. 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 of a lower risk of GDM. In high-risk SA women total n-6 fatty acids and the n-6 fatty acid, LA appears to increase postprandial glucose levels suggesting that fatty acids may be partially responsible for the increased GDM burden experienced by this population. Future work in a larger sample (potentially utilising a two-sample MR) should aim to utilise a larger panel of fatty acid metabolites in SAs to confirm these findings, ideally over the course of a pregnancy in order to aid in GDM prevention in this high-risk population.

## 5.7 Summary

- Metabolites are causally associate with postprandial glucose in both SAs and WEs.
- Cholesterols and lipoproteins are the most common type of metabolite to be causally associated with postprandial glucose in WEs.
- Fatty acids are the most common type of metabolite to be causally associated with GDM in SAs.
- Distinct metabolite classes are associated with postprandial glucose in both ethnicities, with members of the L-HDL and S-HDL classes being associated in both ethnicities.

# Chapter 6: Conclusions and Future Work

## 6.1 Main findings

The main aim of this thesis was to determine whether metabolites are causally associated with GDM in an ethnic-specific manner, with a focus on high-risk SA women. In addition, as diet is also the mainstay in GDM prevention and can alter the metabolome, this study also aimed to examine the presence of associations between diet and GDM in distinct ethnic populations.

The first research chapter of this thesis (Chapter 3) reviewed the role of diet on GDM incidence in ethnically distinct populations, evaluating evidence from both randomised and observational settings in order to determine whether the effect of diet on GDM incidence was similar across ethnic groups. Through conducting a thorough systematic review and meta-analysis, 38 studies (32 observational studies and 6 RCTs) that reported on the association between diet and GDM were identified. Encompassing 5 ethnicities, 16 individual dietary patterns and 2 overarching dietary patterns (healthy and unhealthy), this review did not find evidence to support the use of healthy dietary interventions during pregnancy as a means of preventing GDM, agreeing with previous reviews (Griffith et al., 2020). However, an association between healthy diets and unhealthy diets assessed in an observational setting was identified in WE populations, with a healthy diet reducing an individual's odds of GDM by 24%, while an unhealthy diet increased an individual's odds by 59%. In addition, associations were also identified for the western, fried/fast food, a high meat and high protein dietary patterns and increased GDM risk in WEs. However, in Asian populations, no associations were identified between any dietary pattern and GDM incidence, before or after stratifying the Asian population by east and south Asian subgroups despite adequate power to detect even small changes in effect. Three macronutrients, expressed as a percentage of total energy intake (fat, animal protein and carbohydrate) were however associated with GDM incidence in both WE and Asian populations, with increased intake of fat and

animal protein raising the odds of GDM while increased carbohydrate intake decreased the odds. In subsequent sensitivity analyses, BMI and maternal age were not found to change the significance or magnitude of identified associations in WEs. Furthermore, no dose-response relationship was identified for any exposure. When considering observational studies that assessed dietary intake during pregnancy, no association was identified between diet and GDM in WEs, agreeing with the absence of an association identified between RCTs implemented during pregnancy and GDM incidence, collectively suggesting that for healthy diets to have a protective effect against GDM they must precede pregnancy. Finally, in a *post-hoc* power analysis in the Asian population power was consistently found to be adequate (at the 80% level) to detect the changes in odds of GDM identified in WEs. This means that, despite the lower number of studies identified in Asians compared to WEs, the analysis should have been adequately powered to detect even small associations between diet and GDM incidence in Asian populations. With the majority of identified studies conducted in non-white populations utilising validated culturally tailored interventions (66% of non-white RCTs) or dietary assessment tools (87.5% of non-white observational studies), it can be assumed that diet was captured to the same level of accuracy in both populations. In summary, evidence from this review suggests that diet may have differing impacts on GDM development in WEs and Asians and it is possible that the absence of observed associations in Asians is not due to a lack of culturally specific interventions or dietary assessment tools utilised in Asian studies. This suggests that the absence of an effect between diet and GDM in Asians may be due to biological differences between the populations.

In an attempt to establish whether any potential differences in the association between diet and GDM between ethnicities was as a result of ethnic-specific biological differences, Chapter 4 aimed to identify whether the metabolomic profiles of GDM cases and non-GDM cases differed in an ethnically diverse cohort (50% WEs and 50% SAs), due to the well-established relationship between diet and the metabolome. Furthermore, this chapter aimed to investigate whether the metabolic profiles of GDM cases and non-cases differ by ethnicity to see if the absence of association observed

between diet and GDM in Asians in Chapter 1 could also be observed at the metabolite level in a SA population.

In an overall analysis of the BiB cohort, metabolites explained 11.4 % of the variation in GDM status after accounting for known clinical GDM risk factors (i.e., maternal BMI, maternal age, parity, multiple pregnancy, smoking status, and ethnicity). In this analysis, 7 metabolite values (lactate, mean diameter of VLDL, TotFA, MUFAs, LA, SFA and esterified cholesterol) were considered important (VIP ≥1) to characterise GDM status, all of which are dietary related metabolites. Interestingly, in Chapter 3 fat intake (expressed as a % of energy intake) was one of the only dietary components found to associate with GDM incidence in Asians and was the most common class of metabolites to be determined to be important in distinguishing GDM in SA from the BiB cohort, albeit to a lesser extent than in WEs.

After stratifying the cohort by ethnicity, these metabolites remained important along with glycoprotein acetyls in characterising GDM status in both ethnicities. A further 6 metabolites (FAw6, PUFAs, alanine, glutamine, serum cholesterol, and citrate) were deemed important to characterise GDM status in WEs, all of which are also dietary related and adding further support to the associations between diet and GDM in WEs identified in Chapter 3. Despite these additional associations in WEs, no unique metabolite-GDM associations were identified in SAs. It was also noted that when applied to the SA ethnicity the models explained less of the variation in GDM status than they did in WEs (20% vs 26%). Also, many of the metabolite values identified as characteristic of GDM in both ethnicities' VIPs were significantly higher in WEs than in SAs (MW p value ≤ 0.05). Taken together, it appears that the maternal metabolome characterised in BiB is more closely associated with GDM case status in WEs than SAs. This builds on the last chapter suggesting that diet may have a more important impact in determining GDM status in WE compared to Asian populations. In this chapter, dietary related metabolites characteristic of GDM in both ethnicities (including total fatty acids, SFAs and MUFAs) were more characteristic of GDM in WEs, suggesting that diet may be more important in determining GDM status in WEs.

Finally, this chapter observed that low risk (i.e., non-smokers who carried a singleton pregnancy and had no previous children) healthy-weight

(BMI ≤ 23kg/$m^2$) SA women with GDM have a distinct metabolic profile when compared to all other low-risk SA and WE cases and individuals. This pattern identified in healthy-weight SA women suggests that low-risk SAs have a unique metabolic profile that is associated with an increased GDM risk. Future work should aim to confirm these findings in a larger cohort while also aiming to identify drivers of this elevated risk profile in healthy-weight SAs.

All statistical analysis in Chapter 4 utilised data from the prospective BiB birth cohort. Despite the fact that all analyses were adjusted for known GDM risk factors, residual confounding is likely to still be present in this analysis due to the fact that BiB is an observational cohort. In an attempt to bypass this issue of residual confounding, Chapter 5 of this thesis utilised an MR design that can account for residual confounding. Herein, the presence of causal associations between metabolite values and postprandial glucose measures (key components of a GDM diagnosis) in both ethnic groups in BiB using one-sample MR. By doing so, causal associations between metabolites identified to be important in distinguishing GDM cases from non-cases in a multivariate approach could be confirmed and the direction and magnitude of effects assessed, something that could also not be achieved in the utilised multivariate approaches.

In both ethnicities, GWASs identified SNPs that were associated (p value ≤ 1x $10^{-5}$) with the majority of metabolite values. Interestingly, for a range of metabolites, shared and unique genomic regions were found to be associated with metabolite levels between SAs and WEs. This suggests that shared and unique genetic markers of metabolism exist in SAs and WEs, relating to the previous findings from Chapter 4 that that found common and distinct features in the metabolomic profiles of GDM cases and non-cases between the ethnicities.

The MR analysis found that 14 metabolite values were causally associated with fasting glucose or 2-hour post glucose in WEs whilst two metabolite values (leucine and mean diameter of HDLs) were associated with both postprandial glucose measures in WEs. In WEs, cholesterol and lipoprotein-related metabolites were the most represented class of metabolites to be associated with postprandial glucose measures. In the SA population, 11 metabolite values were causally associated with postprandial

glucose, including three fatty acid measures (FAw6, LA, FAw3). The identification of more causal associations in WEs compared to SAs agrees with the findings of Chapter 4 which suggested that metabolites may have a more influential role in determining GDM status in WEs compared to SAs.

Building on from Chapter 4 which accounted for correlation between metabolites utilising multivariate statistical approaches, in Chapter 5 MR analyses were also performed utilising PC1 from the PCA analysis of highly correlated metabolite classes (i.e., classes where PC1 explained ≥70% of the class variation). In this analysis the fatty acid class was associated with 2-hour post glucose in SAs while the XL-HDL, M-LDL, S-LDL and All LDL classes were associated with either fasting glucose or 2-hour post glucose in WEs. This finding of fatty acids being the only metabolite class to be causally associated with postprandial glucose agrees with the finding from Chapter 4 that found fatty acids to be the most frequent class of metabolites to be classified as being important in determining GDM case status in SAs.

The majority of associations were robust, with only negligible shifts following the successive removal of individual SNPs in each instrument (leave-one-out sensitivity analysis). F-statistics were constant across each leave-out-one out analyses indicating that individual SNPs were not driving the identified associations and all potential confounders identified in Phenoscanner and GWAS Catalog are plausibly believed to be in vertical rather than horizontal pleiotropy. Therefore, all assumptions of MR were supported in the present analysis where it was demonstrated that a range of metabolite values (including dietary-related metabolites) are causally associated with pregnancy dysglycemia in an ethnic-specific manner. The identification of causal associations between fatty acid metabolites and postprandial glucose measures in SAs agrees with the finding of fatty acid metabolites being the most frequent class of metabolites identified in multivariate analyses in Chapter 4, and the association between fat intake and GDM in Asians identified in Chapter 3. Likewise, in WEs, leucine and numerous measures of cholesterol intake were found to be causally associated with postprandial glucose in WEs, all of which are dietary related and could help explain the observed associations between diet and GDM

identified in Chapter 3. A summary of all major findings of this thesis can be seen in **Figure 6.1**.

Taken together, these findings suggest that biological pathways involved in fatty acid, cellular (i.e., glycolytic metabolites) and cholesterol metabolism may all be important in determining GDM status and future work should focus on better understanding the mechanisms that link dysregulation in these pathways to GDM. Clinically, the findings of this project emphasise the fact that individuals from different ethnicities may have underlining biological risk differences to GDM and highlights the need of personalised care. Furthermore, the absence of any detected effect between diet and GDM in Asian populations in Chapter 3 of this thesis further empathises the need for culturally sensitive dietary assessment tools and nutritional advice to be developed and utilised. In addition, these findings have also shown the importance of adopting healthy dietary changes as early as possible during pregnancy, and ideally pre-conception for women of child bearing age in all ethnicities. From a public health perspective it is critical that the importance of pre-conception diet to help ensure a health pregnancy is sufficiently communicated to women and that the importance of ethnicity as a clinical risk factor is understood by those conducting research. Ethnically disaggregated data should be collected whenever possible and ideally recruitment strategies should allow for sufficient numbers in minority ethnicities to prevent the grouping of culturally distinct ethnic subgroups into broader ethnic categories (i.e., Asian rather than EA or SA), which may obscure important cultural and biological differences between the groups. For women, the findings of this thesis also highlights the importance of pre-pregnancy diet in determining conception health but also how modifiable risk factors (such as fatty acid intake) may contribute to their disease risk.

## 6.2 Limitations and future work

The main aim of this PhD was to test whether metabolic profiles, as reflected by metabolites, play a causal role in GDM across and within ethnically distinct populations – namely WEs and SAs – in order to better understand the role of metabolism in GDM prevention. To build upon the

conclusions and limitations of this work, recommended future investigations are summarised in **Table 6.1**. More detail on the future work that could be conducted to build upon each research chapter of this thesis is also outlined below.

### 6.2.1 Chapter 3

Chapter 3 of this thesis did not find evidence that dietary interventions during pregnancy are effective in preventing GDM in any ethnicity, agreeing with the most recent Cochrane review (Griffith et al., 2020) . Furthermore, this study also found that when stratifying observational studies by the time of dietary assessment (before vs during pregnancy) diets assessed in an observational setting were not effective in preventing GDM either, suggesting that adopting a healthy diet during pregnancy may not be sufficient in preventing the metabolic changes that contribute to GDM. Evidence from Chapter 3 does, however, support the association between pre-pregnancy diet and GDM in WEs. To assess the importance of diet pre- and peri-conception on GDM risk: (i) a prospective cohort evaluating dietary intake at multiple time points before pregnancy could be used to assess how long a healthy diet would need to be adopted before pregnancy to achieve effective GDM prevention, or (ii) an RCT of women attending family planning clinics may be prescribed healthy diets until or throughout conception. One limitation of this review was that it utilised FFQ results to assess adherence to dietary patterns. Adherence to dietary patterns is likely to fluctuate over time which would not be captured by a single FFQ, so individuals classified as following a healthy diet may actually be misclassified, biasing any observed association between diet and GDM. Future cohorts should aim to conduct multiple FFQs throughout pregnancy to better understand how adherence to dietary patterns changes during pregnancy and in the pre-conception period, especially as it is known that women often alter their diets when trying to become pregnant/ once pregnant.

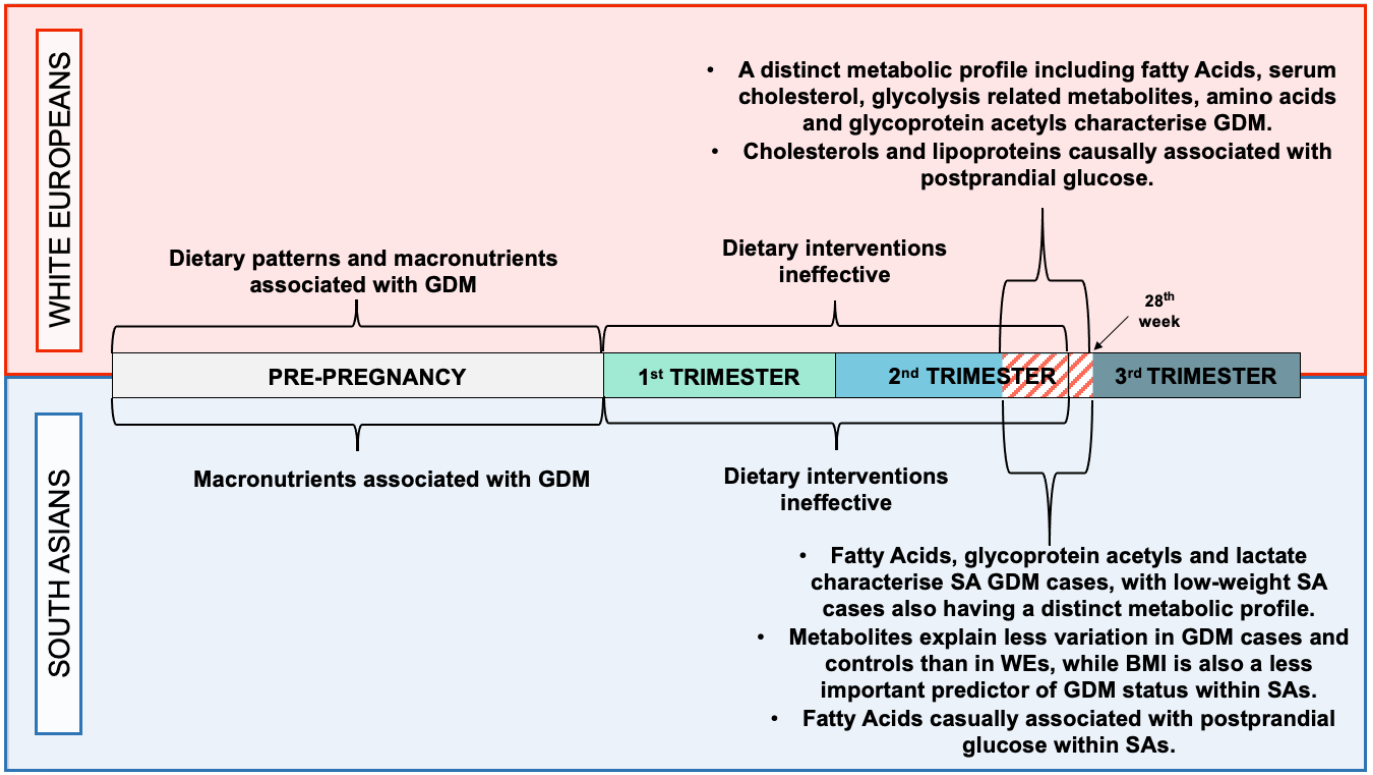



**Figure 6.1: Schematic of the main findings of this thesis.** Dashed lines represent the time of sample collection and GDM diagnosis via a 75g OGTT in Born in Bradford. SA: South Asians. WE: White Europeans.

A further limitation of this review was that a very limited number of studies were conducted in SA populations, with the majority of Asian studies involving Chinese populations. Although sensitivity analyses were conducted to assess differences in the results between SAs and EAs, it is possible that these analyses were underpowered to detect any difference in effect (indeed, it is known that EAs and SAs experience differing GDM risks). Ideally, future cohorts would be adequately powered to investigate ethnic-specific associations in a range high-risk SA populations. This may be achievable through utilising a cohort including a SA immigrant population which could allow for multiple SA groups to be assessed in a single country or research site. Furthermore, by directly comparing an established non-WE cohort to a cohort of non-immigrant WEs, dietary patterns may be more comparable between groups due to the groups having a more similar environment when residing in the same geographical location. In addition, through the assessment of individual foodstuffs (rather overall dietary patterns) it may be easier to compare association between diet and GDM between ethnicities due to the fact that foodstuffs may be more comparable between ethnicities, although preparation methods may differ. Through the consideration of individual foodstuffs it may be possible to compare the impact of diet across populations while also minimising heterogeneity: a limitation of the systematic review and meta-analysis conducted in Chapter 3.

It is imperative that any future study examining the role of diet in GDM incidence in SAs utilises a culturally sensitive GDM assessment tool that considers the consumption of culturally specific food items as well as any difference in preparation methods. Ideally, any tool utilised in a SA population would collect as much nutritional information as feasibly possible when assessing intake, since the composition of foods frequently consumed in non-white ethnicities may differ considerably from commonly consumed foodstuffs in WEs. Furthermore, if a suitable culturally sensitive dietary tool was produced that could be tailored for use in different ethnicities it may help in the comparison of the effect of dietary patterns across ethnicities; herein, the use of PCA in individual studies to assess adherence to dietary patterns likely contributed significantly to the high heterogeneity estimates seen in the meta-

analyses, which may be reducible by the use of an appropriate standardised tool.

### 6.2.2 Chapter 4

This study provided evidence of distinct metabolomic profiles present at the time of OGTT in WE and SA women. Due to the fact that these samples were taken at the time of OGTT, it is possible that they represent changes in the metabolome that have occurred concurrently with GDM development. In order to bypass this limitation and better understand the temporal relationship between metabolites and GDM (and whether this changes during the time course of pregnancy) a prospective cohort where fasting serum samples are repeatedly collected in early pregnancy would offer a better understanding of the role of metabolites in GDM development, a finding have could have significant clinical value. These analyses would also build on the finding that healthy-weight SA cases had distinct metabolic profiles compared to other SA cases which, if replicated during early pregnancy, could help identify SAs who are at high risk of GDM even though they have a healthy BMI.

In addition to collecting metabolite samples at a wider range of time points during pregnancy, utilising a different metabolite panel to quantity the metabolite profiles in the samples may also be beneficial. The Nightingale Health © panel utilised by BiB has the advantage of being affordable and comprehensive, covering a range of metabolite classes making it ideal for exploratory analyses. However, this thesis has identified a range of metabolites as being of particular importance in determining GDM in both Chapters 4 and 5. Utilising a metabolite panel that includes more metabolites in the identified classes would be beneficial. For example, fatty acids were consistently identified to be important in determining GDM cases from non GDM cases in multivariate analyses (Chapter 4) and MR (Chapter 5), however, most measures of fatty acids in BiB encompass a class of fatty acids (for example, total n-3 fatty acids and total PUFAs) rather than individual fatty acids. This limitation means it is not possible to determine which individual members of each fatty acid class are responsible for the identified associations, information that could aid in the development of culturally specific dietary interventions. By utilising a more comprehensive panel of fatty

acids it may be possible to better understand the biological pathways which link fatty acids to GDM development and to identify which fatty acid should be targeted in dietary interventions in GDM prevention. Based upon the findings of Chapters 4 and 5, in addition to fatty acids, cholesterols, amino acids, and glycolysis related metabolites should also be explored in more depth in relation to GDM also.

### 6.2.3 Chapter 5

In this thesis, a one-sample MR was conducted to evaluate the presence of causal associations between a range of metabolites and two measures of postprandial glucose in 2 ethnicities. Although this sample had the benefit of including a large number of SAs (particularly when compared to the current literature) a sample size of approximately 3,500 is still comparatively small compared to large genetic consortia which may prevent the identification of weaker associations identified due to a lack of power. Furthermore, as a result of this sample size (and with it the limited number of GDM cases), it was not possible to investigate GDM as an outcome via MR. In order to examine genetic risk factors for GDM in SAs, larger genetic consortia are needed in SAs populations, ideally including multiple SA groups due to the unique genetic architecture present in some SA populations as a result of the cultural practice of marrying in Baradaris which could have limited the generalisability of the findings of this thesis. Likewise, larger cohorts of GDM cases of WE ancestry are also needed, particularly due to the lower prevalence of GDM in this population.

Generalisability of the findings may also be limited due to the fact that a one-sample MR approach was utilised meaning only one cohort was considered in the analysis. Additional GWASs in relation to both metabolites and GDM would also allow for two-sample MR to be conducted in relation to this topic, hopefully allowing for the findings of Chapter 5 of this thesis to be confirmed in multiple cohorts. In addition to larger cohorts and additional GWAS studies, it is also important to assess the causality of the relationship between metabolite measures taken during early pregnancy and postprandial glucose measures due to the fact that the metabolome varies during pregnancy (Monni et al., 2021). Confirmation of whether the same

associations between metabolite values and postprandial glucose measures identified in this study are also identified in early pregnancy may aid in the development of dietary interventions designed to be implemented in early pregnancy.

### 6.2.4 All chapters

A consistent theme arising from this thesis is the potential role of fatty acids in modulating GDM risk in SAs: in Chapter 3, fat (expressed as % total energy intake) was one of the only dietary-related associations identified in Asians, in Chapter 4 fatty acids were the most represented metabolite class in the metabolite panel that characterised GDM in SAs, and in Chapter 5 causal associations were identified for a range of fatty acid measures and postprandial glucose measures in SAs. When considering all results, a RCT aiming to reduce fat consumption in SA women of child-bearing age may be effective in reducing GDM incidence, although this may be methodologically challenging to implement due to difficulties that arise when trying to implement a RCT for an extended time period. Finally, it is important to note that throughout this thesis it appears that metabolites may make a smaller contribution to GDM risk in SAs than WEs. Although metabolites (particularly fatty acids) may be partially responsible for GDM incidence in SAs, it is important that other avenues are explored for potential explanations for the high GDM burden in SAs, such as differences in inflammatory factors or pregnancy hormones, especially as GDM incidence continues to rise globally.

### 6.2.5 Conclusions

In conclusion, this thesis has found significant evidence that pre-pregnancy diet impacts GDM incidence in WE populations, however this same evidence currently does not exist in Asian populations. Furthermore, this thesis has also established that distinct metabolic profiles characterise GDM risk in SAs and WEs. Finally, this thesis has identified ethnic-specific causal associations between a range of metabolites and postprandial glucose measures. Overall it appears that the metabolome may have a greater impact in determining GDM risk in WEs, despite the lower risk of GDM in this population. Despite this, fatty acids appear to be the most important metabolite class in determining GDM risk in SAs. Future work should aim to better understand the relationship between fatty acids and GDM in SAs. in order to aid in the development of culturally sensitive dietary interventions tailored towards this high risk population.

| Related Chapter | Aim | Proposed Methods |
|---|---|---|
| **3** | To determine the temporal relationship between pre-pregnancy diet and GDM in a multi-ethnic population<br>To investigate whether pre-pregnancy diet is associated with GDM in a large SA population, using well defined culturally sensitive dietary assessment tools that also capture the intake of individual foodstuffs, allowing for a better comparison between populations | A prospective multi-ethnic cohort study in a multi-ethnic examining the dietary intake in women of childbearing age at multiple time points before pregnancy, ideally including women of different ethnic ancestries residing in the same country. |
| **4** | To determine how the relationship between metabolites and GDM changes from early pregnancy in WEs and SAs<br>To determine how fatty acid profiles differ in SAs and WEs during pregnancy in relation to GDM status | Prospective cohort of fasting serum samples taken at multiple time points during pregnancy. Metabolite quantification using a more comprehensive panel of fatty acid metabolites |
| **5** | Confirm findings from MR analyses of the relationship between metabolite values and in a secondary cohort ideally including multiple SA populations, potentially utilising a two-sample MR design<br>Perform MR analyses to examine the relationship between metabolite values and GDM status | Genetic sequencing of a larger sample of SAs, including more GDM cases alongside frequent literature searches of published GWASs related to metabolites, GDM or postprandial glucose in WEs and SAs. |
| **3+4+5** | To determine whether dietary interventions focusing on the reduction of fats are effective in preventing GDM in SAs | An RCT initiated before pregnancy in women planning to conceive or as early as feasibly possible in a cohort of SAs and WEs |

**Table 6.1 Summary of proposed future work from this thesis.**

# Chapter 7: Scientific Contributions from this PhD

## 7.1 Published peer review articles

1. **Fuller, H.**, Moore, J.B., Iles, M.M., Zulyniak, M.A. Ethnic-specific associations between dietary consumption and gestational diabetes mellitus incidence: a meta-analysis, *PLOS Global Health.* 2(5): e0000250. https://doi.org/10.1371/journal.pgph.0000250

## 7.2 Peer review articles under review

1. **Fuller, H.**, Iles M.M., Moore, J.B., Zulyniak M.A., A unique metabolic profile in South Asian women characterizes an elevated risk for gestational diabetes: An Analysis of the Born in Bradford Cohort. *medRxiv* 2022. https://doi.org/10.1101/2022.04.11.22273658

## 7.3 Peer review articles in progress

1. **Fuller, H.**, Iles M.M., Moore, J.B., Zulyniak M.A. Metabolic drivers of dysglycemia in pregnancy: A Mendelian Randomisation study within a multi-ethnic birth cohort.

## 7.4 Conference oral presentations and invited talks

1. **European Congress on Obesity, 2020:** Current practice of GDM prevention through dietary intervention is ineffective in highest risk ethnic groups. **Fuller, H.,** Moore, J.B., Iles, M.M. and Zulyniak, M.A. (doi.org/10.1111/obr.13115)

2. **The 7th PhD Conference in Food Science and Nutrition, 2020:** Serum dietary-related metabolites are predictive of GDM risk in an ethnic specific manner: A multivariate analysis of the Born in Bradford cohort. **Fuller, H.**, Iles, M.M., Moore, J.B. and Zulyniak, M.A.

3. **American Society of Nutrition, 2021:** Distinct Serum Metabolic Profiles in Early Pregnancy Characterise Gestational Diabetes Mellitus

Incidence in High-risk Women. **Fuller, H.**, Iles, M.M., Moore, J.B. and Zulyniak, M.A. (doi.org/10.1093/cdn/nzab046_045)

4. **The 8th PhD Conference in Food Science and Nutrition, 2021:** The role of the Metabolome in the development of Gestational Diabetes Mellitus in high-risk minority women. **Fuller, H.**, Iles, M.M., Moore, J.B. and Zulyniak, M.A.

5. **NUTRIGEN Fellows Seminar, McMaster University, December 2021**: The Role of the Metabolome in the Development of Gestational Diabetes Mellitus in High-risk Minority Women.

## 7.5 Conference poster presentations

1. **UK Association of the Study of Obesity, 2019:** The impact of diet on gestational diabetes mellitus incidence within distinct and diverse ethnic populations: A systematic review and meta-analysis. **Fuller, H.,** Moore, J.B., Iles, M.M. and Zulyniak, M.A. (doi/ 10.1530/obabs.01.P58)

2. **The 6th PhD Conference in Food Science and Nutrition, 2019:** Current Practice of GDM Prevention Through Dietary Intervention is Ineffective in Highest Risk Ethnic Groups. **Fuller, H.**, Iles, M.M., Moore, J.B. and Zulyniak, M.A.

3. **Nutrition Society Summer Conference, 2020:** Predictive metabolomic profiling within a diverse population of pregnant women at low or high risk of gestational diabetes. **Fuller, H.**, Iles, M.M., Moore, J.B. and Zulyniak, M.A. (doi.org/10.1017/S002966512000765X)

4. **Nutrition Society Summer Conference, 2021:** Serum dietary-related metabolites are predictive of GDM risk in an ethnic specific manner: A multivariate analysis of the Born in Bradford cohort. **Fuller, H.**, Iles, M.M., Moore, J.B. and Zulyniak, M.A. (doi.org/10.1017/S0029665121003268).

5. **Society of Social Medicine and Population Health, 2021:** The role of the serum metabolome in driving GDM in white Europeans and high-risk Pakistani women: A multivariate analysis of the Born in Bradford cohort. **Fuller, H.**, Iles, M.M., Moore, J.B. and Zulyniak, M.A.

6. **European Congress on Obesity, 2022:** Metabolic differences in a multi-ethnic birth cohort contribute to disparity of risk for gestational diabetes: A Mendelian Randomisation analysis of metabolites. **Fuller, H.**, Iles, M.M., Moore, J.B. and Zulyniak, M.A.

## 7.6 Submitted conference abstracts

1. **International Society for Developmental Origins of Health and Disease, 2022:** Metabolic drivers of dysglycemia in pregnancy: A Mendelian Randomisation study within a multi-ethnic birth cohort. **Fuller, H.**, Iles, M.M., Moore, J.B. and Zulyniak, M.A.

## 7.7 Prizes, awards and additional outputs.

1. Best Poster, 7th PhD Conference in Food Science and Nutrition (2020).
2. Young Investigators Bursary for the European Association for the Study of Liver Conference (2022).
3. School of Food Science and Nutrition, 2020 conference organising committee.

## 7.8 Scientific collaborations

### 7.8.1 Published peer review articles

1. Zulyniak, M.A., **Fuller, H.**, Iles, M.M. Investigation of the causal association between long-chain n6 polyunsaturated fatty acid synthesis and risk of type-2 diabetes: a Mendelian randomisation analysis. *Lifestyle Genomics.* 2020.

### 7.8.2 Peer review articles in progress

1. Zhang, Z., **Fuller, H**., Burrows, K., Abeysekera, K., Thorne, J. L., Lewis, S. J., Zulyniak, M. A. & Moore, J. B. Non-alcoholic fatty liver disease and vitamin D in the UK biobank: a two-sample bi-directional Mendelian randomisation.
2. **Fuller, H.,** Tittanegro, T.H., Maini A.A., China, L., Thorne, J.L, Moore, J.B., O'Brien A. Lipidomics analyses of ATTIRE trial patients' plasma at day 1 demonstrates that reduced cholesterol esterification predicts development of hospital acquired infection.

### 7.8.3 Conference abstracts

1. **Nutrition Society Spring Conference, 2021:** Characterisation of microRNAs regulated by vitamin D and lipid loading in immortalised hepatocytes. Zhang, Z., Moon, R., **Fuller, H**., Thorne, J.L. and Moore, J.B. (doi.org/10.1017/S0029665121000689)

2. **Nutrition Society Summer Conference, 2021:** Characterisation of microRNAs regulated by vitamin D and lipid loading in immortalised hepatic stellate cells. Zhang, Z., Moon, R., **Fuller, H**., Tan, X., Holmes, M.J., Thorne, J.L. and Moore, J.B.

3. **European Association for the Study of Liver, 2022:** Lipidomics analyses of ATTIRE trial patients' plasma at day 1 demonstrates that reduced cholesterol esterification predicts development of hospital acquired infection. **Fuller, H.,** Tittanegro, T.H., Maini A.A., China, L., Thorne, J.L, Moore, J.B., O'Brien A. **(Oral presentation)**

# Appendix A
# Supplementary data for Chapter 3

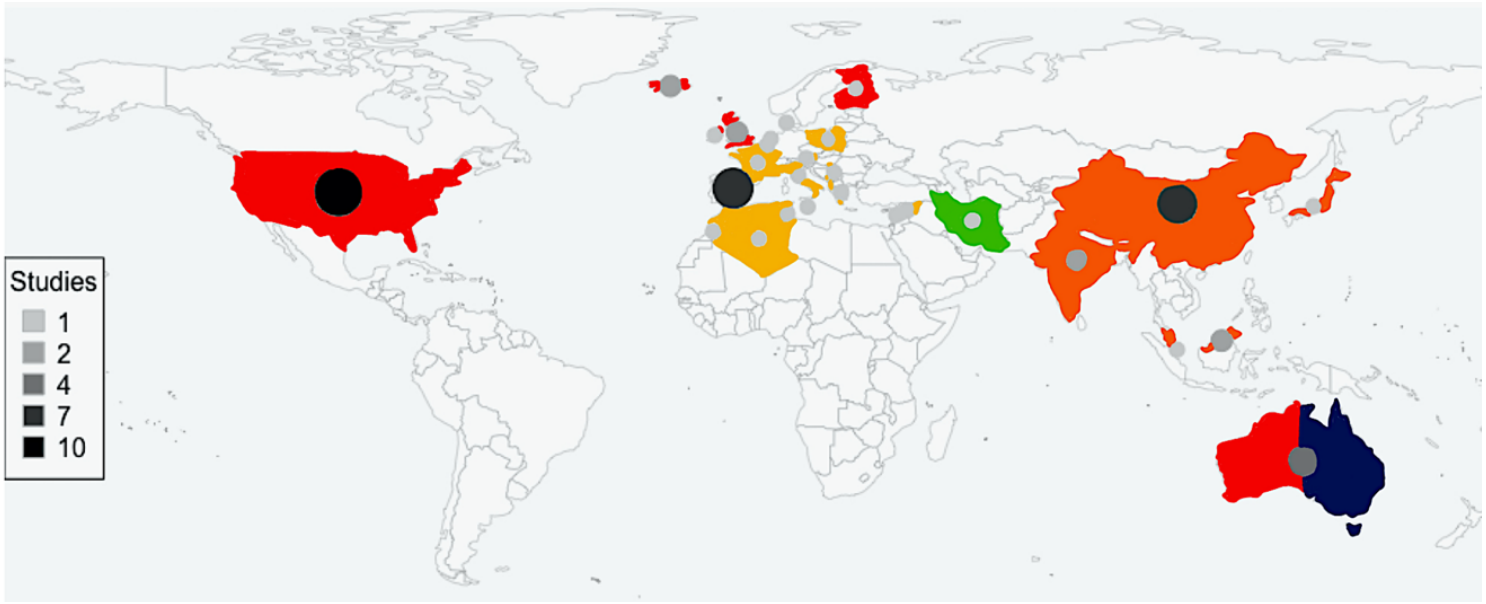



**Figure A.1: Map of studies included in the systematic review.**

Studies shown on the map are greater than the number of studies as some studies were multinational. The countries shading represents the ethnic subgroup that the country was grouped as. Red: WE, yellow: Mediterranean, green: Iranian, orange: Asian, navy: Australian Nationals. (~60% WE and ~40% Asian). Base map was obtained from the *rworld* map package in R studio.

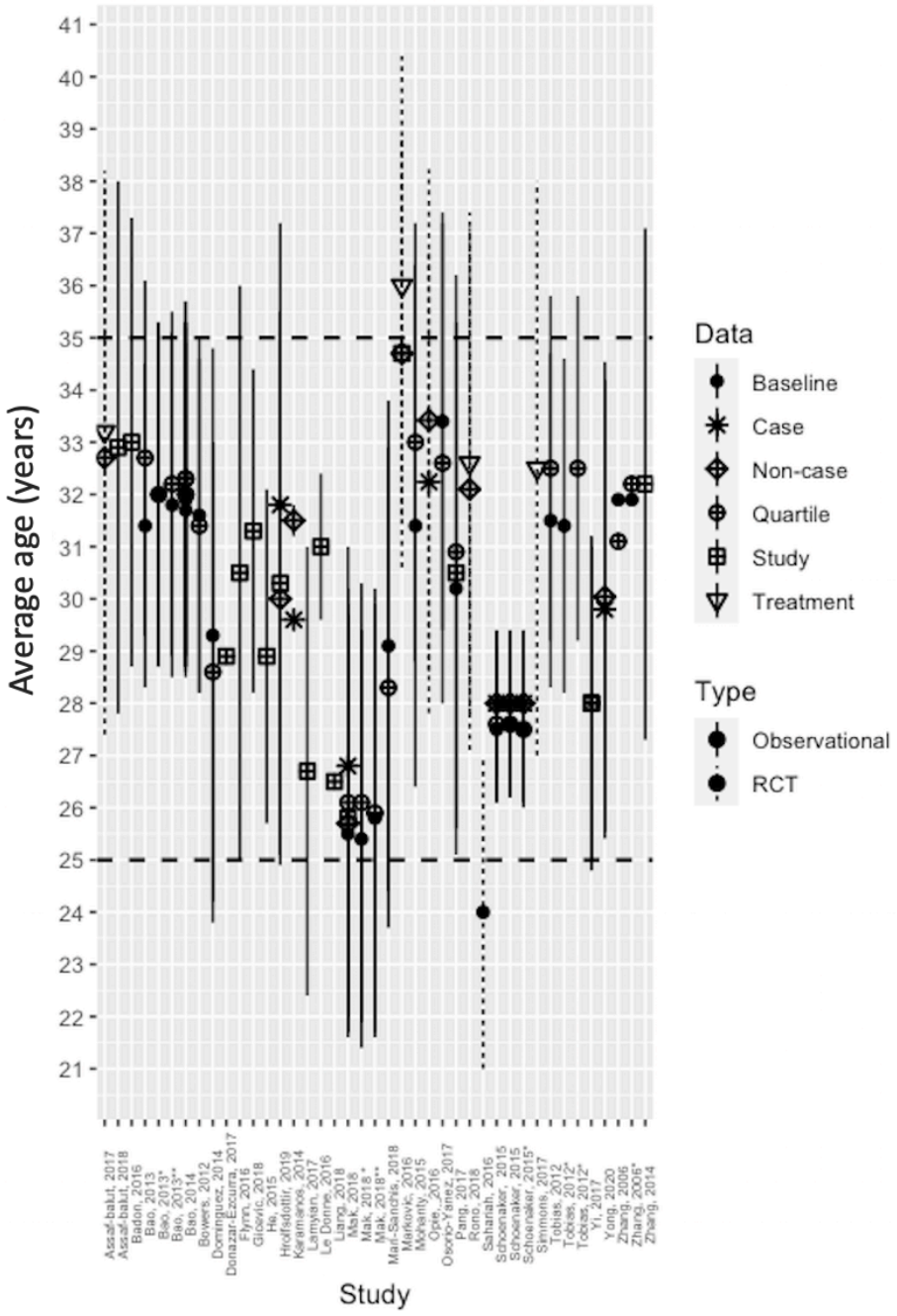



**Figure A.2: Plots showing average age of study participants.**

Plots of demographics of included studies. Dotted lines represent commonly cited cut-offs of GDM risk. 6 studies were excluded from this plot due to not including average age estimates. When more than one average was stated in a study all were plotted. * Represents studies with more than one exposure included in this review.

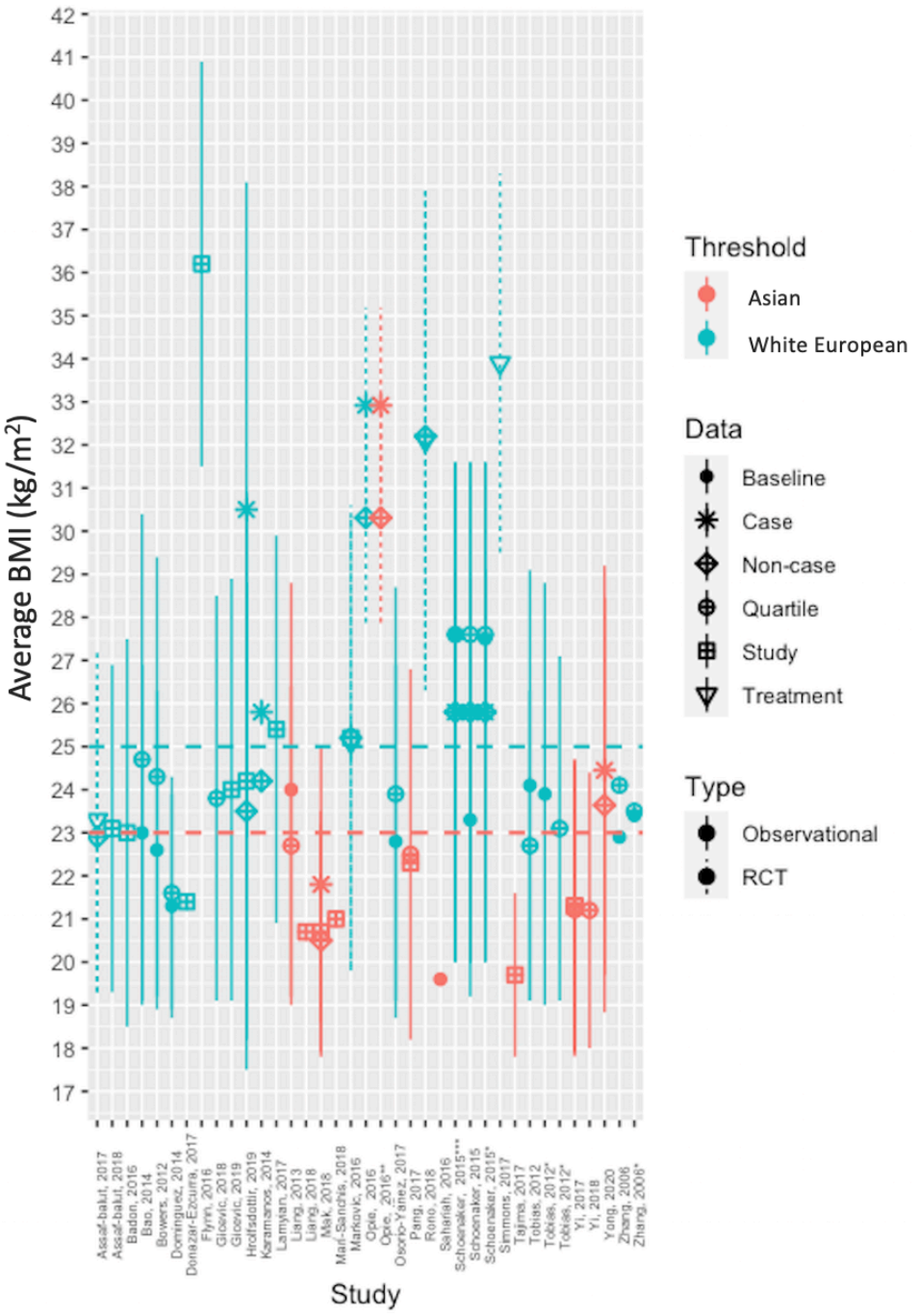



**Figure A.3: Plots showing average BMI of study participants.**

Dotted lines represent ethnic specific BMI cut-offs for overweight status. When more than one average was stated in a study all were plotted. * Represents studies with more than one exposure included in this review.

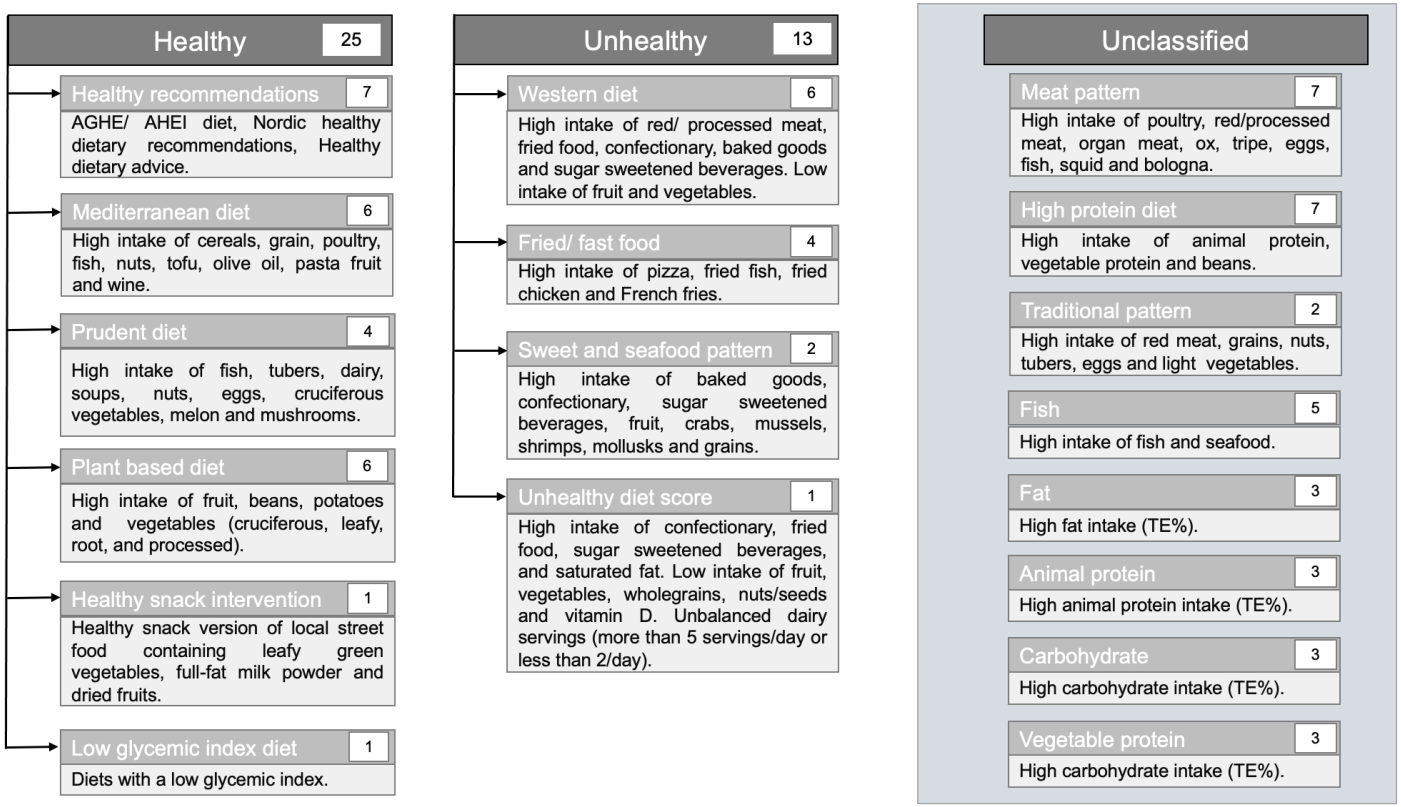



**Figure A.4: Schematic highlighting exposure characteristics.**

Schematic of the number of studies and characterises of each exposure. 2 overarching exposures, healthy (n=25, 6 subsets) and unhealthy (n=13, 4 subsets) were included in this review. Exposures which did not match to either the healthy or unhealthy pattern were classed as unclassified and were analysed independently. TE: Total Energy (%).

## Table A.1 : Summary of utilised GDM diagnostic criteria in observational studies.

**A**

| Diagnosis criteria | Australian Nationals | Mediterranean | Iranian | Asian | WEs |
|---|---|---|---|---|---|
| **Criteria stated** | 1 | 3 | 1 | 9 | 11 |
| **No criteria stated** | - | 2 | - | - | 5 |
| | **Chi squared p value = 0.73** | | | | |
| | | | | **Chi squared p value = 0.48** | |

**B**

| OGGT use | Australian Nationals | Mediterranean | Iranian | Asian | WEs |
|---|---|---|---|---|---|
| **OGGT used** | 1 | 1 | 1 | 9 | 8 |
| **No mention of OGGT** | - | 4 | - | - | 8 |
| | **Chi squared p value = 0.07** | | | | |
| | | | | **Chi squared p value = 0.16** | |

**A**: Categorisation of studies based on if GDM diagnostic criteria was stated. **B** Categorisation of studies based upon if specific OGTT cut-offs were stated. P values for the Chi-squared test to determine if significant differences between the observed and expected frequencies between all ethnicities and between the Asian and WE groups. OGTT: oral glucose tolerance test. WEs: White Europeans.

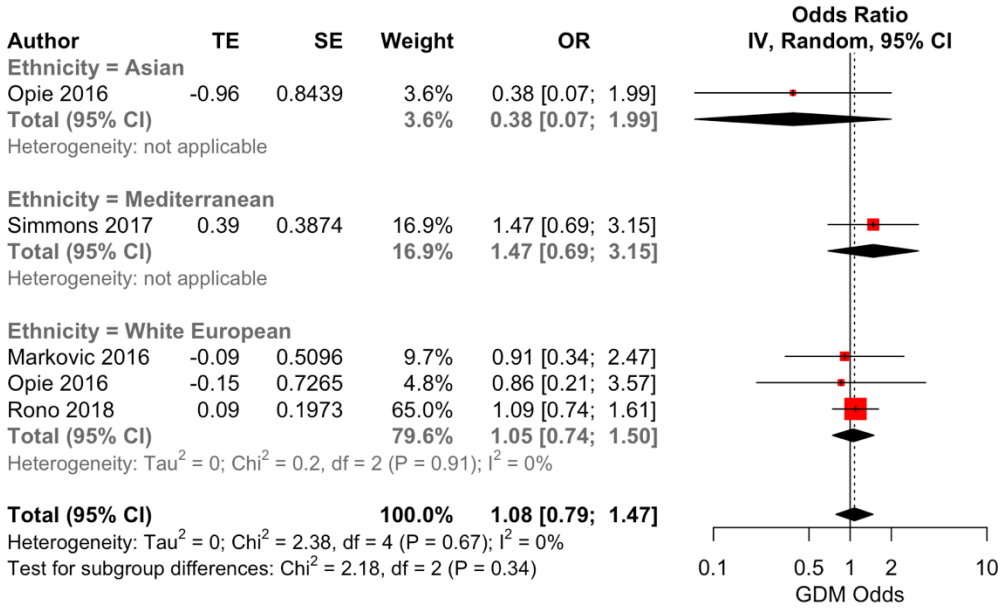
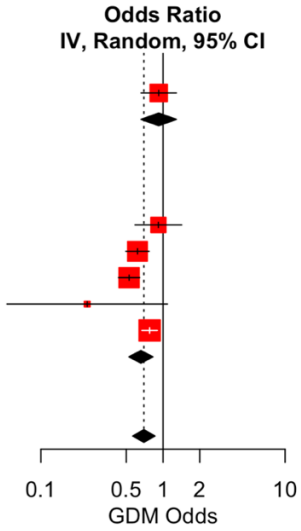



**Figure A.5: Forest plot of the effect of RCTs utilising healthy dietary recommendations on GDM.**

Associations between healthy dietary interventions and GDM in RCTs stratified by ethnicity subgroups using a DerSimonian and Laird (DL) random effects model. TE: treatment effect, SE: standard error, IV: inverse variance method, CI: Confidence interval.

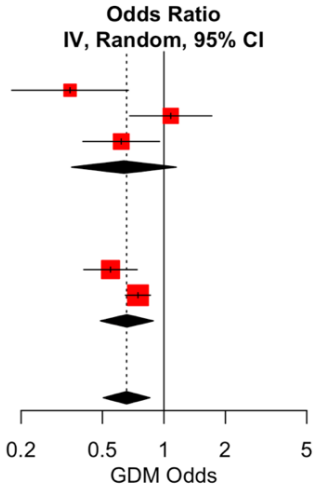
**A**

| Study or Subgroup | TE | SE | Weight | Odds Ratio IV, Random, 95% CI |
|---|---|---|---|---|
| **Ethnicity = Australian National** | | | | |
| Looman 2018 | -0.08 | 0.1697 | 16.6% | 0.92 [0.66; 1.29] |
| **Total (95% CI)** | | | **16.6%** | **0.92 [0.66; 1.29]** |
| Heterogeneity: not applicable | | | | |
| **Ethnicity = White European** | | | | |
| Badon 2016 | -0.09 | 0.2246 | 12.6% | 0.92 [0.59; 1.42] |
| Gicevic 2018 | -0.48 | 0.1121 | 21.6% | 0.62 [0.50; 0.77] |
| Tobias 2012 (aHEI) | -0.64 | 0.1016 | 22.6% | 0.53 [0.43; 0.64] |
| Tryggvadottir 2016 (HEI) | -1.43 | 0.7717 | 1.8% | 0.24 [0.05; 1.09] |
| Zhang 2014 | -0.25 | 0.0748 | 24.9% | 0.78 [0.67; 0.90] |
| **Total (95% CI)** | | | **83.4%** | **0.66 [0.53; 0.83]** |
| Heterogeneity: $Tau^2 = 0.0390$; $Chi^2 = 13.7$, df = 4 ($P < 0.01$); $I^2 = 71\%$ | | | | |
| **Total (95% CI)** | | | **100.0%** | **0.70 [0.56; 0.86]** |
| Heterogeneity: $Tau^2 = 0.0407$; $Chi^2 = 16.88$, df = 5 ($P < 0.01$); $I^2 = 70\%$ | | | | |
| Residual heterogeneity: $Tau^2$ = NA; $Chi^2 = 13.70$, df = 4 ($P < 0.01$); $I^2 = 71\%$ | | | | |



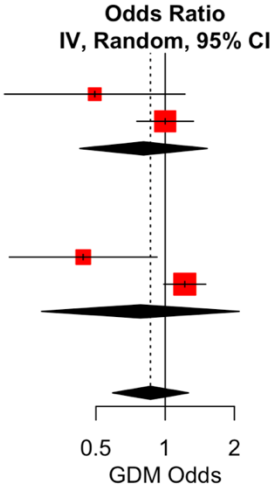

**B**

| Study or Subgroup | TE | SE | Weight | Odds Ratio IV, Random, 95% CI |
|---|---|---|---|---|
| **Ethnicity = Mediterranean** | | | | |
| Assaf-Balut 2018 | -1.06 | 0.3353 | 10.9% | 0.35 [0.18; 0.67] |
| Donazar-Ezcurra 2017 | 0.08 | 0.2360 | 16.6% | 1.08 [0.68; 1.72] |
| Karamanos 2014 | -0.48 | 0.2207 | 17.7% | 0.62 [0.40; 0.95] |
| **Total (95% CI)** | | | **45.2%** | **0.64 [0.35; 1.15]** |
| Heterogeneity: $Tau^2 = 0.2014$; $Chi^2 = 8.04$, df = 2 ($P = 0.02$); $I^2 = 75\%$ | | | | |
| **Ethnicity = White European** | | | | |
| Schoenaker 2015 | -0.60 | 0.1533 | 23.6% | 0.55 [0.41; 0.74] |
| Tobias 2012 (aMED) | -0.29 | 0.0719 | 31.2% | 0.75 [0.65; 0.86] |
| **Total (95% CI)** | | | **54.8%** | **0.66 [0.49; 0.89]** |
| Heterogeneity: $Tau^2 = 0.0335$; $Chi^2 = 3.33$, df = 1 ($P = 0.07$); $I^2 = 70\%$ | | | | |
| **Total (95% CI)** | | | **100.0%** | **0.66 [0.50; 0.85]** |
| Heterogeneity: $Tau^2 = 0.0522$; $Chi^2 = 11.41$, df = 4 ($P = 0.02$); $I^2 = 65\%$ | | | | |
| Residual heterogeneity: $Tau^2$ = NA; $Chi^2 = 11.38$, df = 3 ($P < 0.01$); $I^2 = 74\%$ | | | | |

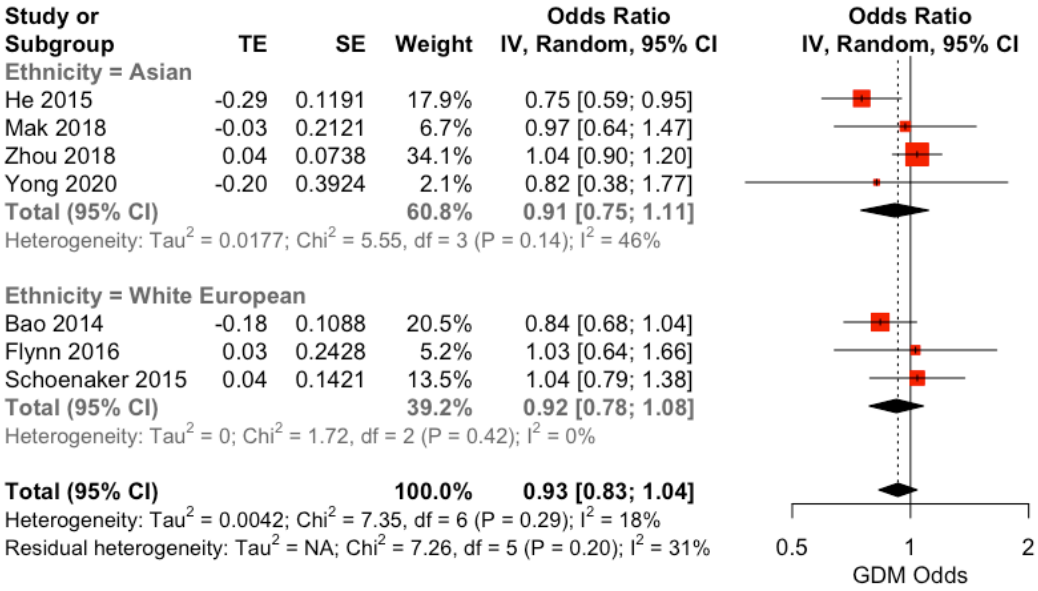



**C**

| Study or Subgroup | TE | SE | Weight | Odds Ratio IV, Random, 95% CI |
|---|---|---|---|---|
| **Ethnicity = Asian** | | | | |
| Yi 2017 | -0.71 | 0.4613 | 12.4% | 0.49 [0.20; 1.22] |
| He 2015 | 0.00 | 0.1446 | 34.1% | 1.00 [0.75; 1.33] |
| **Total (95% CI)** | | | **46.5%** | **0.81 [0.43; 1.52]** |
| Heterogeneity: $Tau^2 = 0.1319$; $Chi^2 = 2.13$, df = 1 ($P = 0.14$); $I^2 = 53\%$ | | | | |
| **Ethnicity = White European** | | | | |
| Tryggvadottir 2016 | -0.82 | 0.3774 | 16.2% | 0.44 [0.21; 0.92] |
| Zhang 2006 | 0.20 | 0.1071 | 37.3% | 1.22 [0.99; 1.50] |
| **Total (95% CI)** | | | **53.5%** | **0.78 [0.29; 2.10]** |
| Heterogeneity: $Tau^2 = 0.4408$; $Chi^2 = 6.73$, df = 1 ($P < 0.01$); $I^2 = 85\%$ | | | | |
| **Total (95% CI)** | | | **100.0%** | **0.86 [0.59; 1.26]** |
| Heterogeneity: $Tau^2 = 0.0887$; $Chi^2 = 10.00$, df = 3 ($P = 0.02$); $I^2 = 70\%$ | | | | |
| Residual heterogeneity: $Tau^2$ = NA; $Chi^2 = 8.86$, df = 2 ($P = 0.01$); $I^2 = 77\%$ | | | | |



**D**




## Figure A.6: Forest plot of the effect of healthy dietary exposures assessed in an observational setting on GDM.

Association*s* between subsets of the healthy dietary exposure and GDM assessed in observational setting utilising DerSimonian and Laird (DL) random effects models. **A**: Healthy recommendations. **B**: Mediterranean diet. **C**: Prudent diet. **D**: Plant-based pattern. TE: treatment effect, SE: standard error, IV: inverse variance method, CI: Confidence interval.

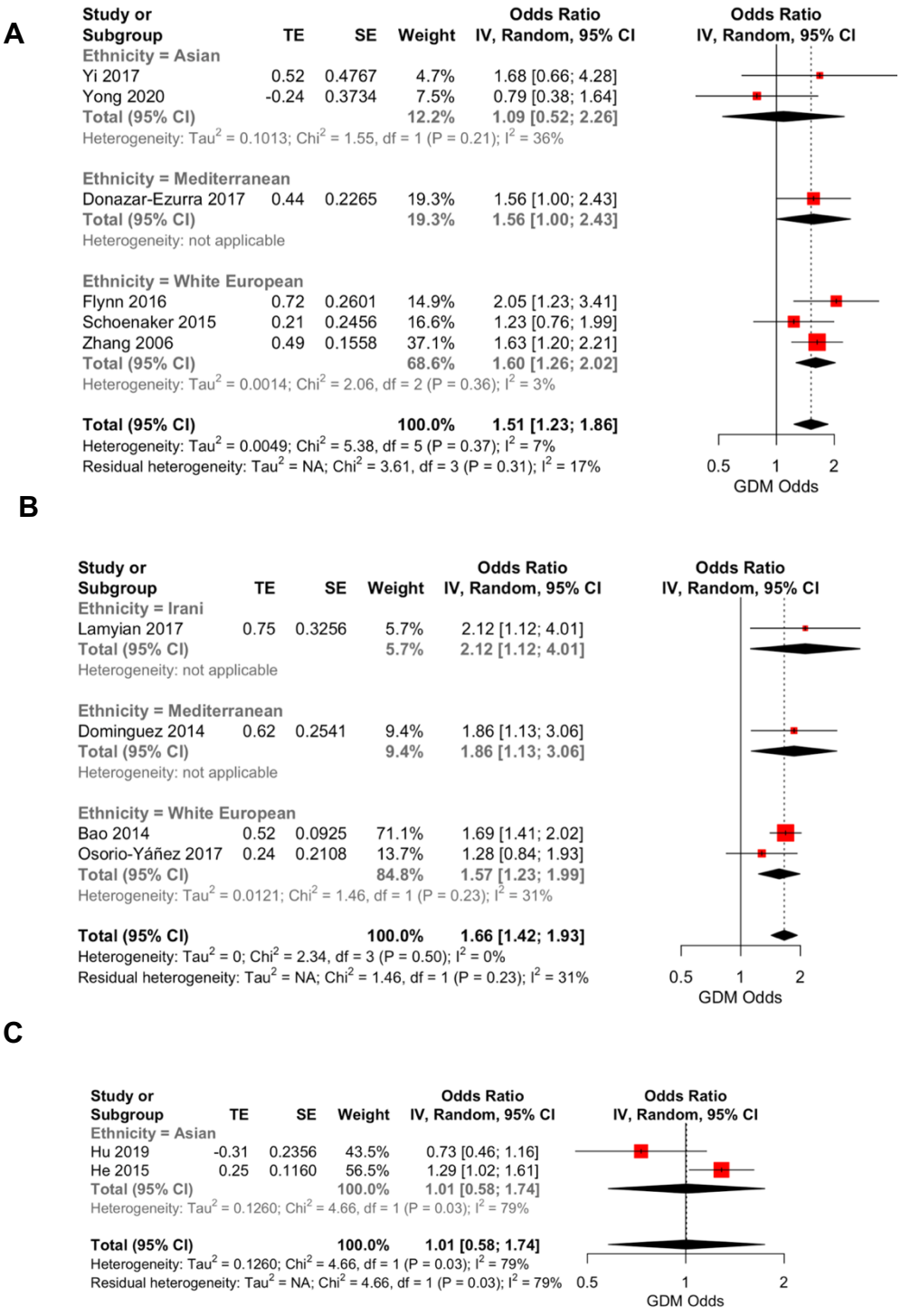
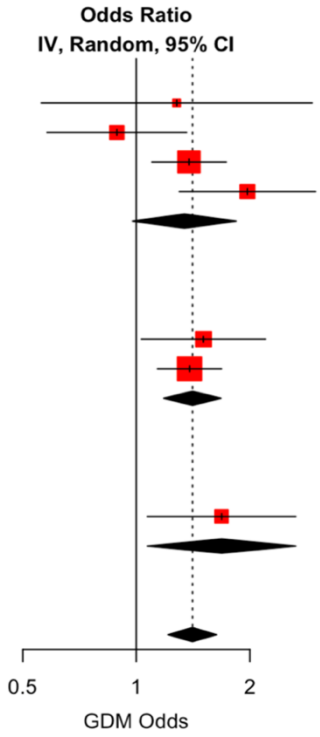



**Figure A.7: Forest plot of the effect of unhealthy dietary exposures assessed in an observational setting on GDM.**

Associations between subsets of the unhealthy dietary exposure and GDM assessed in observational setting utilising DerSimonian and Laird (DL) random effects models. **A**: Western diet. **B**: Fried/ fast food. **C**: Sweets and Seafood pattern.

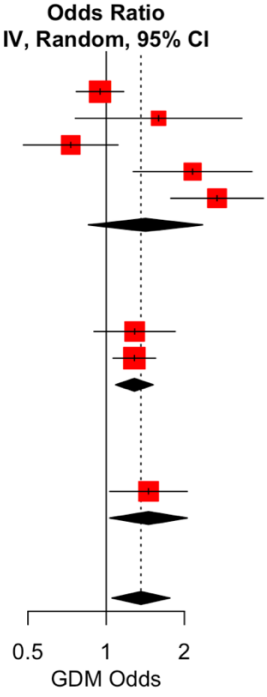
**A**

| Study or Subgroup | TE | SE | Weight | Odds Ratio IV, Random, 95% CI |
|---|---|---|---|---|
| **Ethnicity = Asian** | | | | |
| Liang 2018 | 0.25 | 0.4218 | 3.0% | 1.28 [0.56; 2.93] |
| Mak 2018 | -0.12 | 0.2174 | 10.1% | 0.89 [0.58; 1.36] |
| Pang 2017 | 0.32 | 0.1155 | 25.4% | 1.38 [1.10; 1.73] |
| Zhou 2018 | 0.68 | 0.2121 | 10.5% | 1.97 [1.30; 2.99] |
| **Total (95% CI)** | | | **48.9%** | **1.34 [0.98; 1.84]** |
| Heterogeneity: $Tau^2 = 0.0555$; $Chi^2 = 6.93$, df = 3 (P = 0.07); $I^2 = 57\%$ | | | | |
| **Ethnicity = White European** | | | | |
| Bao 2013 | 0.41 | 0.1937 | 12.2% | 1.51 [1.03; 2.20] |
| Bao 2014 | 0.32 | 0.0999 | 29.9% | 1.38 [1.14; 1.68] |
| **Total (95% CI)** | | | **42.0%** | **1.41 [1.18; 1.68]** |
| Heterogeneity: $Tau^2 = 0$; $Chi^2 = 0.15$, df = 1 (P = 0.70); $I^2 = 0\%$ | | | | |
| **Ethnicity = Mediterranean** | | | | |
| Marí-Sanchis 2018 | 0.52 | 0.2314 | 9.0% | 1.68 [1.07; 2.65] |
| **Total (95% CI)** | | | **9.0%** | **1.68 [1.07; 2.65]** |
| Heterogeneity: not applicable | | | | |
| **Total (95% CI)** | | | **100.0%** | **1.41 [1.22; 1.63]** |
| Heterogeneity: $Tau^2 = 0.0090$; $Chi^2 = 7.83$, df = 6 (P = 0.25); $I^2 = 23\%$ | | | | |
| Residual heterogeneity: $Tau^2$ = NA; $Chi^2 = 7.08$, df = 4 (P = 0.13); $I^2 = 44\%$ | | | | |

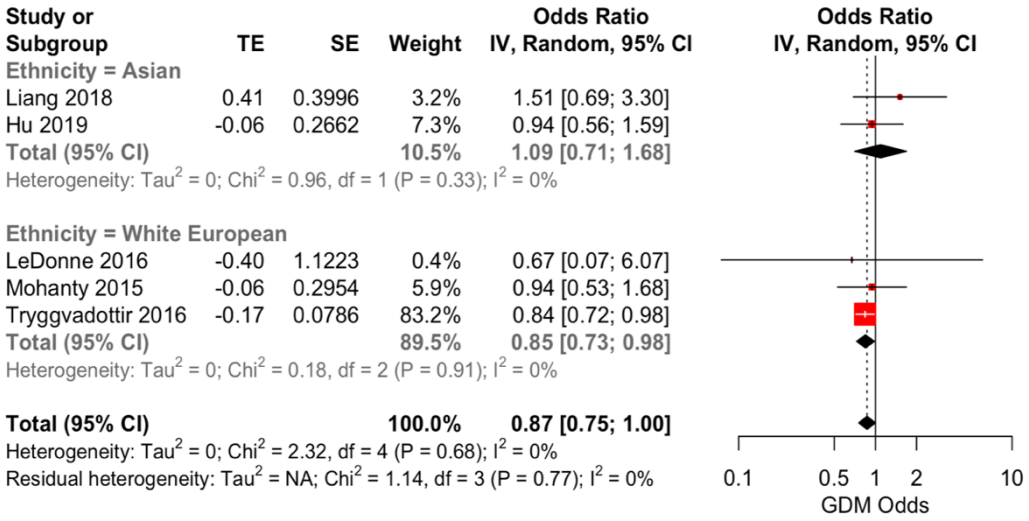



**B**

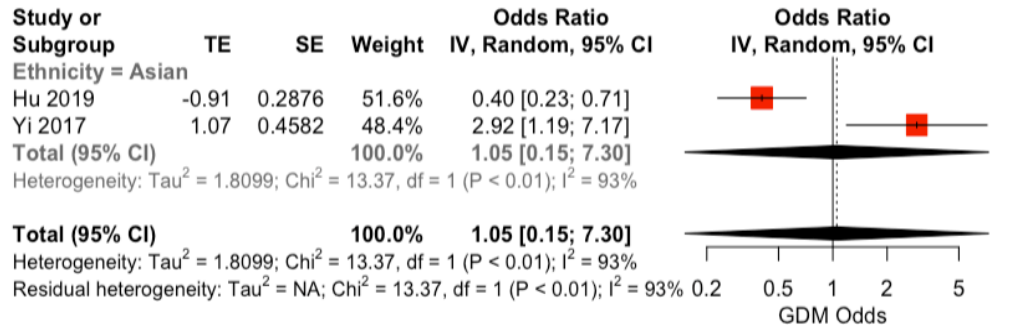
| Study or Subgroup | TE | SE | Weight | Odds Ratio IV, Random, 95% CI |
|---|---|---|---|---|
| **Ethnicity = Asian** | | | | |
| He 2015 | -0.05 | 0.1075 | 15.8% | 0.95 [0.77; 1.17] |
| Liang 2018 | 0.46 | 0.3766 | 7.2% | 1.59 [0.76; 3.33] |
| Mak 2018 | -0.31 | 0.2139 | 12.0% | 0.73 [0.48; 1.11] |
| Pang 2017 | 0.77 | 0.2686 | 10.1% | 2.15 [1.27; 3.64] |
| Zhou 2018 | 0.98 | 0.2093 | 12.2% | 2.67 [1.77; 4.02] |
| **Total (95% CI)** | | | **57.3%** | **1.42 [0.85; 2.35]** |
| Heterogeneity: $Tau^2 = 0.2752$; $Chi^2 = 30.08$, df = 4 (P < 0.01); $I^2 = 87\%$ | | | | |
| **Ethnicity = White European** | | | | |
| Bao 2013 | 0.25 | 0.1836 | 13.1% | 1.29 [0.90; 1.84] |
| Bao 2014 | 0.25 | 0.0963 | 16.1% | 1.28 [1.06; 1.55] |
| **Total (95% CI)** | | | **29.3%** | **1.28 [1.09; 1.52]** |
| Heterogeneity: $Tau^2 = 0$; $Chi^2 = 0$, df = 1 (P = 0.99); $I^2 = 0\%$ | | | | |
| **Ethnicity = Australian National** | | | | |
| Looman 2018 | 0.37 | 0.1752 | 13.4% | 1.45 [1.03; 2.05] |
| **Total (95% CI)** | | | **13.4%** | **1.45 [1.03; 2.05]** |
| Heterogeneity: not applicable | | | | |
| **Total (95% CI)** | | | **100.0%** | **1.36 [1.05; 1.76]** |
| Heterogeneity: $Tau^2 = 0.0968$; $Chi^2 = 31.49$, df = 7 (P < 0.01); $I^2 = 78\%$ | | | | |
| Residual heterogeneity: $Tau^2$ = NA; $Chi^2 = 30.08$, df = 5 (P < 0.01); $I^2 = 83\%$ | | | | |



**C**




**D**

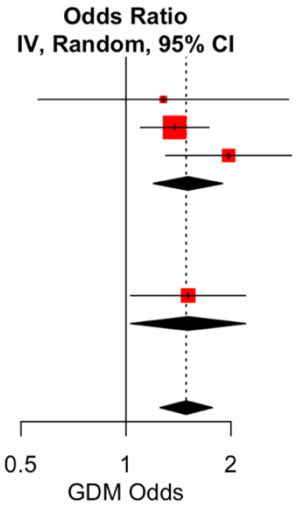



## Figure A.8: Forest plot of the effect of unclassified dietary exposures assessed in an observational setting on GDM.

Associations between unclassified dietary patterns assessed in an observational setting and GDM in an observational setting utilising DerSimonian and Laird (DL) random effects models: **A**: Meat pattern. **B**: High-protein pattern. **C**: Fish. **D**: Traditional Asian.

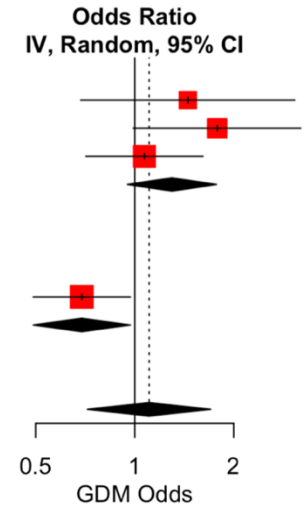
**A**

| Study or Subgroup | TE | SE | Weight | Odds Ratio IV, Random, 95% CI |
|---|---|---|---|---|
| **Ethnicity = Asian** | | | | |
| Liang 2018 | 0.25 | 0.4218 | 4.3% | 1.28 [0.56; 2.93] |
| Pang 2017 | 0.32 | 0.1155 | 57.9% | 1.38 [1.10; 1.73] |
| Zhou 2018 | 0.68 | 0.2121 | 17.2% | 1.97 [1.30; 2.99] |
| **Total (95% CI)** | | | **79.4%** | **1.51 [1.20; 1.89]** |
| Heterogeneity: $Tau^2 = 0.0071$; $Chi^2 = 2.31$, df = 2 (P = 0.32); $I^2 = 13\%$ | | | | |
| **Ethnicity = White European** | | | | |
| Bao 2013 | 0.41 | 0.1937 | 20.6% | 1.51 [1.03; 2.20] |
| **Total (95% CI)** | | | **20.6%** | **1.51 [1.03; 2.20]** |
| Heterogeneity: not applicable | | | | |
| **Total (95% CI)** | | | **100.0%** | **1.49 [1.25; 1.77]** |
| Heterogeneity: $Tau^2 = 0$; $Chi^2 = 2.31$, df = 3 (P = 0.51); $I^2 = 0\%$ | | | | |
| Residual heterogeneity: $Tau^2$ = NA; $Chi^2 = 2.31$, df = 2 (P = 0.32); $I^2 = 13\%$ | | | | |



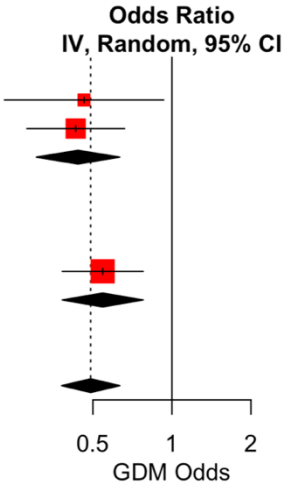

**B**

| Study or Subgroup | TE | SE | Weight | Odds Ratio IV, Random, 95% CI |
|---|---|---|---|---|
| **Ethnicity = Asian** | | | | |
| Liang 2018 | 0.37 | 0.3829 | 17.8% | 1.45 [0.68; 3.07] |
| Pang 2017 | 0.58 | 0.2993 | 22.5% | 1.78 [0.99; 3.20] |
| Zhou 2018 | 0.07 | 0.2093 | 28.6% | 1.07 [0.71; 1.61] |
| **Total (95% CI)** | | | **68.9%** | **1.30 [0.95; 1.77]** |
| Heterogeneity: $Tau^2 = 0.0020$; $Chi^2 = 2.05$, df = 2 (P = 0.36); $I^2 = 2\%$ | | | | |
| **Ethnicity = White European** | | | | |
| Bao 2013 | -0.37 | 0.1738 | 31.1% | 0.69 [0.49; 0.97] |
| **Total (95% CI)** | | | **31.1%** | **0.69 [0.49; 0.97]** |
| Heterogeneity: not applicable | | | | |
| **Total (95% CI)** | | | **100.0%** | **1.11 [0.72; 1.70]** |
| Heterogeneity: $Tau^2 = 0.1262$; $Chi^2 = 9.27$, df = 3 (P = 0.03); $I^2 = 68\%$ | | | | |
| Residual heterogeneity: $Tau^2$ = NA; $Chi^2 = 2.05$, df = 2 (P = 0.36); $I^2 = 2\%$ | | | | |

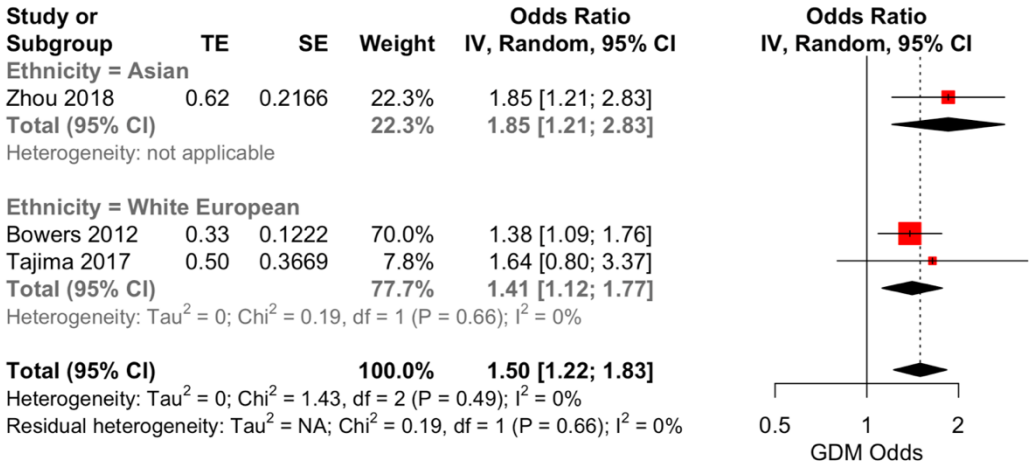



**C**

| Study or Subgroup | TE | SE | Weight | Odds Ratio IV, Random, 95% CI |
|---|---|---|---|---|
| **Ethnicity = Asian** | | | | |
| Tajima 2017 | -0.77 | 0.3564 | 13.3% | 0.46 [0.23; 0.93] |
| Zhou 2018 | -0.84 | 0.2187 | 35.3% | 0.43 [0.28; 0.66] |
| **Total (95% CI)** | | | **48.5%** | **0.44 [0.30; 0.63]** |
| Heterogeneity: $Tau^2 = 0$; $Chi^2 = 0.03$, df = 1 (P = 0.86); $I^2 = 0\%$ | | | | |
| **Ethnicity = White European** | | | | |
| Looman 2018 | -0.61 | 0.1810 | 51.5% | 0.54 [0.38; 0.78] |
| **Total (95% CI)** | | | **51.5%** | **0.54 [0.38; 0.78]** |
| Heterogeneity: not applicable | | | | |
| **Total (95% CI)** | | | **100.0%** | **0.49 [0.38; 0.63]** |
| Heterogeneity: $Tau^2 = 0$; $Chi^2 = 0.73$, df = 2 (P = 0.70); $I^2 = 0\%$ | | | | |
| Residual heterogeneity: $Tau^2$ = NA; $Chi^2 = 0.03$, df = 1 (P = 0.86); $I^2 = 0\%$ | | | | |



**D**




## Figure A.9: Forest plot of the effect of macronutrient dietary exposures assessed in an observational setting on GDM.

Associations between macronutrient-defined diets and GDM in an observational setting utilising DerSimonian and Laird (DL) random effects models: **A**: Animal protein. **B**: Vegetable protein. **C**: Carbohydrate **D**: Fat. All macronutrient exposures were assessed as a % of total energy intake (TE).

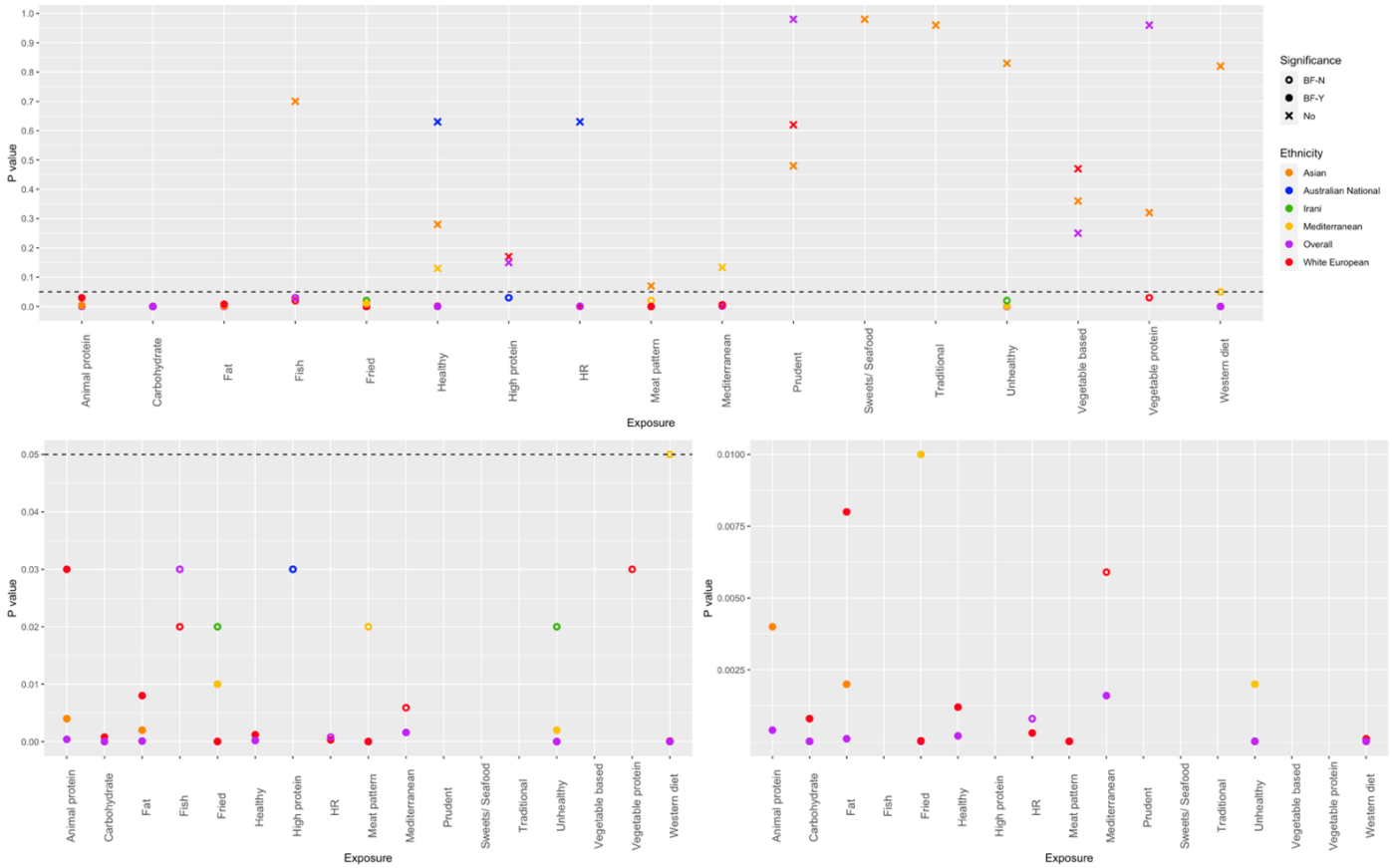



**Figure A.10: Significance of observational results following a Bonferroni correction.**

**Top**: P values from observational analyses and their significance after a Bonferroni correction. **Bottom left**: P value ≤ 0.05. **Bottom right**: P value ≤ 0.01. BF-Y: Significant following a Bonferroni correction. BF-N: Non-significant following a Bonferroni correction.

## Table A.3 : Random effects meta-analysis: HKSJ approach.

| | | | | Random effects | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | D-L | | | HKSJ | | |
| **Exposure** | **Study type** | **Subclass** | **N** | **OR (95% CI)** | **P value** | **Tau²** | **OR (95% CI)** | **P value** | **Tau²** |
| **Healthy** | RCT | Overall | 7* | 0.86 (0.69 - 1.08) | 0.19 | 0.01 | 0.88 (0.65 - 1.18) | 0.32 | 0.06 |
| | | Asian | 2 | 0.76 (0.48 - 1.20) | 0.24 | 0 | 0.72 (0.03 - 20.6) | 0.43 | 0.79 |
| | | Australian National | 1 | 0.91 (0.34 - 2.47) | 0.85 | NA | 0.91 (0.34 - 2.47) | 0.85 | NA |
| | | Mediterranean | 2 | 0.92 (0.43 - 1.96) | 0.83 | 0.22 | 0.92 (0.01 - 123.9) | 0.86 | 0.20 |
| | | White European | 2 | 0.69 (0.24 - 1.98) | 0.50 | 0.46 | 0.70 (0.01 - 630) | 0.63 | 0.41 |
| | Observational | Overall | 22*** | 0.79 (0.70 - 0.89) | ≤0.001 | 0.05 | 0.78 (0.67 - 0.90) | ≤0.01 | 0.09 |
| | | Asian | 6* | 0.91 (0.42 - 1.53) | 0.28 | 0.01 | 0.91 (0.73 - 1.11) | 0.26 | 0.03 |
| | | Australian National | 1 | 0.92 (0.66 - 1.29) | 0.63 | NA | 0.92 (0.66 - 1.29) | 0.63 | NA |
| | | Mediterranean | 3 | 0.64 (0.35 - 1.15) | 0.14 | 0.20 | 0.64 (0.16 - 2.54) | 0.29 | 0.24 |
| | | White European | 12** | 0.76 (0.64 - 0.89) | ≤0.001 | 0.06 | 0.75 (0.61 - 0.93) | 0.01 | 0.12 |
| | Combined | Overall | 29*** | 0.75 (0.66 - 0.86) | ≤0.0001 | 0.05 | 0.80 (0.70 - 0.90) | 0.007 | 0.08 |
| | | Asian | 8* | 0.90 (0.78 - 1.04) | 0.18 | 0.01 | 0.87 (0.73 - 1.05) | 0.13 | 0.04 |
| | | Australian National | 2 | 0.92 (0.67 - 1.26) | 0.60 | 0.00 | 0.92 (0.88 - 0.97) | 0.03 | 0 |
| | | Mediterranean | 5 | 0.75 (0.50 - 1.07) | 0.11 | 0.12 | 0.74 (0.39 - 1.40) | 0.25 | 0.21 |
| | | White European | 14** | 0.77 (0.65 - 0.90) | ≤0.001 | 0.06 | 0.72 (0.58 - 0.90) | ≤0.01 | 0.11 |
| **HR** | RCT | Overall | 4 | 0.62 (0.29 - 1.30) | 0.20 | 0.35 | 0.60 (0.17 - 2.15) | 0.29 | 0.46 |
| | | Asian | 1 | 0.16 (0.03 - 0.86) | 0.03 | NA | 0.16 (0.03 - 0.86) | 0.03 | NA |
| | | White European | 3 | 0.79 (0.42 - 1.50) | 0.47 | 0.18 | 0.78 (0.19 - 3.17) | 0.53 | 0.19 |
| | Observational | Overall | 7 | 0.72 (0.60 - 0.86) | ≤0.001 | 0.04 | 0.72 (0.55 - 0.94) | 0.02 | 0.09 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Australian National | 1 | 0.92 (0.66 - 1.29) | 0.63 | NA | 0.92 (0.66 - 1.29) | 0.63 | NA |
| | | White European | 6 | 0.69 (0.57 - 0.84) | ≤0.001 | 0.04 | 0.69 (0.50 - 0.94) | 0.03 | 0.10 |
| | Combined | Overall | 11 | 0.73 (0.61 - 0.87) | ≤0.001 | 0.05 | 0.67 (0.52 - 0.94) | 0.02 | 0.19 |
| | | Australian National | 1 | 0.92 (0.67 - 1.26) | 0.63 | NA | 0.92 (0.67 - 1.26) | 0.63 | NA |
| | | Asian | 1 | 0.16 (0.031 - 0.86) | 0.03 | NA | 0.16 (0.031 - 0.86) | 0.03 | NA |
| | | White European | 9 | 0.72 (0.60 - 0.87) | ≤0.001 | 0.04 | 0.71 (0.54 - 0.94) | 0.02 | 0.11 |
| **Mediterranean** | Observational | Overall | 5 | 0.66 (0.50 - 0.85) | ≤0.01 | 0.05 | 0.65 (0.41 - 1.036) | 0.06 | 0.11 |
| | | Mediterranean | 3 | 0.64 (0.35 - 1.15) | 0.14 | 0.20 | 0.64 (0.16 - 2.54) | 0.29 | 0.24 |
| | | White European | 2 | 0.66 (0.49 - 0.89) | ≤0.01 | 0.04 | 0.66 (0.10 - 4.55) | 0.22 | 0.03 |
| | Combined | Overall | 6 | 0.66 (0.54 - 0.82) | ≤0.0001 | 0.0344 | 0.65 (0.46 - 0.92) | 0.03 | 0.08 |
| | | Mediterranean | 4 | 0.66 (0.45 - 0.95) | 0.03 | 0.0811 | 0.65 (0.32 - 1.31) | 0.15 | 0.14 |
| | | White European | 2 | 0.66 (0.49 - 0.89) | 0.01 | 0.0335 | 0.66 (0.10 - 4.55) | 0.22 | 0.66 |
| **Prudent diet** | Observational | Overall | 4 | 0.86 (0.59 - 1.26) | 0.39 | 0.0887 | 0.80 (0.36 - 1.77) | 0.44 | 0.18 |
| | | Asian | 2 | 0.81 (0.43 - 1.52) | 0.50 | 0.132 | 0.80 (0.01 - 51.50) | 0.63 | 0.81 |
| | | White European | 2 | 0.78 (0.29 - 2.10) | 0.61 | 0.441 | 0.77 (0 - 486.71) | 0.70 | 0.77 |
| **Plant-based diet** | Observational | Overall | 6 | 0.93 (0.82 - 1.06) | 0.30 | 0.0086 | 0.93 (0.79 - 1.10) | 0.32 | 0.01 |
| | | Asian | 3 | 0.91 (0.73 - 1.15) | 0.45 | 0.0253 | 0.92 (0.58 - 1.45) | 0.50 | 0.02 |
| | | White European | 3 | 0.94 (0.79 - 1.11) | 0.46 | 0.0028 | 0.95 (0.65 - 1.39) | 0.60 | 0.01 |
| **Unhealthy diets** | Observational | Overall | 13 | 1.44 (1.25 - 1.67) | ≤0.0001 | 0.0257 | 1.43 (1.20 - 1.72) | ≤0.001 | 0.06 |
| | | Asian | 4 | 1.04 (0.72 - 1.51) | 0.83 | 0.0701 | 1.04 (0.59 - 1.83) | 0.84 | 0.08 |
| | | Iranian | 1 | 2.12 (1.12 - 4.01) | 0.02 | NA | 2.12 (1.12 - 4.01) | 0.02 | NA |
| | | Mediterranean | 2 | 1.69 (1.21 - 2.35) | ≤0.01 | 0 | 1.68 (0.54 - 5.24) | 0.11 | 0.002 |
| | | White European | 6 | 1.59 (1.41 - 1.81) | ≤0.0001 | 0 | 1.57 (1.33 - 1.84) | ≤0.001 | 0.01 |
| **Fried/ fast food** | Observational | Overall | 4 | 1.66 (1.42 - 1.93) | ≤0.0001 | 0 | 1.65 (1.24 - 2.19) | ≤0.01 | 0.02 |
| | | Iranian | 1 | 2.12 (1.12 - 4.01) | 0.02 | NA | 2.12 (1.12 - 4.01) | 0.02 | NA |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Mediterranean | 1 | 1.86 (1.13 - 3.06) | ≤0.01 | NA | 1.86 (1.13 - 3.06) | ≤0.01 | NA |
| | | White European | 2 | 1.57 (1.23 - 1.99) | 0.0004 | 0.01 | 1.55 (0.31 - 7.82) | 0.18 | 0.02 |
| **Western diet** | Observational | Overall | 7 | 1.51 (1.23 - 1.86) | ≤0.0001 | 0.0049 | 1.48 (1.09 - 2.01) | 0.02 | 0.05 |
| | | Asian | 2 | 1.09 (0.52 - 2.26) | 0.82 | 0.10 | 1.09 (0.01 - 129.8) | 0.86 | 0.13 |
| | | Mediterranean | 1 | 1.56 (1.00 - 2.43) | 0.05 | NA | 1.55 (1.00 - 2.42) | 0.05 | NA |
| | | White European | 3 | 1.60 (1.26 - 2.02) | ≤0.0001 | 0.0014 | 1.60 (0.90 - 2.82) | 0.07 | 0.03 |
| **Sweets and Seafood pattern** | Observational | Asian | 2 | 1.01 (0.58 - 1.74) | 0.98 | 0.13 | 1.01 (0.03 - 34.14) | 0.98 | 0.18 |
| **Traditional Asian** | Observational | Asian | 2 | 1.05 (0.15 - 7.30) | 0.96 | 1.81 | 1.05 (0 - 301,857) | 0.97 | 1.71 |
| **Meat pattern** | Observational | Overall | 7 | 1.41(1.22 - 1.63) | ≤0.0001 | 0.01 | 1.41 (1.14 - 1.75) | ≤0.01 | 0.03 |
| | | Asian | 4 | 1.34 (0.98 - 1.84) | 0.07 | 0.06 | 1.34 (0.79 - 2.23) | 0.18 | 0.07 |
| | | Mediterranean | 1 | 1.68 (1.07 - 2.65) | 0.02 | NA | 1.68 (1.07 - 2.65) | 0.02 | NA |
| | | White European | 2 | 1.41(1.18 - 1.68) | ≤0.0001 | 0 | 1.40 (0.88 - 2.24) | 0.07 | 0.0003 |
| **High protein diet** | Observational | Overall | 7 | 1.19 (0.94 - 1.52) | 0.15 | 0.06 | 1.20 (0.88 - 1.65) | 0.20 | 0.08 |
| | | Asian | 5 | 1.13 (0.82 - 1.56) | 0.45 | 0.09 | 1.15 (0.68 - 1.93) | 0.50 | 0.36 |
| | | Australian National | 1 | 1.45(1.03 - 2.05) | 0.18 | NA | 1.45 (1.03 - 2.04) | 0.18 | NA |
| | | White European | 1 | 1.28 (0.9 - 1.84) | 0.17 | NA | 1.28 (0.90 - 1.84) | 0.17 | NA |
| **Animal protein** | Observational | Overall | 4 | 1.49 (1.25 - 1.77) | < 0.0001 | 0 | 1.51 (1.16 - 1.97) | 0.02 | 0.01 |
| | | Asian | 3 | 1.51 (1.20 - 1.89) | ≤0.001 | ≤0.01 | 1.53 (0.91 - 2.57) | 0.07 | 0.02 |
| | | White European | 1 | 1.51 (1.03 - 2.20) | 0.03 | NA | 1.51 (1.03 - 2.20) | 0.03 | NA |
| **Vegetable protein** | Observational | Overall | 4 | 1.11 (0.72 - 1.70) | 0.67 | 0.12 | 1.10 (0.56 - 2.16) | 0.67 | 0.12 |
| | | Asian | 3 | 1.30 (0.95 - 1.77) | 0.10 | 0.0021 | 1.33 (0.67 - 2.63) | 0.22 | 0.03 |
| | | White European | 1 | 0.69 (0.49 - 0.97) | 0.03 | NA | 0.69 (0.97 - 90.49) | 0.03 | NA |
| **Fat** | Observational | Overall | 3 | 1.50 (1.22 - 1.83) | ≤0.0001 | 0 | 1.53 (1.04 - 2.25) | 0.04 | 0.01 |
| | | Asian | 1 | 1.85 (1.21 - 2.83) | ≤0.01 | NA | 1.86 (1.22 - 2.84) | ≤0.01 | NA |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | White European | 2 | 1.41 (1.12 - 1.77) | ≤0.01 | 0 | 1.42 (0.73 - 2.76) | 0.20 | 0.04 |
| **Carbohydrate** | Observational | Overall | 3 | 0.49 (0.38 - 0.63) | ≤0.0001 | 0 | 0.49 (0.35 - 0.68) | 0.01 | 0.0026 |
| | | Asian | 2 | 0.44 (0.30 - 0.63) | ≤0.0001 | 0 | 0.44 (0.30 - 0.65) | 0.02 | <0.0001 |
| | | White European | 1 | 0.54 (0.38 -0.78) | ≤0.001 | NA | 0.54 (0.38 - 0.77) | ≤0.001 | NA |
| **Fish** | Observational | Overall | 5 | 0.87 (0.75 - 1.00) | 0.05 | 0 | 0.92 (0.73 - 1.15) | 0.35 | 0.02 |
| | | Asian | 2 | 1.09 (0.71 - 1.68) | 0.70 | 0 | 1.11 (0.06 - 19.30) | 0.72 | 0.04 |
| | | White European | 3 | 0.85 (0.75 - 1.00) | 0.03 | 0 | 0.85 (0.77 - 0.94) | 0.02 | 0.0014 |

NA represents ethnic subgroup with only one study meaning a random effects meta-analysis was not required. DL: DerSimonian and Laird HKSJ: Hartung-Knapp-Sidik-Jonkman HR: Healthy recommendations

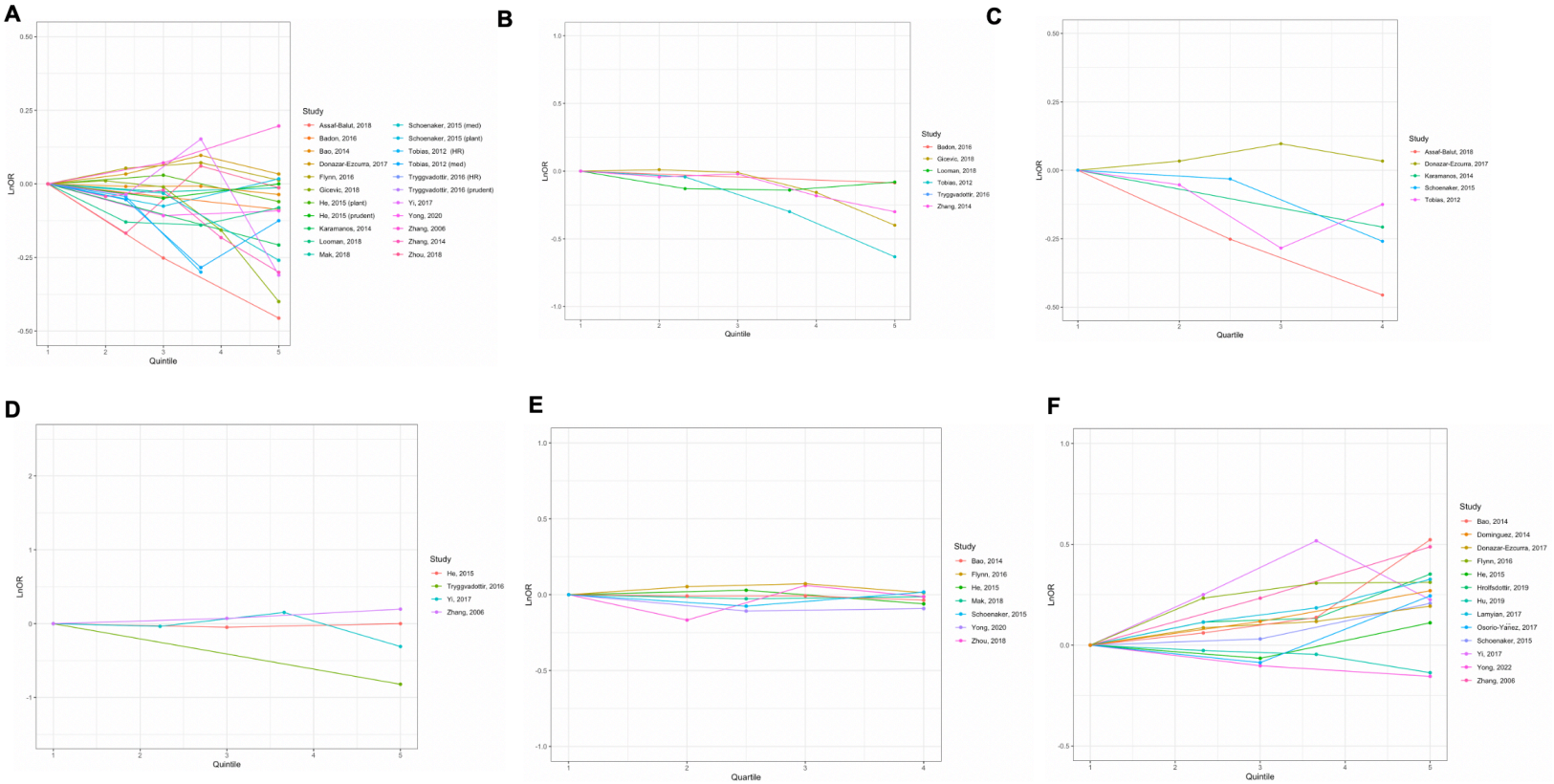



**Figure A.11: Exposures indicating trends between quantiles of adherence in observational studies.**

**A**: Healthy diets. **B**: Healthy recommendations. **C**: Mediterranean diet **D**: Prudent. **E:** Plant-based pattern. **F:** Unhealthy diets.

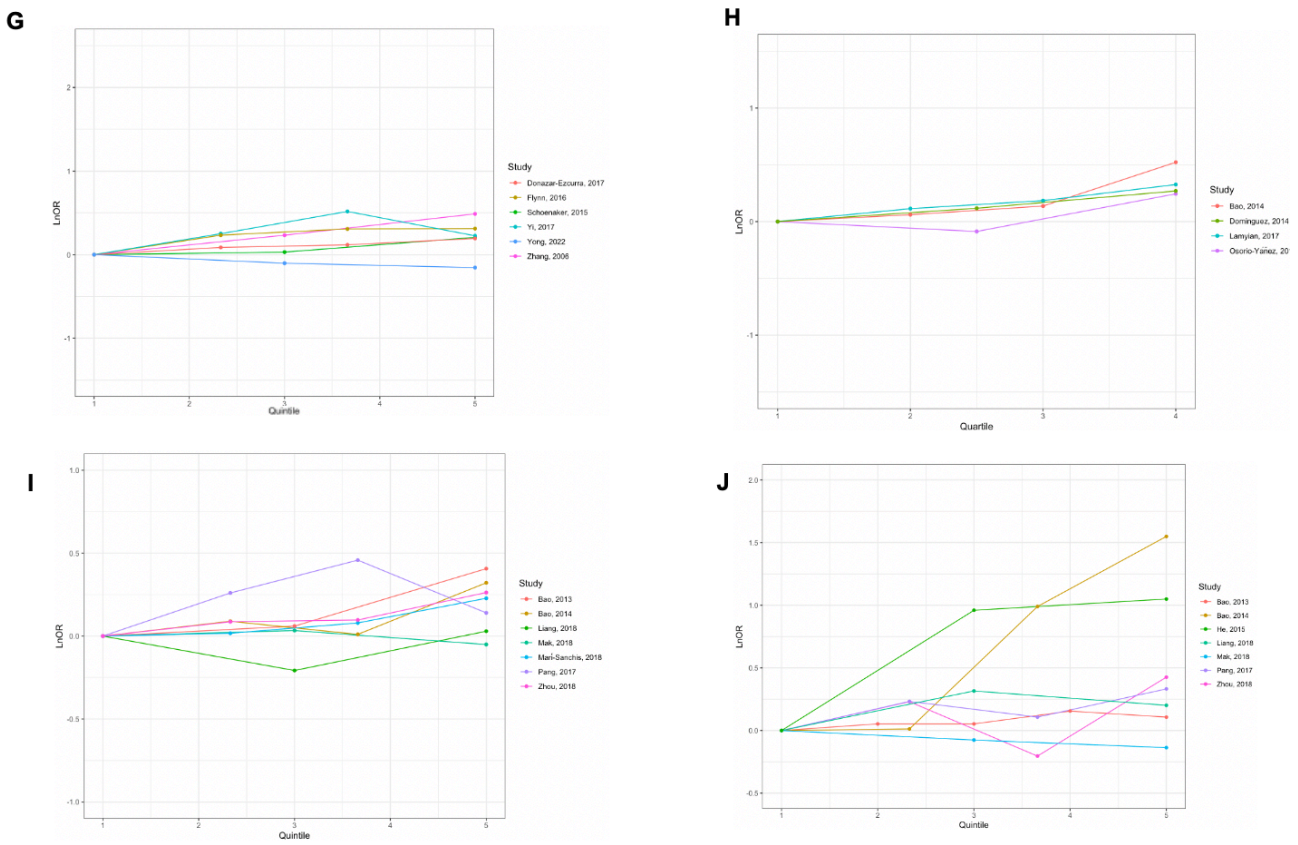



**Figure A.11 (cont): Exposures indicating trends between quantiles of adherence in observational studies.**
**G:** Western diet. **H:** Fried/fast food. **I:** Meat pattern. **J:** High-protein diets.

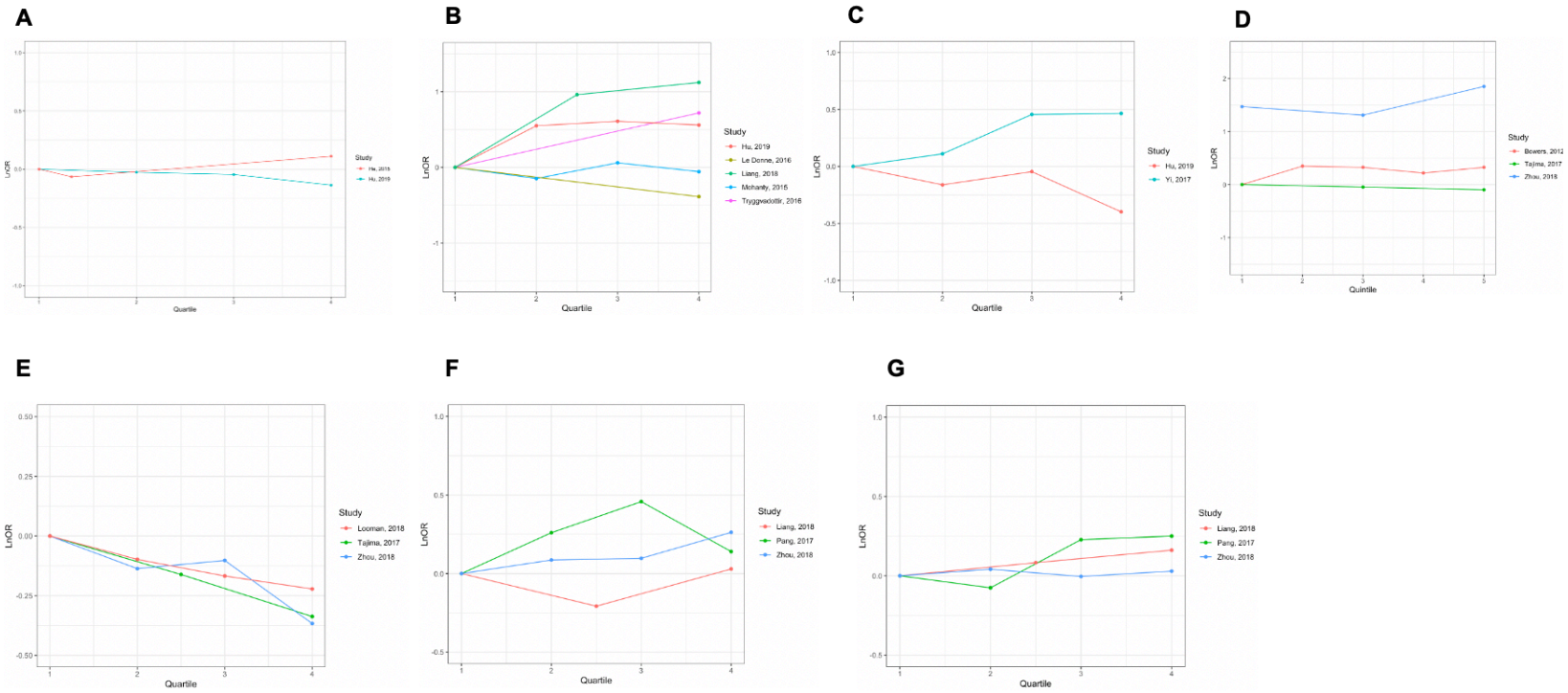



**Figure A.12: Exposures with no indication trends between quantiles of adherence in observational studies.**

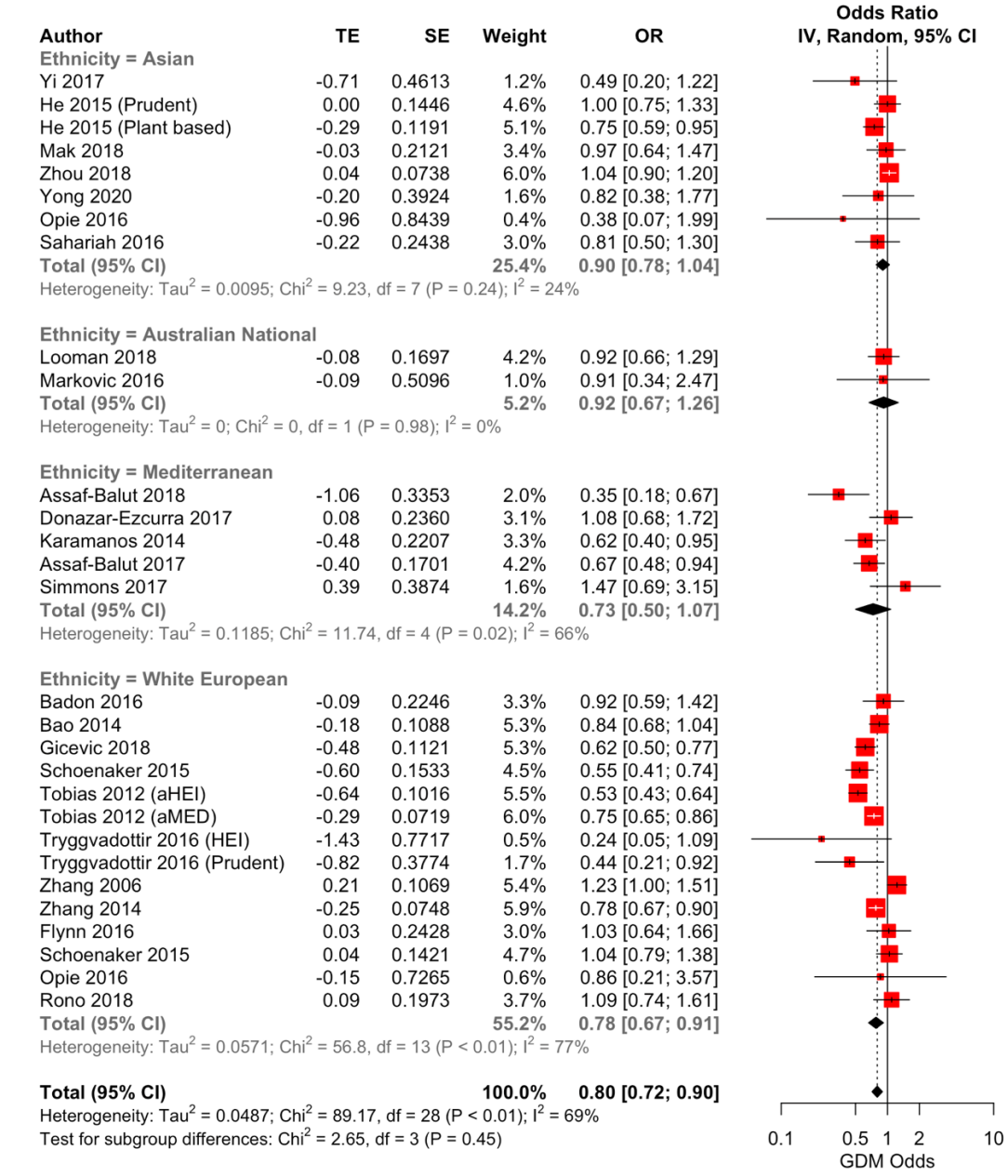
Exposures with no apparent trends or missing case/control data for all quantiles meaning dose response relationships were not investigated. **A:** Sweet and seafood**. B:** Fish **C:** Traditional Asian **D:** Fat **E:** Carbohydrate**. F:** Animal protein **G:** Vegetable protein.

## Table A.3 : Dose response analyses.

| | | | Linear | Quadratic | Cubic Spline - 3K | | Cubic Spline - 4K | |
|---|---|---|---|---|---|---|---|---|
| **Exposure** | **Unit** | **Subclass** | **LogLik** | **LogLik** | **LogLik** | **Wald P-value** | **LogLik** | **Wald P-value** |
| Healthy | Quintile | White European* | 7.422 | **29.244** (0.205) | - | - | - | - |
| | | Asian | 4.777 | **10.103** (0.814) | 5.876 | 0.96 | - | - |
| | | Med | 0.622 | **2.416** (0.984) | 1.009 | 0.83 | - | - |
| HR | Quartile | White European* | 4.174 | **11.380** (0.007) | - | - | - | - |
| Mediterranean | Quartile | White European | 1.752 | **3.273** (0.903) | 2.158 | 6.1e-06 | - | - |
| | | Med | 0.334 | **1.549** (0.984) | 0.429 | 0.83 | - | - |
| Prudent | Quintile | White European* | 1.224 | **2.401** (0.952) | - | - | - | - |
| | | Asian | 1.214 | **2.418** (0.954) | 0.969 | 0.95 | - | - |
| Plant based | Quintile | White European | 1.909 | **3.490** (0.798) | 2.383 | 0.95 | - | - |
| | | Asian | 0.917 | **2.673** (0.845) | 1.554 | 0.98 | - | - |
| Unhealthy | Quintile | White European | 6.612 | **14.760** (0.099) | 9.123 | 1.2e-07 | - | - |
| | | Asian | 1.101 | **1.914** (0.185) | 0.552 | 0.41 | - | - |
| | | Mediterranean | 1.819 | **3.902** (0.827) | 2.480 | 0.027 | - | - |
| Western | Quintile | White European | 1.157 | **3.297** (0.663) | 1.911 | 0.023 | - | - |
| Fried/ fast food | Quartile | White European | 1.190 | **2.458** (0.023) | 1.348 | 1e-04 | - | - |
| Meat pattern | Quintile | White European | 1.846 | **4.136** (0.078) | 2.742 | 0.0012 | 0.566 | 0.051 |
| | | Asian | 1.196 | **2.977** (0.061) | 1.570 | 0.79 | - | - |
| High protein | Quintile | Asian | -0.091 | **1.785** (0.345) | 0.398 | 0.62 | - | - |

Table showing dose response analyses results.* Represent models were the quadratic* model was performed. Bolded LogLik values illustrate the best fitting model. P-values of the best fitting models are shown in brackets. No identified associations were significant following a Bonferroni correction. Wald ratios illustrate deviation from a linear model. 4K cubic spline models were performed when each study presented at least 4 measures of adherence due to the intractability of the algorithm if less than 4 were included.

**A**


Author
TE
SE
Weight
OR
Odds Ratio
IV, Random, 95% CI
Ethnicity = Asian
Yi 2017 -0.71 0.4613 1.2% 0.49 [0.20; 1.22]
He 2015 (Prudent) 0.00 0.1446 4.6% 1.00 [0.75; 1.33]
He 2015 (Plant based) -0.29 0.1191 5.1% 0.75 [0.59; 0.95]
Mak 2018 -0.03 0.2121 3.4% 0.97 [0.64; 1.47]
Zhou 2018 0.04 0.0738 6.0% 1.04 [0.90; 1.20]
Yong 2020 -0.20 0.3924 1.6% 0.82 [0.38; 1.77]
Opie 2016 -0.96 0.8439 0.4% 0.38 [0.07; 1.99]
Sahariah 2016 -0.22 0.2438 3.0% 0.81 [0.50; 1.30]
Total (95% CI) 25.4% 0.90 [0.78; 1.04]
Heterogeneity: Tau² = 0.0095; Chi² = 9.23, df = 7 (P = 0.24); I² = 24%
Ethnicity = Australian National
Looman 2018 -0.08 0.1697 4.2% 0.92 [0.66; 1.29]
Markovic 2016 -0.09 0.5096 1.0% 0.91 [0.34; 2.47]
Total (95% CI) 5.2% 0.92 [0.67; 1.26]
Heterogeneity: Tau² = 0; Chi² = 0, df = 1 (P = 0.98); I² = 0%
Ethnicity = Mediterranean
Assaf-Balut 2018 -1.06 0.3353 2.0% 0.35 [0.18; 0.67]
Donazar-Ezcurra 2017 0.08 0.2360 3.1% 1.08 [0.68; 1.72]
Karamanos 2014 -0.48 0.2207 3.3% 0.62 [0.40; 0.95]
Assaf-Balut 2017 -0.40 0.1701 4.2% 0.67 [0.48; 0.94]
Simmons 2017 0.39 0.3874 1.6% 1.47 [0.69; 3.15]
Total (95% CI) 14.2% 0.73 [0.50; 1.07]
Heterogeneity: Tau² = 0.1185; Chi² = 11.74, df = 4 (P = 0.02); I² = 66%
Ethnicity = White European
Badon 2016 -0.09 0.2246 3.3% 0.92 [0.59; 1.42]
Bao 2014 -0.18 0.1088 5.3% 0.84 [0.68; 1.04]
Gicevic 2018 -0.48 0.1121 5.3% 0.62 [0.50; 0.77]
Schoenaker 2015 -0.60 0.1533 4.5% 0.55 [0.41; 0.74]
Tobias 2012 (aHEI) -0.64 0.1016 5.5% 0.53 [0.43; 0.64]
Tobias 2012 (aMED) -0.29 0.0719 6.0% 0.75 [0.65; 0.86]
Tryggvadottir 2016 (HEI) -1.43 0.7717 0.5% 0.24 [0.05; 1.09]
Tryggvadottir 2016 (Prudent) -0.82 0.3774 1.7% 0.44 [0.21; 0.92]
Zhang 2006 0.21 0.1069 5.4% 1.23 [1.00; 1.51]
Zhang 2014 -0.25 0.0748 5.9% 0.78 [0.67; 0.90]
Flynn 2016 0.03 0.2428 3.0% 1.03 [0.64; 1.66]
Schoenaker 2015 0.04 0.1421 4.7% 1.04 [0.79; 1.38]
Opie 2016 -0.15 0.7265 0.6% 0.86 [0.21; 3.57]
Rono 2018 0.09 0.1973 3.7% 1.09 [0.74; 1.61]
Total (95% CI) 55.2% 0.78 [0.67; 0.91]
Heterogeneity: Tau² = 0.0571; Chi² = 56.8, df = 13 (P < 0.01); I² = 77%
Total (95% CI) 100.0% 0.80 [0.72; 0.90]
Heterogeneity: Tau² = 0.0487; Chi² = 89.17, df = 28 (P < 0.01); I² = 69%
Test for subgroup differences: Chi² = 2.65, df = 3 (P = 0.45)
0.1
0.5
1
2
10
GDM Odds

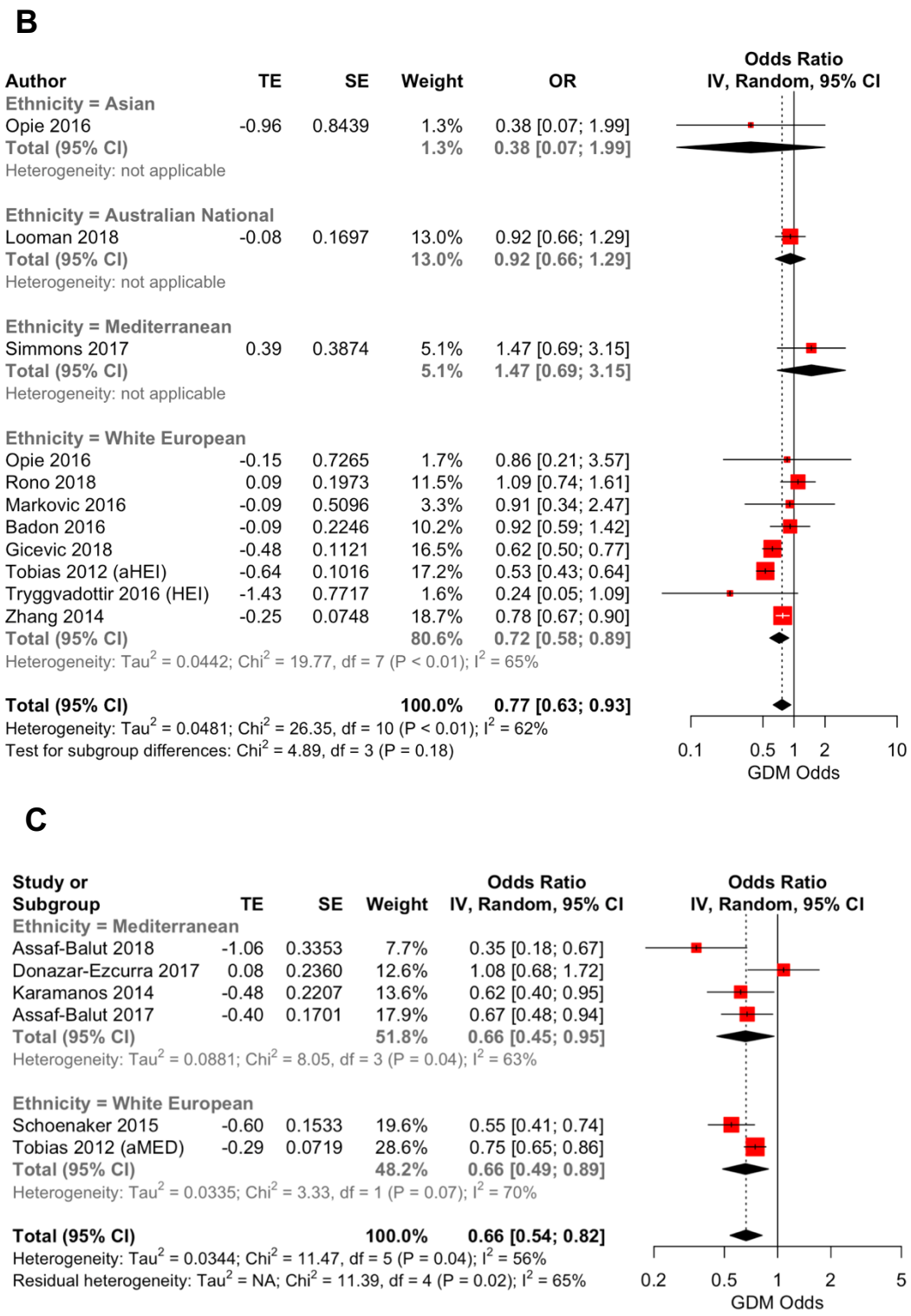



**Figure A.13: Forest plot for the combined analyses of RCTs and observational studies.**

Associations between healthy diets and GDM stratified assessed in either a RCT or observational study assessed via a DerSimonian and Laird (DL) random effects meta-analysis: **A**: Overall Healthy diet **B**: Healthy recommendations **C**: Mediterranean diet.

## Table A.4: Sensitivity analysis results.

| | | | | Random effects | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | DL | | | HKSJ | | |
| **Exposure** | **Study type** | **Subclass** | **N** | **OR (95% CI)** | **P value** | **Tau²** | **OR (95% CI)** | **P value** | **Tau²** |
| **Healthy diets** | During pregnancy | Overall | 14*** | 0.84(0.71 - 1.00) | 0.05 | 0.05 | 0.81 (0.63 - 1.03) | 0.08 | 0.14 |
| | | Asian | 5* | 0.93 (0.75 - 1.10) | 0.39 | 0.01 | 0.93 (0.78 -1.10) | 0.39 | 0.14 |
| | | Mediterranean | 3 | 0.68 (0.35 - 1.32) | 0.26 | 0.25 | 0.69 (0.12 - 3.88) | 0.45 | 0.38 |
| | | White European | 5* | 0.78 (0.50 - 1.22) | 0.28 | 0.11 | 0.74 (0.37 - 1.51) | 0.31 | 0.22 |
| | | Australian National | 1 | 0.91(0.33 - 2.48) | 0.86 | NA | 0.91 (0.33 - 2.48) | 0.86 | NA |
| | Obstetric adjustments | Overall | 17** | 0.76 (0.66 - 0.87) | ≤0.0001 | 0.06 | 0.75 (0.63 - 0.89) | ≤0.01 | 0.10 |
| | | Asian | 4 | 0.91 (0.75 - 1.11) | 0.36 | 0.02 | 0.92 (0.70 -1.20) | 0.38 | 0.01 |
| | | Mediterranean | 2 | 0.51 (0.27 -0.97) | 0.04 | 0.15 | 0.52 (0.01 - 31.0) | 0.28 | 0.13 |
| | | Australian National | 1 | 0.91 (0.34 - 2.48) | 0.86 | NA | 0.91 (0.34 - 2.48) | 0.86 | NA |
| | | White European | 10** | 0.73 (0.61 - 0.87) | ≤0.001 | 0.06 | 0.72 (0.56 -0.92) | 0.02 | 0.12 |
| | BMI under | Overall | 17* | 0.81 (0.71 - 0.93) | ≤0.01 | 0.05 | 0.81 (0.71 - 0.93) | ≤0.01 | 0.05 |
| | | Asian | 6 | 0.89 (0.75 -1.07) | 0.22 | 0.02 | 0.88 (0.68 - 1.14) | 0.26 | 0.05 |
| | | Mediterranean | 3 | 0.82 (0.62 - 1.08) | 0.15 | 0.02 | 0.82 (0.46 -1.48) | 0.29 | 0.03 |
| | | Australian National | 1 | 0.91 (0.34 - 2.48) | 0.86 | NA | 0.91 (0.34 - 2.48) | 0.86 | NA |
| | | White European | 7* | 0.76 (0.61 - 0.95) | ≤0.01 | 0.07 | 0.76 (0.57 - 1.00) | 0.05 | 0.08 |
| | BMI over | Overall | 12** | 0.68 (0.49 – 0.94) | 0.02 | 0.16 | 0.64 (0.40 - 1.02) | 0.06 | 0.46 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Asian | 2 | 0.72 (0.36 – 1.44) | 0.34 | 0 | 0.69 (0.01 - 39.9) | 0.45 | 0.07 |
| | | Mediterranean | 3 | 0.54 (0.17 -1.69) | 0.29 | 0.73 | 0.44 (0.01 - 26.25) | 0.48 | 2.28 |
| | | Australian National | 1 | 0.91 (0.34 - 2.48) | 0.86 | NA | 0.91 (0.34 - 2.48) | 0.86 | NA |
| | | White European | 6* | 0.66 (0.41 - 1.04) | 0.07 | 0.19 | 0.64 (0.35 - 1.20) | 0.13 | 0.25 |
| | Older mothers | Overall | 23*** | 0.76 (0.68 -0.84) | <0.0001 | 0.02 | 0.76 (0.67 - 0.87) | 0.0003 | 0.08 |
| | | Asian | 6 | 0.66 (0.43-1.00) | 0.03 | 0 | 0.83 (0.70 - 0.99) | 0.03 | 0 |
| | | Mediterranean | 4 | 0.66 (0.43 - 1.00) | 0.05 | 0.11 | 0.66 (0.28 - 1.59) | 0.23 | 0.24 |
| | | Australian National | 1 | 0.91 (0.34 - 2.48) | 0.86 | NA | 0.91 (0.34 - 2.48) | 0.86 | NA |
| | | White European | 12** | 0.75 (0.66 -0.86) | <0.0001 | 0.03 | 0.76 (0.63 - 0.90) | 0.005 | 0.07 |
| **HR** | During pregnancy | Overall | 7** | 0.77 (0.63 - 0.93) | 0.01 | 0.05 | 0.78 (0.60 - 1.00) | 0.049 | 0.11 |
| | | Asian | 1 | 0.38 (0.07 - 2.00) | 0.25 | NA | 0.38 (0.07 - 2.00) | 0.25 | NA |
| | | Mediterranean | 1 | 1.48 (0.69 - 3.16) | 0.31 | NA | 1.48 (0.69 - 3.16) | 0.31 | NA |
| | | White European | 5* | 0.72 (0.58 – 0.89) | 0.03 | 0.04 | 0.73 (0.55 - 0.96) | 0.03 | 0.09 |
| | Obstetric adjustments | Overall | 5 | 0.67 (0.53 - 0.84) | ≤0.001 | 0.04 | 0.66 (0.44 - 1.00) | 0.05 | 0.11 |
| | | Australian National | 1 | 0.92 (0.66 - 1.29) | 0.64 | NA | 0.92 (0.66 - 1.29) | 0.64 | NA |
| | | White European | 4 | 0.62 (0.49 - 0.80) | ≤0.001 | 0.04 | 0.65 (0.37 - 0.98) | 0.05 | 0.10 |
| | BMI Healthy/ underweight | Overall | 5 | 0.71 (0.58 -0.87) | ≤0.001 | 0.04 | 0.70 (0.50 - 0.98) | 0.03 | 0.11 |
| | | Australian National | 1 | 0.92 (0.66 - 1.29) | 0.64 | NA | 0.92 (0.66 - 1.29) | 0.64 | NA |
| | | White European | 4 | 0.67 (0.54 - 0.84) | ≤0.001 | 0.04 | 0.67 (0.47 - 0.97) | 0.04 | 0.04 |
| | BMI over | Overall | 6* | 0.96 (0.65 - 1.42) | 0.85 | 0.04 | 0.86 (0.46 - 1.63) | 0.58 | 0.23 |
| | | Asian | 1 | 0.38 (0.07 - 2.00) | 0.26 | NA | 0.38 (0.07 - 2.00) | 0.26 | NA |
| | | Mediterranean | 1 | 1.48 (0.69 - 3.16) | 0.31 | NA | 1.48 (0.69 - 3.16) | 0.31 | NA |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | White European | 4 | 0.90 (0.56 - 1.45) | 0.66 | 0.06 | 0.81 (0.32 - 2.06) | 0.53 | 0.22 |
| | Older mothers | Overall | 10* | 0.73 (0.59 - 0.90) | 0.003 | 0.05 | 0.73 (0.55 -0.97) | 0.03 | 0.13 |
| | | Asian | 1 | 0.38 (0.07 - 2.00) | 0.26 | NA | 0.38 (0.07 - 2.00) | 0.26 | NA |
| | | Australian National | 1 | 0.91 (0.34 - 2.48) | 0.86 | NA | 0.91 (0.34 -2.48) | 0.86 | NA |
| | | Mediterranean | 1 | 1.48 (0.69 - 3.16) | 0.31 | NA | 1.48 (0.69 - 3.16) | 0.31 | NA |
| | | White European | 7 | 0.72 (0.58 - 0.89) | 0.003 | 0.04 | 0.73 (0.55 - 0.96) | 0.03 | 0.09 |
| **Mediterranean diet** | During pregnancy | Mediterranean | 2 | 0.51 (0.27 - 0.97) | 0.04 | 0.15 | 0.52 (0.01 - 31.5) | 0.29 | 0.13 |
| | Obstetric adjustments | Overall | 4 | 0.62 (0.48 - 0.79) | ≤0.0001 | 0.04 | 0.60 (0.38 - 0.95) | 0.04 | 0.06 |
| | | Mediterranean | 2 | 0.66 (0.49 - 0.89) | ≤0.01 | 0.03 | 0.52 (0.01 - 31.2) | 0.29 | 0.36 |
| | | White European | 2 | 0.51 (0.27 - 0.97) | 0.04 | 0.15 | 0.66 (0.10 - 4.55) | 0.22 | 0.17 |
| | BMI under | Overall | 4 | 0.73 (0.56 -0.96) | ≤0.01 | 0.03 | 0.74 (0.49 - 1.12) | 0.10 | 0.04 |
| | | Mediterranean | 3 | 0.83 (0.63 -1.09) | 0.18 | 0.01 | 0.82 (0.46 - 1.48) | 0.29 | 0.03 |
| | | White European | 1 | 0.59 (0.47 -0.47) | ≤ 0.0001 | NA | 0.59 (0.47 -0.47) | ≤0.0001 | NA |
| | BMI over | Overall | 3 | 0.48 (0.27-0.83) | ≤0.01 | 0.13 | 0.33 (0.016 - 6.89) | 0.26 | 1.25 |
| | | Mediterranean | 2 | 0.22 (0.02 - 2.46) | 0.22 | 2.59 | 0.22 (0 - >1000) | 0.44 | 2.32 |
| | | White European | 1 | 0.51 (0.37 - 0.70) | ≤ 0.0001 | NA | 0.51 (0.37 - 0.70) | ≤0.0001 | NA |
| | Older mothers | Overall | 4 | 0.65 (0.52 - 0.82) | ≤ 0.001 | 0.02 | 0.62 (0.40 - 0.99) | 0.05 | 0.06 |
| | | Mediterranean | 3 | 0.58 (0.42 - 0.80) | ≤ 0.001 | 0.03 | 0.56 (0.25 - 1.25) | 0.09 | 0.07 |
| | | White European | 1 | 0.75 (0.65 - 0.86) | ≤ 0.0001 | NA | 0.75 (0.65 - 0.86) | ≤ 0.0001 | NA |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Prudent diet** | During pregnancy | Overall | 3 | 0.66 (0.36 - 1.22) | 0.19 | 0.19 | 0.68 (0.21 - 2.21) | 0.30 | 0.13 |
| | | Asian | 2 | 0.81 (0.42 - 1.52) | 0.51 | 0.13 | 0.81 (0.13 - 49.8) | 0.63 | 0.13 |
| | | White European | 1 | 0.44 (0.21 - 0.92) | 0.03 | NA | 0.44 (0.21 - 0.92) | 0.03 | NA |
| | BMI under | Overall | 3 | 1.04 (0.80 - 1.35) | 0.77 | 0.03 | 0.95 (0.32 - 2.80) | 0.85 | 0.16 |
| | | Asian | 2 | 0.78 (0.37 - 1.65) | 0.51 | 0.22 | 0.78 (0.01 - 98.47) | 0.64 | 0.20 |
| | | White European | 1 | 1.22 (0.99 - 1.50) | 0.07 | NA | 1.22 (0.99 -1.50) | 0.07 | NA |
| | BMI over | White European | 1 | 0.31 (0.13 - 0.74) | ≤ 0.001 | NA | 0.31 (0.13 - 0.74) | ≤ 0.001 | NA |
| | Older mothers | Overall | 4 | 0.80 (0.63 - 1.03) | 0.08 | 0.02 | 0.81 (0.52 - 1.26) | 0.22 | 0.05 |
| | | Asian | 2 | 1.05 (0.72 - 1.53) | 0.80 | 0 | 1.05 (1.02 - 1.09) | 0.03 | ≤ 0.001 |
| | | White European | 2 | 0.72 (0.55 - 0.92) | ≤0.01 | 0.01 | 0.70 (0.11 - 4.48) | 0.25 | 0.02 |
| **Plant based pattern** | During pregnancy | Overall | 3 | 0.92 (0.72 - 1.17) | 0.51 | 0.03 | 0.92 (0.58 - 1.48) | 0.54 | 0.02 |
| | | Asian | 2 | 0.90 (0.65 - 1.23) | 0.50 | 0.04 | 0.90 (0.11 - 7.03) | 0.62 | 0.04 |
| | | White European | 1 | 1.03 (0.64 - 1.66) | 0.90 | NA | 1.03 (0.64 - 1.66) | 0.90 | NA |
| | BMI under | Overall | 4 | 0.88 (0.76 - 1.03) | 0.10 | ≤ 0.01 | 0.89 (0.68 -1.15) | 0.25 | 0.02 |
| | | Asian | 3 | 0.88 (0.70 - 1.12) | 0.31 | 0.02 | 0.89 (0.52 - 1.51) | 0.44 | 0.03 |
| | | White European | 1 | 0.91 (0.73 - 1.13) | 0.37 | NA | 0.91 (0.73 -1.13) | 0.37 | NA |
| | BMI over | Overall | 2 | 0.97 (0.65 - 1.45) | 0.87 | 0 | 0.97 (0.26 - 3.56) | 0.79 | ≤ 0.01 |
| | | Asian | 1 | 1.03 (0.64 - 1.66) | 0.90 | NA | 1.03 (0.64 - 1.66) | 0.90 | NA |
| | | White European | 1 | 0.82 (0.38 - 1.77) | 0.61 | NA | 0.82 (0.38 - 1.77) | 0.61 | NA |
| | Older mothers | Overall | 5 | 0.88 (0.77 - 1.00) | 0.05 | 0 | 0.87 (0.76 - 1.00) | 0.06 | ≤ 0.01 |
| | | Asian | 3 | 0.79 (0.66 - 0.96) | 0.02 | 0 | 0.79 (0.71 - 0.89) | ≤0.01 | 0 |
| | | White European | 2 | 0.96 (0.80 - 1.16) | 0.70 | 0 | 0.98 (0.42 - 2.30) | 0.42 | 0 |
| **Unhealthy diets** | | Overall | 6 | 1.31 (1.03 – 1.67) | 0.03 | 0 | 0.98 (0.42 - 2.30) | 0.42 | 0 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | During pregnancy | Asian | 3 | 1.01 (0.71 - 1.71) | 0.67 | 0.09 | 1.10 (0.42 - 2.90) | 0.71 | 0.10 |
| | | White European | 3 | 1.52 (1.19 - 1.94) | ≤0.001 | ≤0.001 | 1.54 (0.88 -2.68) | 0.08 | 0.02 |
| | BMI under | Overall | 9 | 1.73 (1.34 - 2.24) | ≤0.0001 | 0.12 | 1.73 (1.32 - 2.27) | ≤0.01 | 0.09 |
| | | Asian | 3 | 1.29 (1.21 - 1.38) | ≤0.0001 | 0 | 1.29 (0.81 - 2.04) | 0.15 | 0.03 |
| | | Mediterranean | 2 | 1.60 (1.16 - 2.21) | ≤0.01 | 0 | 1.60 (0.41 - 6.29) | 0.14 | ≤0.01 |
| | | White European | 4 | 2.17 (1.59 - 2.96) | ≤0.0001 | 0.07 | 2.16 (1.23 - 3.79) | 0.02 | 0.09 |
| | BMI over | Overall | 3 | 1.32 (0.80 - 1.21) | 0.28 | 0.11 | 1.31 (0.42 - 4.13) | 0.41 | 0.14 |
| | | Asian | 1 | 0.79 (0.38 - 1.64) | 0.53 | NA | 0.79 (0.38 - 1.64) | 0.53 | NA |
| | | White European | 2 | 1.58 (0.96 - 2.60) | 0.07 | 0.07 | 1.58 (0.06 - 40.16) | 0.34 | 0.07 |
| | Older mothers | Overall | 9 | 1.45 (1.24 - 1.70) | ≤0.0001 | 0.02 | 1.46 (1.17 - 1.82) | ≤0.01 | 0.06 |
| | | Asian | 4 | 1.28 (1.16 - 1.41) | ≤0.0001 | ≤0.01 | 1.20 (0.77 - 1.87) | 0.28 | 0.06 |
| | | Mediterranean | 1 | 1.37 (0.89 - 2.13) | 0.15 | NA | 1.37 (0.89 - 2.13) | 0.15 | NA |
| | | White European | 4 | 1.68 (1.34 - 2.11) | ≤0.0001 | 0.02 | 1.69 (1.16 - 2.48) | 0.02 | 0.03 |
| **Fried/ fast food** | During pregnancy | White European | 1 | 1.28 (0.84 - 1.93) | 0.25 | NA | 1.28 (0.84 - 1.93) | 0.25 | NA |
| | BMI under | Overall | 3 | 2.03 (1.17 - 3.52) | 0.01 | 0.19 | 2.03 (0.64 -6.48) | 0.12 | 0.16 |
| | | Mediterranean | 1 | 1.80 (1.11 - 2.91) | 0.02 | | 1.80 (1.11 - 2.91) | 0.02 | NA |
| | | White European | 2 | 2.13 (0.88 - 5.15) | 0.10 | 0.36 | 2.13 (0.01 - 657.1) | 0.34 | 0.33 |
| **Western diet** | During pregnancy | Overall | 2 | 1.96 (0.13 - 3.06) | ≤0.01 | 0 | 1.96 (0.68 - 5.66) | 0.08 | 0.04 |
| | | Asian | 1 | 1.68 (0.66 - 4.27) | 0.28 | NA | 1.68 (0.66 - 4.27) | 0.28 | NA |
| | | White European | 1 | 2.05 (1.23 - 3.41) | ≤0.01 | NA | 2.05 (1.23 - 3.41) | ≤0.01 | NA |
| | BMI under | Overall | 3 | 1.90 (1.55 -2.38) | ≤0.0001 | 0 | 1.85 (1.16- 2.97) | 0.03 | 0.01 |
| | | Asian | 1 | 1.94 (0.80 - 4.96) | 0.14 | NA | 1.94 (0.80 - 4.96) | 0.14 | NA |
| | | Mediterranean | 1 | 1.45 (0.94 - 2.25) | 0.10 | NA | 1.45 (0.94 - 2.25) | 0.10 | NA |
| | | White European | 1 | 2.05 (1.62 - 2.61) | ≤0.001 | NA | 2.05 (1.62 - 2.61) | ≤0.001 | NA |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | BMI over | Overall | 3 | 1.32 (0.80 -1.21) | 0.28 | 0.11 | 1.31 (0.42 - 4.13) | 0.41 | 0.14 |
| | | Asian | 1 | 0.79 (0.38 - 1.64) | 0.53 | NA | 0.79 (0.38 - 1.64) | 0.53 | NA |
| | | White European | 2 | 1.58 (0.96 - 2.60) | 0.07 | 0.07 | 1.58 (0.062 - 40.16) | 0.33 | 0.07 |
| | Older mothers | Overall | 6 | 1.48 (1.23 - 1.78) | ≤0.0001 | 0.01 | 1.47 (1.09 - 1.96) | 0.02 | 0.06 |
| | | Asian | 2 | 1.19 (0.50 - 2.87) | 0.53 | 0.23 | 1.19 (0.004 - 352.62) | 0.76 | 0.22 |
| | | Mediterranean | 1 | 1.37 (0.89 - 2.13) | 0.10 | NA | 1.37 (0.89 - 2.13) | 0.10 | NA |
| | | White European | 3 | 1.55 (1.26 - 1.91) | ≤0.0001 | ≤0.01 | 1.56 (0.95 - 2.55) | 0.06 | 0.02 |
| **High protein diet** | During pregnancy | Asian | 4 | 1.69 (0.92 - 3.09) | 0.09 | 0.32 | 1.20 (0.77 - 1.87) | 0.28 | 0.06 |
| | Older mothers | Overall | 5 | 1.36 (1.01 - 1.83) | 0.04 | 0.08 | 1.36 (0.85 - 2.19) | 0.14 | 0.11 |
| | | Asian | 3 | 1.45 (0.72 - 2.90) | 0.30 | 0.32 | 1.45 (0.37 - 5.64) | 0.36 | 0.24 |
| | | White European | 2 | 1.28 (1.09 - 1.52) | ≤0.01 | 0 | 1.28 (1.28 - 1.29) | ≤0.001 | ≤0.0001 |
| **Vegetable protein** | During pregnancy | Asian | 3 | 1.20 (0.84 - 1.73) | 0.32 | 0.03 | 1.21 (0.53- 2.76) | 0.43 | 0.06 |
| | Older mothers | Overall | 3 | 0.94 (0.64 - 1.40) | 0.76 | 0.07 | 0.94 (0.41 - 2.18) | 0.78 | 0.07 |
| | | Asian | 2 | 1.15 (0.82 - 1.610 | 0.43 | 0 | 1.15 (0.31 - 4.20) | 0.40 | ≤0.01 |
| | | White European | 1 | 0.68 (0.49 - 0.95) | 0.02 | NA | 0.68 (0.49 - 0.95) | 0.02 | NA |

Table showing the comparison between initial (DL) random effects model compared to the Hartung-Knapp-Sidik-Jonkman (HKSJ) model the sensitivity analyses. Ethnic-specific BMI thresholds of 23 kg/m$^2$ for BMI in Asians and 25 kg/m$^2$ in white Europeans to classify studies as BMI under or BMI over. * Represents studies including two relevant exposures. The random exclusion of one exposure from each of these studies was not found to impact the overall effect estimate for all exposures. HR: Healthy recommendations

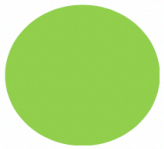
## Table A.5: Sensitivity analyses results of Asian subgroups.

| | | | | Random effect model type | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | D-L | | | HKSJ | | |
| **Exposure** | **Analysis** | **Subclass** | **N** | **OR (95% CI)** | **P value** | **Tau²** | **OR (95% CI)** | **P value** | **Tau²** |
| **Healthy** | During pregnancy | East Asian | 4* | 0.94 (0.79 - 1.11) | 0.45 | 0.01 | 0.94 (0.73 -1.20) | 0.47 | ≤0.01 |
| | Studies adjusted/ accounting for obstetric risk factors | East Asian | 4* | 0.94 (0.79 - 1.11) | 0.45 | 0.01 | 0.94 (0.73 - 1.20) | 0.47 | 0.012 |
| | | South/ South-east Asian | 1 | 0.82 (0.38 - 1.77) | 0.61 | NA | 0.82 (0.38 - 1.77) | 0.61 | NA |
| | BMI under | East Asian | 4 | 0.94 (0.78 - 1.13) | 0.48 | 0.02 | 0.94 (0.70 - 1.26) | 0.52 | 0.02 |
| | | South/ South-east Asian | 1 | 0.81 (0.50 - 1.30) | 0.38 | NA | 0.81 (0.50 - 1.30) | 0.38 | NA |
| | Older mothers | East Asian | 3* | 0.84 (0.69 - 1.02) | 0.08 | 0 | 0.86 (0.55 - 1.34) | 0.29 | 0.01 |
| | BMI over | South/ South-east Asian | 1 | 0.82 (0.38 - 1.77) | 0.61 | NA | 0.82 (0.38 - 1.77) | 0.61 | NA |
| **Plant based** | During pregnancy | East Asian † | 2 | 0.90 (0.65 - 1.23) | 0.50 | 0.04 | 0.90 (0.11 - 7.03) | 0.62 | 0.04 |
| | Studies adjusted/ accounting for obstetric risk factors | East Asian | 3 | 1.02 (0.90 - 1.17) | 0.72 | 0 | 1.02 (0.89 - 1.19) | 0.61 | ≤0.01 |
| | | South/ South-east Asian | 1 | 0.82 (0.38 - 1.77) | 0.61 | NA | 0.82 (0.38 - 1.77) | 0.61 | NA |
| | BMI under | East Asian † | 3 | 0.88 (0.70 - 1.12) | 0.31 | 0.02 | 0.90 (0.69-1.18) | 0.34 | 0.04 |
| | BMI over | South/ South-east Asian | 1 | 0.82 (0.38 - 1.77) | 0.61 | NA | 0.82 (0.38 - 1.77) | 0.61 | NA |
| | Older mothers | East Asian | 2 | 0.84 (0.71 – 1.00) | 0.05 | 0 | 0.85 (0.68 – 1.07) | 0.11 | 0.01 |
| | | South/ South-east Asian | 1 | 0.82 (0.38 - 1.77) | 0.61 | NA | 0.82 (0.38 - 1.77) | 0.61 | NA |
| **Prudent** | BMI under | East Asian † | 2 | 0.78 (0.37 - 1.65) | 0.51 | 0.22 | 0.78 (0.01 - 98.47) | 0.64 | 0.19 |
| | Older mothers | East Asian † | 2 | 1.05 (0.72 - 1.53) | 0.80 | 0 | 1.05 (1.02 - 1.09) | 0.03 | ≤0.001 |
| **Unhealthy** | During pregnancy | East Asian † | 3 | 1.01 (0.71 - 1.71) | 0.67 | 0.09 | 1.10 (0.42 - 2.90) | 0.71 | 0.10 |
| | BMI under | East Asian † | 3 | 1.29 (1.21 - 1.38) | ≤0.0001 | 0 | 1.29 (0.81 - 2.04) | 0.15 | 0.03 |
| | BMI over | South/ South-east Asian † | 1 | 0.79 (0.38 - 1.64) | 0.53 | NA | 0.79 (0.38 - 1.64) | 0.53 | NA |

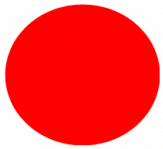
| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Older mothers | East Asian | 3 | 1.29 (1.21 -1.38) | ≤0.0001 | 0 | 1.29 (0.81 - 2.04) | 0.15 | 0.03 |
| | | South/ South-east Asian | 1 | 0.79 (0.38 - 1.64) | 0.53 | NA | 0.79 (0.38 - 1.64) | 0.53 | NA |
| **High protein diet** | During pregnancy | East Asian | 4 | 1.28 (0.73 - 2.63) | 0.39 | 0.2813 | 1.28 (0.51 - 3.25) | 0.46 | 0.28 |
| | | South/ South-east Asian | 1 | 0.79 (0.38 - 1.64) | 0.53 | NA | 0.79 (0.38 - 1.64) | 0.53 | NA |
| | Older mothers | East Asian † | 3 | 1.45 (0.72 -2.90) | 0.30 | 0.324 | 1.45 (0.37 - 5.64) | 0.36 | 0.24 |
| **Vegetable protein** | During pregnancy | East Asian | 2 | 1.15 (0.80 – 1.64) | 0.45 | 0 | 1.16 (0.22 – 6.20) | 0.47 | 0.01 |
| | | South/ South-east Asian | 1 | 1.78 (0.99 - 3.20) | 0.05 | NA | 1.78 (0.99 - 3.20) | 0.05 | NA |
| | Older mothers | East Asian † | 2 | 1.15 (0.82 - 1.610 | 0.43 | 0 | 1.15 (0.31 - 4.20) | 0.40 | ≤0.01 |

Table showing the comparison between initial (DL) random effects model compared to the Hartung-Knapp-Sidik-Jonkman (HKSJ) model the sensitivity analyses of Asian subgroups. Ethnic-specific BMI thresholds of 23 kg/m$^2$ for BMI in Asians and 25 kg/m$^2$ in white Europeans to classify studies as BMI under or BMI over.* Represent studies including two relevant exposures. † Represents exposures where all Asian studies were categorised as east or South/ South-east Asian. The random exclusion of one exposure from each of these studies was not found to impact the overall effect estimate for all exposures.

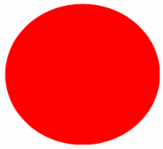
| Validity question | | Sub-questions | | |
|---|---|---|---|---|
| V4 | Was the method of handling withdrawals described? | 4.1 | Were the follow up methods the same for all groups? | ✓ |
| | | 4.2 | Was the number and characteristics of withdrawals (loss of follow up, attrition rates, drop outs) described? Was the follow up goal of 80% of a strong study met? | ✓ |
| | | **4.3** | **Were all enrolled individuals in the original sample accounted for?** | ✓ |
| | | 4.4 | Were reasons for withdrawal similar across groups? | ✓ |
| | | 4.5 | NA | NA |



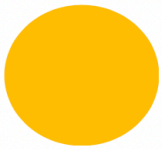
| Validity question | | Sub-questions | | |
|---|---|---|---|---|
| V4 | Was the method of handling withdrawals described? | 4.1 | Were the follow up methods the same for all groups? | ✓ |
| | | 4.2 | Was the number and characteristics of withdrawals (loss of follow up, attrition rates, drop-outs) described? Was the follow up goal of 80% of a strong study met? | ✓ |
| | | **4.3** | **Were all enrolled individuals in the original sample accounted for?** | ✗ |
| | | 4.4 | Were reasons for withdrawal similar across groups? | ✓ |
| | | 4.5 | NA | NA |



| Validity question | | Sub-questions | | |
|---|---|---|---|---|
| V4 | Was the method of handling withdrawals described? | 4.1 | Were the follow up methods the same for all groups? | ✗ |
| | | 4.2 | Was the number and characteristics of withdrawals (loss of follow up, attrition rates, drop outs) described? Was the follow up goal of 80% of a strong study met? | ✗ |
| | | **4.3** | **Were all enrolled individuals in the original sample accounted for?** | ✓ |
| | | 4.4 | Were reasons for withdrawal similar across groups? | ✗ |
| | | 4.5 | NA | NA |



| Validity question | | Sub-questions | | |
|---|---|---|---|---|
| V4 | Was the method of handling withdrawals described? | 4.1 | Were the follow up methods the same for all groups? | ✗ |
| | | 4.2 | Was the number and characteristics of withdrawals (loss of follow up, attrition rates, drop-outs) described? Was the follow up goal of 80% of a strong study met? | ✓ |
| | | **4.3** | **Were all enrolled individuals in the original sample accounted for?** | ✓ |
| | | 4.4 | Were reasons for withdrawal similar across groups? | ✗ |
| | | 4.5 | NA | NA |



## Figure A.14: Guide for ADA Risk of Bias tool.

Guide to how a question in the ADA tool was assigned to having a high, neutral or low ROB. Bolded sub-questions are sub-questions assigned as having a greater importance in the ADA tool. Coloured circles represent the score given to the overall question in each of the above scenarios. Green = positive, yellow = neutral, red = negative. For a study to be excluded 6 questions had to be assigned as negative, as stated in the ADA guidance.

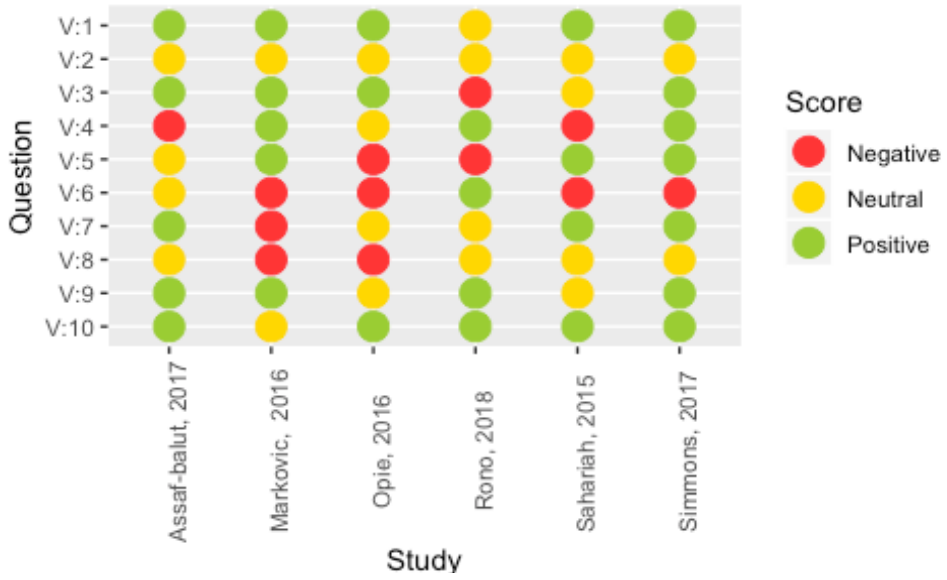



**Figure A.15: Risk of Bias scores for RCTs.**

V1-V10 correspond to the questions represented on the ADA tool. In brief, questions can be summarised as follows: V1: Clear research question, V2: Participant selection. V3: Comparable study groups. V4: Managing withdrawals. V5: Blinding. V6: Comparisons and exposures. V7: Outcome measurement. V8: Statistical analyses. V9: Conclusions and limitations. V10: Funding and sponsorship.


**Figure A.16: Risk of Bias scores for observational studies.**

V1-V10 correspond to the questions represented on the ADA tool. In brief, questions can be summarised as follows: V1: Clear research question, V2: Participant selection. V3: Comparable study groups. V4: Managing withdrawals. V5: Blinding. V6: Comparisons and exposures. V7: Outcome measurement. V8: Statistical analyses. V9: Conclusions and limitations. V10: Funding and sponsorship.

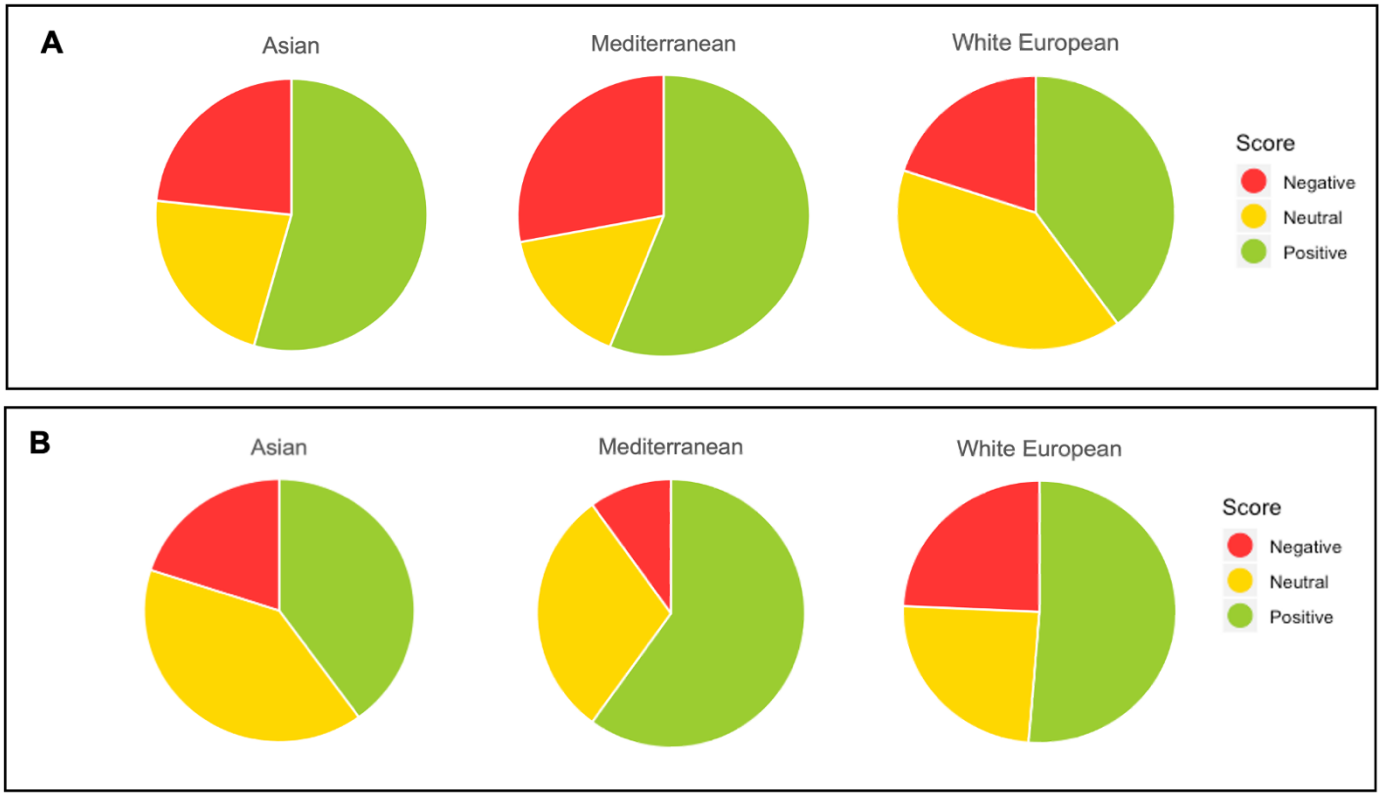



**Figure A.17: Breakdown of Risk of Bias scores by ethnicity.**

Average number of positive, negative or neutral scores per study by ethnicity and study type. Only ethnicities with ≥2 studies are represented. **A**: RCTs. **B**: Observational studies.

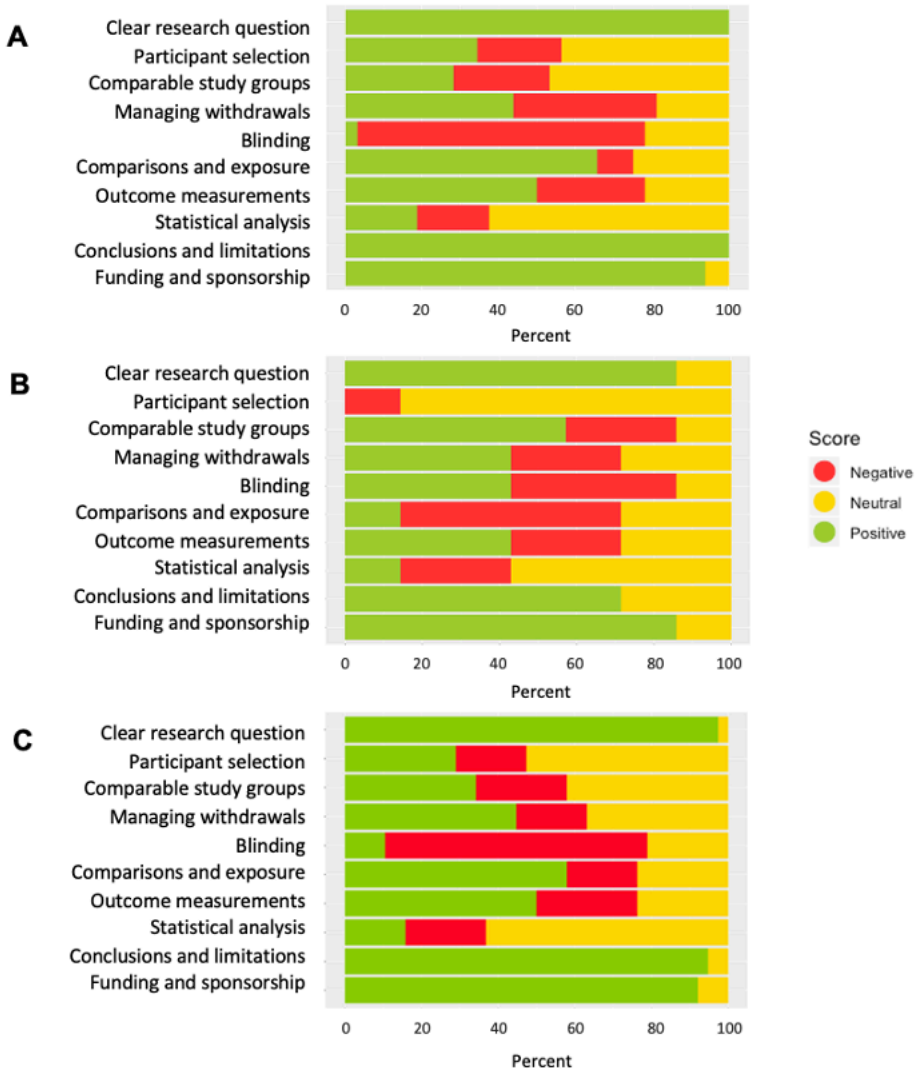



**Figure A.18: Risk of Bias scores by study type.**

V1-V10 correspond to the questions represented on the ADA tool. In brief, questions can be summarised as follows: V1: Clear research question. V2: Participant selection. V3: Comparable study groups. V4: Managing withdrawals. V5: Blinding/ V6- Comparisons and exposures. V7: Outcome measurement. V8: Statistical analyse. V9: Conclusions and limitations. V10: Funding and sponsorship. **A**: Score breakdown for RCTs. **B**: Score breakdown for observational studies. **C**: Score breakdown for combined analyses.

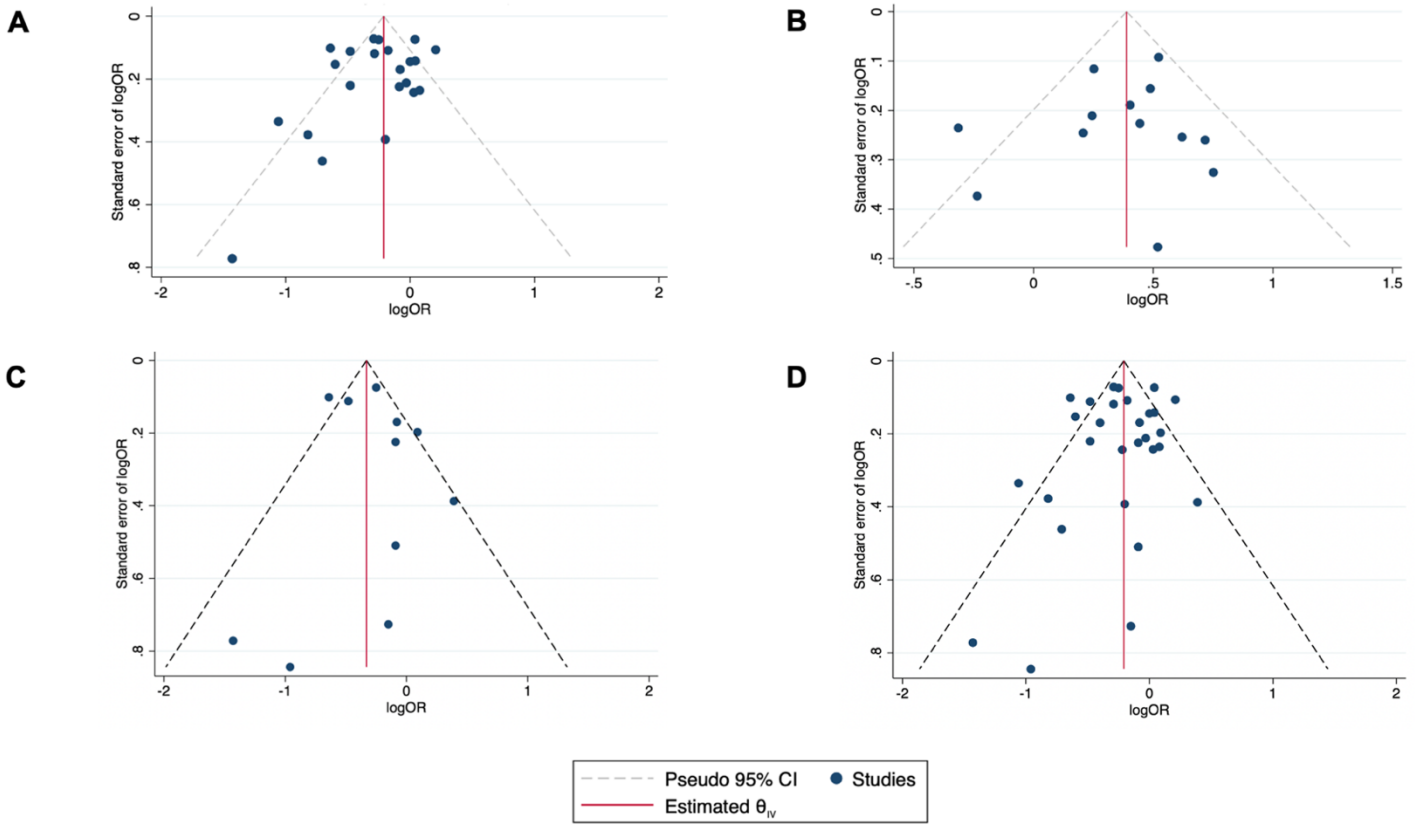



**Figure A19: Assessment of publication bias.**

**A:** Healthy (Observational) **B:** Unhealthy **C:** Healthy Recommendations (Combined) **D:** Healthy (Combined).

## Table A.6: Power analyses.

### A

| Exposure | Group | OR (95% CI) | Detectable ES |
|---|---|---|---|
| **Healthy** | Overall | 0.86 (0.69 - 1.08) | 0.04 |
| | Asian | 0.76 (0.48 - 1.20) | 0.17 |
| | Mediterranean | 0.92 (0.43 – 1.96) | 0.09 |
| | White European | 0.69 (0.24 – 1.98) | 0.18 |
| **HR** | Overall | 0.77 (0.43 – 1.39) | 0.10 |
| | White European | 0.79 (0.42 – 1.49) | 0.10 |

### B

| Exposure | Group | OR (95% CI) | Detectable ES |
|---|---|---|---|
| **Healthy** | Overall | 0.79 (0.70 - 0.89) | <0.01 |
| | Asian | 0.91 (0.78 - 1.07) | <0.01 |
| | Mediterranean | 0.64 (0.35 - 1.15) | 0.03 |
| | White European | 0.76 (0.64 - 0.90) | <0.01 |
| **HR** | Overall | 0.70 (0.56 - 0.86) | <0.01 |
| | White European | 0.66 (0.53 - 0.83) | <0.01 |
| **Mediterranean diet** | Overall | 0.66 (0.50 - 0.85) | <0.01 |
| | Mediterranean | 0.64 (0.35 - 1.15) | 0.03 |
| | White European | 0.66 (0.49 - 0.89) | 0.02 |
| **Prudent** | Overall | 0.86 (0.59 - 1.26) | 0.03 |
| | Asian | 0.81 (0.43 - 1.52) | 0.02 |
| | White European | 0.78 (0.29 - 2.10) | 0.03 |
| **Plant based** | Overall | 0.93 (0.83 - 1.04) | <0.01 |
| | Asian | 0.91 (0.75 - 1.11) | 0.02 |
| | White European* | 0.92 (0.78 - 1.08) | 0.06 |
| **Unhealthy** | Overall | 1.44 (1.25 - 1.67) | <0.01 |
| | Asian | 1.04 (0.72 - 1.51) | 0.09 |
| | Mediterranean | 1.69 (1.21 - 2.35) | 0.07 |
| | White European | 1.59 (1.41 - 1.81) | 0.04 |
| **Fried/ Fast food** | Overall* | 1.66 (1.42 - 1.93) | 0.07 |
| | White European | 1.57 (1.23 - 1.99) | <0.01 |
| **Western** | Overall | 1.51 (1.23 - 1.86) | <0.01 |
| | Asian | 1.09 (0.52 - 2.26) | 0.03 |
| | White European | 1.60 (1.26 - 2.02) | <0.01 |
| **Sweets and Seafood** | Asian | 1.01 (0.58 - 1.74) | 0.03 |
| **Meat** | Overall | 1.41 (1.22 - 1.63) | <0.01 |
| | Asian | 1.34 (0.98 - 1.84) | 0.01 |
| | White European | 1.41 (1.18 – 1.68) | 0.05 |
| **High protein diet** | Overall | 1.36 (1.05 - 1.76) | 0.02 |
| | Asian | 1.42 (0.85 - 2.35) | 0.05 |
| | White European* | 1.28 (1.06 - 1.55) | 0.05 |
| **Asian Traditional** | Asian | 1.05 (0.15 - 7.30) | 0.18 |
| **Fish** | Overall | 0.87 (0.75 - 1.00) | 0.11 |
| | Asian | 1.09 (0.71 - 1.68) | 0.18 |
| | White European | 0.85 (0.73 - 0.98) | 0.14 |
| **Carbohydrate** | Overall | 0.49 (0.38 - 0.63) | 0.11 |
| | Asian | 0.44 (0.30 - 0.63) | 0.13 |
| **Fat** | Overall* | 1.50 (1.22 - 1.83) | 0.07 |
| | White European* | 1.41 (1.12 - 1.77) | 0.08 |
| **Animal protein** | Overall* | 1.49 (1.25 - 1.77) | 0.05 |
| | Asian* | 1.51 (1.20 - 1.77) | 0.06 |
| **Vegetable protein** | Overall | 1.11 (0.72 - 1.70) | 0.02 |
| | Asian | 1.30 (0.95 - 1.77) | <0.01 |

## C

| Exposure | Group | OR (95% CI) | Detectable ES |
|---|---|---|---|
| **Healthy** | Overall | 0.75 (0.66 -0.86) | <0.01 |
| | Asian | 0.90 (0.78-1.04) | 0.02 |
| | Mediterranean | 0.73 (0.50 - 1.07) | 0.03 |
| | White European | 0.76 (0.64 - 0.90) | <0.01 |
| **HR** | Overall | 0.74 (0.60 - 0.91) | <0.01 |
| | White European | 0.70 (0.56 - 0.87) | <0.01 |
| **Mediterranean diet** | Overall | 0.66 (0.54 - 082) | <0.01 |
| | Mediterranean | 0.66 (0.45 - 0.95) | 0.03 |
| | White European | 0.66 (0.49 - 0.89) | 0.02 |

## D

| Exposure | Analysis | Group | OR (95% CI) | Detectable ES |
|---|---|---|---|---|
| **HR** | During pregnancy | Overall | 0.77 (0.63 – 0.93) | 0.09 |
| | | White European | 0.72 (0.58 – 0.89) | 0.11 |
| | Obstetric adjustments | Overall | 0.67 (0.53 - 0.84) | 0.01 |
| | | White European | 0.62 (0.49 - 0.80) | 0.01 |
| | BMI Healthy/ underweight | Overall | 0.71 (0.58 -0.87) | 0.01 |
| | | White European | 0.67 (0.54 - 0.84) | 0.01 |
| | BMI overweight/ obese | Overall | 0.96 (0.62 – 1.47) | 0.10 |
| | | White European | 0.71 (0.26 – 1.93) | 0.15 |
| | Older mothers | Overall | 0.73 (0.59 -0.90) | 0.01 |
| | | White European | 0.72 (0.58 – 0.89) | 0.02 |
| **Healthy diets** | During pregnancy | Overall | 0.84 (0.71 – 1.00) | 0.01 |
| | | Asian | 0.93(0.75 – 1.10) | 0.02 |
| | | Mediterranean | 0.77 (0.47 - 1.28) | 0.05 |
| | | White European | 0.78 (0.50 – 1.22) | 0.01 |
| | Obstetric adjustments | Overall | 0.76 (0.66 - 0.87) | 0.01 |
| | | Asian | 0.91 (0.75 - 1.11) | 0.01 |
| | | Mediterranean | 0.51 (0.27 - 0.97) | 0.04 |
| | | White European | 0.73 (0.61 - 0.87) | 0.01 |
| | BMI underweight/ healthy | Overall | 0.81 (0.71 - 0.93) | 0.01 |
| | | Asian | 0.89 (0.75 -1.07) | 0.02 |
| | | Mediterranean | 0.82 (0.62 - 1.08) | 0.02 |
| | | White European | 0.76 (0.61 - 0.95) | 0.01 |
| | BMI overweight/ obese | Overall | 0.68 (0.49 -0.95) | 0.05 |
| | | Asian | 0.44 (0.09-2.03) | 0.47 |
| | | Mediterranean | 0.54 (0.17 -1.69) | 0.15 |
| | | White European | 0.66 (0.41 – 1.04) | 0.05 |
| | Older mothers | Overall | 0.76 (0.68 – 0.84) | 0.01 |
| | | Asian* | 0.83 (0.69 – 1.00) | 0.10 |
| | | Mediterranean | 0.66 (0.43 – 1.00) | 0.04 |
| | | White European | 0.75 (0.66 – 0.86) | 0.01 |
| **Mediterranean** | During pregnancy | Mediterranean | 0.51 (0.27 - 0.97) | 0.06 |
| | Obstetric adjustments | Overall | 0.62 (0.48 - 0.79) | 0.01 |
| | | Mediterranean | 0.66 (0.49 - 0.89) | 0.06 |
| | | White European | 0.51 (0.27 - 0.97) | 0.02 |
| | BMI underweight/ healthy | Overall | 0.73 (0.56 -0.96) | 0.01 |
| | | Mediterranean | 0.83 (0.63 -1.09) | 0.02 |
| | BMI overweight/ obese | Overall | 0.48 (0.27-0.83) | 0.02 |
| | | Mediterranean | 0.22 (0.02 - 2.46) | NA |
| | | White European | 0.51 (0.37 - 0.70) | 0.02 |
| | | Overall | 0.65 (0.52 - 0.82) | 0.02 |

**E**

| | | | | |
|---|---|---|---|---|
| | Older mothers | Mediterranean | 0.58 (0.42 - 0.80) | 0.01 |
| | | White European | 0.75 (0.65 - 0.86) | 0.03 |
| **Prudent diet** | During pregnancy | Overall | 0.66 (0.36 - 1.22) | 0.08 |
| | | Asian | 0.81 (0.42 - 1.52) | 0.09 |
| | BMI underweight/ healthy | Overall | 1.04 (0.80 - 1.35) | 0.02 |
| | | Asian | 0.78 (0.37 - 1.65) | 0.05 |
| | Older mothers | Overall | 0.80 (0.63 - 1.03) | 0.01 |
| | | Asian | 1.05 (0.72 - 1.53) | 0.21 |
| | | White European | 0.72 (0.55 - 0.92) | 0.01 |
| **Unhealthy diets** | During pregnancy | Overall | 1.31 (1.03 - 1.67) | 0.02 |
| | | Asian | 1.01 (0.71 - 1.71) | 0.04 |
| | | White European | 1.52 (1.19 -1.94) | 0.01 |
| | BMI underweight/ healthy | Overall | 1.73 (1.34 - 2.24) | 0.01 |
| | | Asian* | 1.29 (1.21 - 1.38) | 0.07 |
| | | Mediterranean* | 1.60 (1.16 - 2.21) | 0.10 |
| | | White European | 2.17 (1.56 - 2.95) | 0.02 |
| | BMI overweight/ obese | Overall | 1.32 (0.80 -1.21) | 0.03 |
| | | White European | 1.58 (0.96 - 2.60) | 0.03 |
| | Older mothers | Overall | 1.45 (1.24 - 1.70) | 0.01 |
| | | Asian | 1.28 (1.16 - 1.41) | 0.01 |
| | | White European | 1.68 (1.34 - 2.11) | 0.01 |
| **Western diet** | During pregnancy | Overall* | 1.96 (0.13 - 3.06) | 0.21 |
| | BMI underweight/ healthy | Overall* | 1.90 (1.55 -2.38) | 0.07 |
| | BMI overweight/ obese | Overall | 1.32 (0.80 -1.21) | 0.03 |
| | | White European | 1.58 (0.96 -2.60) | 0.03 |
| | Older mothers | Overall | 1.48 (1.23 - 1.78) | 0.01 |
| | | Asian | 1.19 (0.50 - 2.87) | 0.11 |
| | | White European | 1.55 (1.26 - 1.91) | 0.01 |
| **Fried/ fast food** | BMI underweight/ healthy | Overall | 2.03 (1.17 - 3.52) | 0.03 |
| | | White European | 2.13 (0.88 - 5.15) | 0.05 |
| **Plant based pattern** | During pregnancy | Overall | 0.92 (0.72 - 1.17 | 0.02 |
| | | Asian | 0.90 (0.65 - 1.23) | 0.03 |
| | Obstetric adjustments | Overall | 0.92 (0.81 - 1.04) | 0.01 |
| | | Asian | 0.91 (0.75 - 1.11) | 0.01 |
| | | White European | 0.92 (0.74 - 1.13) | 0.01 |
| | BMI underweight/ healthy | Overall | 0.88 (0.76 - 1.03) | 0.01 |
| | | Asian | 0.88 (0.70 - 1.12) | 0.02 |
| | BMI overweight/ obese | Overall* | 0.97 (0.65 -1.45) | 0.20 |
| | Older mothers | Overall* | 0.88 (0.77 - 1.00) | 0.05 |
| | | Asian* | 0.79 (0.66 - 0.96) | 0.10 |
| | | White European* | 0.96 (0.80 - 1.16) | 0.01 |
| **High protein diet** | During pregnancy | Asian | 1.69 (0.92 - 3.09) | 0.05 |
| | Older mothers | Overall | 1.36 (1.01 - 1.83) | 0.06 |
| | | Asian | 1.45 (0.72 - 2.90) | 0.06 |
| | | White European* | 1.28 (1.09 - 1.52) | 0.05 |
| **Vegetable protein** | During pregnancy | Asian | 1.20 (0.84 - 1.73) | 0.03 |
| | Older mothers | Overall | 0.79 (0.31 - 2.01) | 0.05 |

| Exposure | Group | OR (95% CI) | Detectable ES |
|---|---|---|---|
| **Healthy** | East Asian | 0.87 (0.72 - 1.06) | 0.02 |
| | South/ South-east Asian | 0.81 (0.54 - 1.22) | 0.13 |
| **Plant based** | East Asian | 0.91 (0.73 - 1.15) | 0.18 |
| **Unhealthy** | East Asian | 1.10 (0.71 - 1.71) | 0.04 |
| **Animal protein** | East Asian | 1.81 (1.25 - 2.62) | 0.06 |

**F**

| Exposure | Group | Analysis | OR (95% CI) | Detectable ES |
|---|---|---|---|---|
| **Healthy** | East Asian | During pregnancy | 0.94 (0.79 - 1.11) | 0.02 |
| | | Studies adjusted/ accounting for obstetric risk factors | 0.94 (0.79 - 1.11) | 0.02 |
| | | BMI under | 0.94 (0.78 - 1.13) | 0.02 |
| | | Older mothers | 0.84 (0.71 - 1.00) | 0.11 |
| **Plant based** | East Asian | Older mothers | 0.79 (0.65 - 0.96) | 0.09 |
| **Unhealthy** | East Asian | Older mothers | 1.29 (1.21 - 1.38) | 0.11 |
| **High Protein Diet** | East Asian | During pregnancy | 1.28 (0.73 - 2.63) | 0.01 |
| **Vegetable protein** | East Asian | During pregnancy | 1.15 (0.80 - 1.64) | 0.06 |

Detectable effect size based at an 80% power level for all exposures/ subgroups containing ≥ 2 studies. Ethnic-specific BMI thresholds of 23 $kg/m^2$ for BMI in Asians and 25 $kg/m^2$ in white Europeans to classify studies as BMI under or BMI over. * Analysis represents power determined under a fixed effect model ($tau^2$ of main DL analysis =0). ES: Effect size HR: Healthy recommendations. **A**: Analysis of RCTs. **B**: Observational studies. **C**: Combined Analysis. **D**: Sensitivity Analyses. **E**: Asian subgroups. **F**: Asian subgroups sensitivity analyses.

## Table A.7: Summary of the cultural sensitivity of interventions and assessment tools utilised in non-white studies.

### A

| Study | Ethnicity | Was the intervention culturally tailored? | Details on how the intervention was ethnically tailored for the individual participants |
|---|---|---|---|
| **Assaf-balut, 2017 (Assaf-Balut et al., 2017)** | Mediterranean | Yes | MedDiet recommendations. Treatment group supplemented with 40ml of EVOO and 25-30g of pistachios daily |
| **Markovic, 2016 (Markovic et al., 2016)** | Australian Nationals | No | Diets were not ethnically tailored but individualised dietary consultations were provided. In the case of noncompliance alternative foods that were suitable for the intervention diets (low GI / high-fiber moderate GI) were suggested. |
| **Opie, 2016 (Opie et al., 2016)** | Asian and White European | Yes | Intervention focused on AGHE guidelines. Sustainable, patient driven goals were provided in a culturally sensitive manner (in relation to foods, aversions, allergies/intolerances, preferences, medical needs and food insecurity) |
| **Sahariah, 2016 (Sahariah et al., 2016)** | Asian | Yes | Daily snack resembling local street food including fresh samosas and fritters fried in sunflower oil |
| **Simmons, 2017 (Simmons et al., 2017)** | Mediterranean populations* | No | Healthy dietary recommendations involving face-to-face coaching sessions. The intervention was not ethnically tailored |

**B**

| Study | Ethnicity | Was dietary assessment culturally sensitive? | Details on how the intervention was ethnically tailored for the individual participants |
|---|---|---|---|
| **Assaf-Balut, 2018 (Assaf-Balut et al., 2018)** | Mediterranean | Yes | DNCT and MEDAS questionnaires validated in older Spanish adults in the PREDIMED to have a moderate correlation |
| **Domingues, 2014 (Domingues et al., 2014)** | Mediterranean | Yes | Previously validated FFQ. Nutrient intake assessed via Spanish food composition tables |
| **Donazar-Ezcurra, 2017 (Donazar-ezcurra et al., 2017)** | Mediterranean | Yes | Previously validated FFQ. Nutrient intake assessed via Spanish food composition tables |
| **He, 2015 (He et al., 2015)** | Asian | No | FFQ validated in a subsample of study participants and not an external cohort |
| **Hu, 2019 (Hu et al., 2019)** | Asian | Yes | Food frequency questionnaire contained 25 food items which represented commonly consumed Chinese dishes. Daily food and nutritional intakes calculated using Chinese food composition databases. Tool not previously validated |
| **Karamanos, 2014 (Karamanos et al., 2014)** | Mediterranean | Yes | Dietary questionnaire previously validated in 8 Mediterranean countries |
| **Lamyian, 2017 (Lamyian et al., 2017)** | Tehranian | Yes | Validated FFQ including food commonly consumed by Iranians. Composition of foods assessed via USDA due to incompleteness of the Iranian Food Composition table. |
| **Liang, 2018 (Liang et al., 2018)** | Asian | Yes | FFQ and 24-hr dietary recall previously validated in a Chinese population. Total energy and nutrient intake reported to reflect Chinese food |
| **Looman, 2018 (Looman et al., 2018)** | Australian Nationals | No | Nutrient intake assessed via the use of national government food database for Australian foods validated in women of reproductive age. No information on ethnicity. |
| **Mak, 2018 (Mak et al., 2018)** | Asian | Yes | FFQ validated in western China, with the exception of 10 oils/ condiments deemed too difficult to estimate |
| **Marí-Sanchis, 2018 (Sanchis et al., 2018)** | Mediterranean | Yes | Previously validated FFQ including dietary habits common to those consuming a Mediterranean diet. |

| | | | |
|---|---|---|---|
| **Pang, 2017 (Pang et al., 2017)** | Asian | Yes | 24-hour recall and 3-day food diary performed by trained professionals. Database of locally available foods used to assess nutrient composition of food along with food labels and nutrient software. |
| **Tajima, 2017 (Tajima et al., 2017)** | Asian | Yes | Weighted 3-day food diary. Nutrient intake assessed via standard tables of food composition from Japan. |
| **Yi, 2017 (Yi et al., 2017)** | Asian | Yes | FFQ validated in pregnant women in Taiwan. Nutrient and energy intake assessed based on estimates from the China Food Composition table |
| **Yong, 2020 (Yong et al., 2020)** | Asia | Yes | Semi quantitative FFQ with a previously shown good reliability in pregnant Malaysian women |
| **Zhou, 2018 (Zhou et al., 2018)** | Asia | Yes | Previously validated semi-quantitative FFQ with shown reliability in urban-dwelling pregnant women from central China |

Table summarising interventions and assessment tools used for non-white studies and evaluation of whether these interventions/tools are culturally sensitive. AGHE: Australian Guide to Healthy Eating. DNCT: Diabetes Nutrition and Complications Trial. EVOO: Extra Virgin Olive Oil. FFQ: Food Frequency Questionnaire. USDA: US Department of Agriculture. **A**: RCTs. **B**: Observational studies.

# Appendix B
# Supplementary data for Chapter 4

**Table B.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) |

*non-HDL and non-LDL cholesterol. ** And other cholines. *** 16:1 and 18:1 monounsaturated fatty acids. **** Mainly a1-acid glycoprotein.

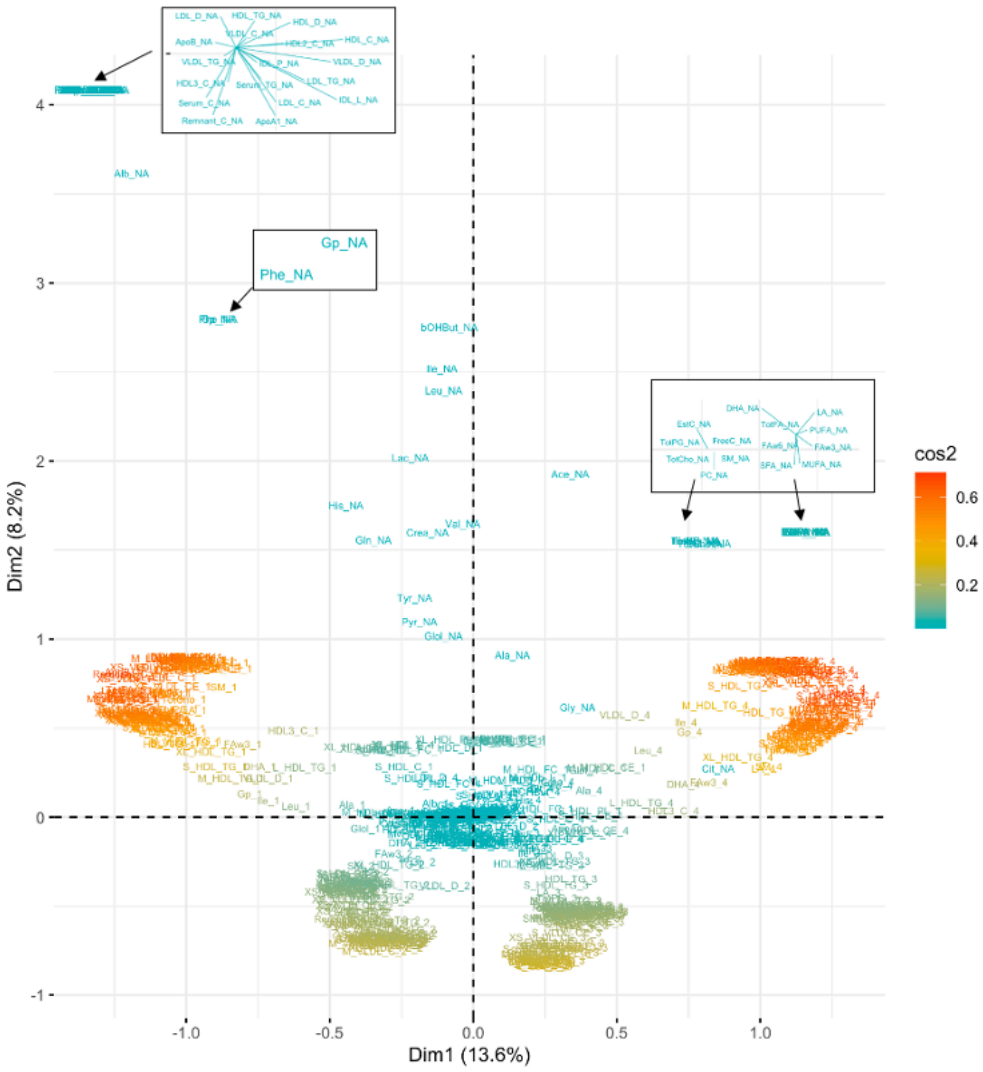



**Figure B.1: MCA plot examining missing data.**

MCA factor maps highlighting the correlation between missing and observed (split into quartiles) categories for each metabolite. Cos2 represents the groups contribution to the map position.

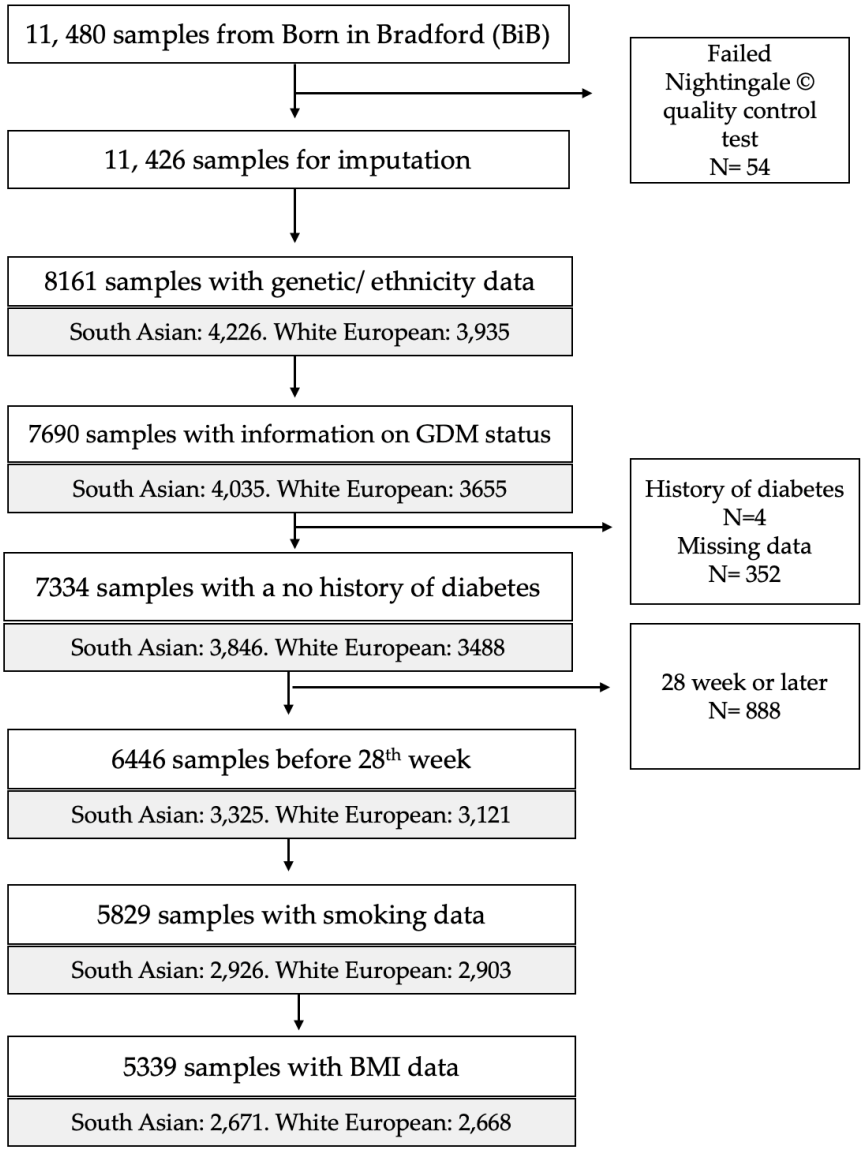



**Figure B.2: Flow of study participants.**

BiB: Born in Bradford. All individuals with BMI data also had information on maternal age, multiple pregnancy and parity.

## Table B.2: Cumulative $R^2Y$ values for PLSDA models characterising GDM status.

| | Optimised component number | | All components | |
|---|---|---|---|---|
| **Iteration** | **White European** | **South Asian** | **White European** | **South Asian** |
| **1** | 0.572 | 0.247 | 0.621 | 0.364 |
| **2** | 0.363 | 0.149 | 0.651 | 0.298 |
| **3** | 0.631 | 0.325 | 0.656 | 0.336 |
| **4** | 0.301 | 0.210 | 0.594 | 0.353 |
| **5** | 0.201 | 0.177 | 0.564 | 0.364 |
| **6** | 0.477 | 0.194 | 0.602 | 0.352 |
| **7** | 0.370 | 0.155 | 0.609 | 0.353 |
| **8** | 0.168 | 0.157 | 0.555 | 0.348 |
| **9** | 0.316 | 0.186 | 0.591 | 0.334 |
| **10** | 0.285 | 0.131 | 0.547 | 0.301 |
| **11** | 0.187 | 0.182 | 0.604 | 0.332 |
| **12** | 0.350 | 0.236 | 0.636 | 0.365 |
| **13** | 0.106 | 0.336 | 0.609 | 0.354 |
| **14** | 0.234 | 0.243 | 0.586 | 0.378 |
| **15** | 0.194 | 0.338 | 0.553 | 0.368 |
| **16** | 0.234 | 0.203 | 0.583 | 0.343 |
| **17** | 0.227 | 0.188 | 0.601 | 0.302 |
| **18** | 0.146 | 0.392 | 0.631 | 0.403 |
| **19** | 0.218 | 0.243 | 0.619 | 0.297 |
| **20** | 0.421 | 0.242 | 0.596 | 0.347 |
| **Median (range)** | **0.26 (0.11 – 0.63)** | **0.20 (0.13 – 0.39)** | **0.60 (0.55 – 0.66)** | **0.35 (0.30-0.40)** |

Optimised component number selected based upon the significance of pR2Y and pQ2, the minimisation of RMSEE and the maximisation of R2Y. Models included the following covariates: BMI (continuous), age (continuous), multiple pregnancy, parity and smoking status.

## Table B.3: Mean VIPs for GDM adjusted covariates in ethnically stratified PLSDA models.

| | Prediction of GDM | | |
|---|---|---|---|
| **Covariate** | **White European** | **South Asian** | **MW P value** |
| **Age (continuous)** | 5.99 (0.27) | 5.84 (0.17) | 0.53 |
| **BMI (continuous)** | 4.33 (0.22) | 7.06 (0.21) | ≤ 0.001 |
| **Parity** | 2.91 (0.14) | 2.72 (0.04) | 0.41 |
| **Multiple Pregnancy** | 1.59 (0.13) | 1.38 (0.09) | 0.10 |
| **Smoking Status** | 2.12 (0.11) | 1.31 (0.10) | ≤ 0.001 |

Standard errors of 20 model iterations are shown in brackets. MW: Mann-Whitney test. PLSDA: Partial Least Squares Discriminatory Analysis.

## Table B.4: Mean VIPs of metabolite measures in PLSDA analysis

| Metabolite Class | Metabolite | Prediction of Case Status | | | Prediction of Ethnicity | | |
|---|---|---|---|---|---|---|---|
| | | White European | South Asian | MW P- value | Cases | Non - Cases | MW P- value |
| XXL-VLDL | Total Lipids | 0.26 (0.03) | 0.28 (0.04) | 0.86 | 0.41 (0.05) | 0.30 (0.00) | 0.29 |
| | Phospholipids | 0.09 (0.01) | 0.08 (0.01) | 0.53 | 0.19 (0.03) | 0.20 (0.00) | 0.03 |
| | Total Cholesterol | 0.13 (0.02) | 0.13 (0.02) | 0.99 | 0.25 (0.04) | 0.24 (0.00) | 0.09 |
| | Cholesterol Esters | 0.13 (0.02) | 0.10 (0.01) | 0.90 | 0.29 (0.06) | 0.28 (0.00) | 0.04 |
| | Free Cholesterol | 0.07 (0.01) | 0.07 (0.01) | 0.76 | 0.18 (0.04) | 0.19 (0.00) | 0.03 |
| | Triglycerides | 0.20 (0.02) | 0.23 (0.04) | 0.97 | 0.39 (0.06) | 0.30 (0.00) | 0.11 |
| XL-VLDL | Total Lipids | 0.39 (0.03) | 0.30 (0.02) | 0.34 | 0.47 (0.05) | 0.37 (0.00) | 0.002 |
| | Phospholipids | 0.18 (0.03) | 0.13 (0.01) | 0.51 | 0.30 (0.05) | 0.29 (0.00) | 0.02 |
| | Total Cholesterol | 0.20 (0.02) | 0.20 (0.02) | 0.95 | 0.35 (0.05) | 0.27 (0.00) | 0.11 |
| | Cholesterol Esters | 0.15 (0.01) | 0.16 (0.02) | 0.76 | 0.34 (0.06) | 0.27 (0.00) | 0.11 |
| | Free Cholesterol | 0.16 (0.03) | 0.13 (0.02) | 0.64 | 0.25 (0.04) | 0.22 (0.00) | 0.03 |
| | Triglycerides | 0.28 (0.02) | 0.23 (0.01) | 0.09 | 0.44 (0.06) | 0.35 (0.00) | 0.17 |
| L-VLDL | Total Lipids | 0.60 (0.01) | 0.49 (0.02) | 0.01 | 0.67 (0.04) | 0.59 (0.00) | ≤0.001 |
| | Phospholipids | 0.25 (0.01) | 0.22 (0.02) | 0.01 | 0.36 (0.04) | 0.34 (0.00) | 0.03 |
| | Total Cholesterol | 0.30 (0.02) | 0.27 (0.02) | 0.13 | 0.41 (0.04) | 0.34 (0.00) | 0.82 |
| | Cholesterol Esters | 0.22 (0.02) | 0.20 (0.01) | 0.50 | 0.37 (0.04) | 0.33 (0.00) | 0.11 |
| | Free Cholesterol | 0.23 (0.02) | 0.17 (0.01) | 0.03 | 0.35 (0.04) | 0.30 (0.00) | 0.01 |
| | Triglycerides | 0.49 (0.01) | 0.37 (0.01) | 0.01 | 0.56 (0.04) | 0.47 (0.00) | ≤0.001 |
| M-VLDL | Total Lipids | 0.66 (0.05) | 0.56 (0.02) | ≤0.01 | 0.76 (0.04) | 0.69 (0.00) | ≤0.001 |
| | Phospholipids | 0.28 (0.03) | 0.22 (0.01) | ≤0.01 | 0.40 (0.04) | 0.39 (0.00) | 0.03 |
| | Total Cholesterol | 0.40 (0.04) | 0.38 (0.03) | 0.09 | 0.49 (0.03) | 0.45 (0.00) | 0.25 |
| | Cholesterol Esters | 0.41 (0.05) | 0.42 (0.05) | 0.88 | 0.50 (0.05) | 0.40 (0.00) | 0.47 |
| | Free Cholesterol | 0.24 (0.02) | 0.18 (0.01) | ≤0.01 | 0.36 (0.04) | 0.32 (0.00) | 0.11 |
| | Triglycerides | 0.54 (0.04) | 0.41 (0.01) | 0.01 | 0.60 (0.04) | 0.50 (0.00) | ≤0.001 |
| S-VLDL | Total Lipids | 0.62 (0.05) | 0.45 (0.03) | ≤0.01 | 0.69 (0.04) | 0.64 (0.00) | ≤0.001 |
| | Phospholipids | 0.43 (0.05) | 0.22 (0.01) | ≤0.01 | 0.41 (0.03) | 0.35 (0.00) | 0.01 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Total Cholesterol | 0.45 (0.0) | 0.35 (0.03) | 0.11 | 0.49 (0.03) | 0.48 (0.00) | 0.66 |
| | Cholesterol Esters | 0.37 (0.04) | 0.31 (0.03) | 0.55 | 0.50 (0.04) | 0.43 (0.00) | 0.68 |
| | Free Cholesterol | 0.28 (0.03) | 0.17 (0.01) | ≤0.01 | 0.33 (0.03) | 0.30 (0.00) | 0.49 |
| | Triglycerides | 0.42 (0.04) | 0.35 (0.03) | 0.04 | 0.55 (0.05) | 0.44 (0.00) | 0.01 |
| **XS-VLDL** | Total Lipids | 0.59 (0.05) | 0.44 (0.03) | ≤0.01 | 0.65 (0.03) | 0.65 (0.00) | 0.29 |
| | Phospholipids | 0.44 (0.06) | 0.26 (0.02) | 0.06 | 0.53 (0.05) | 0.42 (0.00) | 0.43 |
| | Total Cholesterol | 0.52 (0.05) | 0.36 (0.03) | 0.02 | 0.52 (0.04) | 0.47 (0.00) | 0.56 |
| | Cholesterol Esters | 0.56 (0.06) | 0.29 (0.02) | 0.07 | 0.55 (0.05) | 0.41 (0.00) | 0.09 |
| | Free Cholesterol | 0.27 (0.04) | 0.19 (0.02) | 0.24 | 0.43 (0.05) | 0.37 (0.00) | 0.60 |
| | Triglycerides | 0.37 (0.04) | 0.27 (0.02) | 0.18 | 0.49 (0.06) | 0.40 (0.00) | 0.23 |
| **IDL** | Concentration | 0.00 (0.00) | 0.00 (0.00) | 0.12 | 0.00 (0.00) | 0.04 (0.01) | ≤0.001 |
| | Total Lipids | 0.62 (0.05) | 0.48 (0.02) | ≤0.01 | 0.81 (0.04) | 0.93 (0.00) | 0.01 |
| | Phospholipids | 0.35 (0.03) | 0.25 (0.01) | ≤0.01 | 0.51 (0.04) | 0.49 (0.00) | 0.47 |
| | Total Cholesterol | 0.56 (0.05) | 0.40 (0.02) | ≤0.01 | 0.69 (0.04) | 0.74 (0.00) | 0.06 |
| | Cholesterol Esters | 0.55 (0.05) | 0.34 (0.01) | ≤0.01 | 0.62 (0.04) | 0.65 (0.00) | 0.04 |
| | Free Cholesterol | 0.31 (0.04) | 0.21 (0.01) | ≤0.01 | 0.40 (0.03) | 0.42 (0.00) | 0.01 |
| | Triglycerides | 0.40 (0.04) | 0.30 (0.01) | 0.12 | 0.44 (0.04) | 0.39 (0.00) | 0.82 |
| **L-LDL** | Concentration | 0.00 (0.00) | 0.00 (0.00) | ≤0.01 | 0.00 (0.00) | 0.03 (0.01) | ≤0.001 |
| | Total Lipids | 0.70 (0.50) | 0.51 (0.02) | ≤0.01 | 0.85 (0.04) | 1.00 (0.00) | ≤0.001 |
| | Phospholipids | 0.33 (0.03) | 0.23 (0.01) | ≤0.01 | 0.49 (0.04) | 0.49 (0.00) | 0.01 |
| | Total Cholesterol | 0.62 (0.05) | 0.45 (0.02) | ≤0.01 | 0.73 (0.04) | 0.84 (0.00) | ≤0.001 |
| | Cholesterol Esters | 0.55 (0.04) | 0.39 (0.02) | ≤0.01 | 0.65 (0.03) | 0.75 (0.00) | ≤0.001 |
| | Free Cholesterol | 0.30 (0.03) | 0.23 (0.01) | ≤0.01 | 0.42 (0.03) | 0.43 (0.00) | 0.06 |
| | Triglycerides | 0.38 (0.04) | 0.28 (0.02) | 0.07 | 0.44 (0.04) | 0.37 (0.00) | 0.16 |
| **M-LDL** | Concentration | 0.00 (0.00) | 0.00 (0.00) | 0.01 | 0.00 (0.00) | 0.02 (0.00) | ≤0.001 |
| | Total Lipids | 0.58 (0.04) | 0.42 (0.02) | ≤0.01 | 0.68 (0.04) | 0.79 (0.00) | ≤0.001 |
| | Phospholipids | 0.26 (0.03) | 0.17 (0.01) | ≤0.01 | 0.40 (0.04) | 0.41 (0.00) | 0.01 |
| | Total Cholesterol | 0.51 (0.04) | 0.40 (0.02) | ≤0.01 | 0.59 (0.03) | 0.67 (0.00) | 0.004 |
| | Cholesterol Esters | 0.48 (0.04) | 0.36 (0.02) | ≤0.01 | 0.56 (0.03) | 0.61 (0.00) | 0.04 |
| | Free Cholesterol | 0.25 (0.02) | 0.17 (0.01) | ≤0.01 | 0.34 (0.03) | 0.37 (0.00) | 0.01 |
| | Triglycerides | 0.31 (0.03) | 0.20 (0.01) | 0.09 | 0.35 (0.04) | 0.33 (0.00) | 0.09 |
| **S-LDL** | Concentration | 0.00 (0.00) | 0.00 (0.00) | ≤0.01 | 0.00 (0.00) | 0.03 (0.01) | ≤0.001 |
| | Total Lipids | 0.47 (0.04) | 0.33 (0.01) | ≤0.01 | 0.58 (0.03) | 0.64 (0.00) | 0.02 |
| | Phospholipids | 0.27 (0.03) | 0.17 (0.01) | ≤0.01 | 0.39 (0.04) | 0.37 (0.00) | 0.10 |
| | Total Cholesterol | 0.42 (0.04) | 0.31 (0.01) | ≤0.01 | 0.49 (0.03) | 0.54 (0.00) | 0.03 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Cholesterol Esters | 0.38 (0.03) | 0.28 (0.01) | ≤0.01 | 0.44 (0.02) | 0.48 (0.00) | 0.05 |
| | Free Cholesterol | 0.22 (0.02) | 0.16 (0.01) | 0.01 | 0.34 (0.04) | 0.29 (0.00) | 0.13 |
| | Triglycerides | 0.19 (0.02) | 0.14 (0.01) | 0.02 | 0.31 (0.04) | 0.29 (0.00) | 0.05 |
| **XL-HDL** | Concentration | 0.00 (0.00) | 0.00 (0.00) | 0.10 | 0.00 (0.00) | 0.06 (0.01) | ≤0.001 |
| | Total Lipids | 0.81 (0.06) | 0.64 (0.03) | ≤0.01 | 0.80 (0.04) | 0.49 (0.00) | ≤0.001 |
| | Phospholipids | 0.67 (0.06) | 0.46 (0.03) | 0.05 | 0.66 (0.05) | 0.38 (0.00) | ≤0.001 |
| | Total Cholesterol | 0.63 (0.05) | 0.51 (0.02) | 0.06 | 0.64 (0.04) | 0.39 (0.00) | ≤0.001 |
| | Cholesterol Esters | 0.59 (0.05) | 0.41 (0.02) | 0.17 | 0.56 (0.04) | 0.37 (0.00) | ≤0.001 |
| | Free Cholesterol | 0.35 (0.03) | 0.26 (0.01) | 0.16 | 0.41 (0.04) | 0.27 (0.00) | ≤0.001 |
| | Triglycerides | 0.23 (0.02) | 0.17 (0.01) | 0.80 | 0.34 (0.05) | 0.31 (0.00) | 0.03 |
| **L-HDL** | Concentration | 0.00 (0.00) | 0.00 (0.00) | 0.02 | 0.00 (0.00) | 0.08 (0.01) | ≤0.001 |
| | Total Lipids | 0.86 (0.06) | 0.67 (0.03) | ≤0.01 | 0.86 (0.05) | 0.44 (0.01) | ≤0.001 |
| | Phospholipids | 0.59 (0.05) | 0.42 (0.02) | 0.02 | 0.67 (0.06) | 0.39 (0.00) | ≤0.001 |
| | Total Cholesterol | 0.67 (0.05) | 0.51 (0.02) | 0.01 | 0.65 (0.04) | 0.35 (0.00) | ≤0.001 |
| | Total Cholesterol | 0.67 (0.05) | 0.51 (0.02) | 0.01 | 0.65 (0.04) | 0.35 (0.00) | ≤0.001 |
| | Cholesterol Esters | 0.60 (0.05) | 0.44 (0.02) | 0.01 | 0.57 (0.04) | 0.33 (0.00) | ≤0.001 |
| | Free Cholesterol | 0.33 (0.03) | 0.23 (0.01) | 0.05 | 0.35 (0.03) | 0.21 (0.00) | ≤0.001 |
| | Triglycerides | 0.40 (0.06) | 0.23 (0.03) | 0.58 | 0.41 (0.06) | 0.32 (0.00) | 0.29 |
| **M-HDL** | Concentration | 0.00 (0.00) | 0.00 (0.00) | 0.01 | 0.00 (0.00) | 0.09 (0.01) | ≤0.001 |
| | Total Lipids | 0.91 (0.08) | 0.74 (0.03) | ≤0.01 | 0.79 (0.04) | 0.28 (0.00) | ≤0.001 |
| | Phospholipids | 0.66 (0.06) | 0.50 (0.02) | 0.01 | 0.64 (0.06) | 0.30 (0.00) | ≤0.001 |
| | Total Cholesterol | 0.72 (0.05) | 0.56 (0.02) | ≤0.01 | 0.63 (0.04) | 0.28 (0.00) | ≤0.001 |
| | Cholesterol Esters | 0.68 (0.05) | 0.49 (0.02) | 0.01 | 0.60 (0.04) | 0.28 (0.00) | ≤0.001 |
| | Free Cholesterol | 0.33 (0.03) | 0.24 (0.01) | 0.01 | 0.32 (0.03) | 0.20 (0.00) | ≤0.001 |
| | Triglycerides | 0.37 (0.04) | 0.27 (0.03) | 0.66 | 0.42 (0.05) | 0.31 (0.00) | 0.21 |
| **S-HDL** | Concentration | 0.00 (0.00) | 0.00 (0.00) | 0.01 | 0.00 (0.00) | 0.09 (0.01) | ≤0.001 |
| | Total Lipids | 0.97 (0.07) | 0.77 (0.02) | ≤0.01 | 0.79 (0.05) | 0.30 (0.00) | ≤0.001 |
| | Phospholipids | 0.79 (0.06) | 0.62 (0.03) | ≤0.01 | 0.63 (0.05) | 0.32 (0.00) | ≤0.001 |
| | Total Cholesterol | 0.71 (0.05) | 0.55 (0.02) | 0.01 | 0.61 (0.04) | 0.31 (0.00) | ≤0.001 |
| | Cholesterol Esters | 0.69 (0.05) | 0.51 (0.01) | 0.05 | 0.60 (0.04) | 0.32 (0.00) | ≤0.001 |
| | Free Cholesterol | 0.37 (0.04) | 0.24 (0.01) | 0.02 | 0.41 (0.06) | 0.26 (0.00) | 0.13 |
| | Triglycerides | 0.28 (0.03) | 0.19 (0.01) | 0.13 | 0.35 (0.05) | 0.34 (0.00) | 0.01 |
| **Lipoprotein particle size** | Mean diameter VLDL | 1.58 (0.07) | 1.38 (0.07) | 0.04 | 1.51 (0.09) | 0.88 (0.00) | ≤0.001 |
| | Mean diameter LDL | 0.94 (0.07) | 0.88 (0.04) | 0.58 | 0.92 (0.06) | 0.36 (0.00) | ≤0.001 |
| | Mean diameter HDL | 0.87 (0.06) | 0.75 (0.04) | 0.03 | 0.85 (0.05) | 0.44 (0.00) | ≤0.001 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Amino Acids** | Alanine | 1.02 (0.1) | 0.89 (0.05) | 0.09 | 1.10 (0.08) | 0.35 (0.00) | ≤0.001 |
| | Glutamine | 1.08 (0.09) | 0.96 (0.06) | 0.18 | 0.74 (0.08) | 0.34 (0.01) | ≤0.001 |
| | Glycine | 0.89 (0.10) | 0.73 (0.07) | 0.231 | 0.70 (0.06) | 0.39 (0.00) | ≤0.001 |
| | Histidine | 0.55 (0.08) | 0.79 (0.05) | 0.02 | 0.49 (0.06) | 0.36 (0.00) | 0.10 |
| | Isoleucine | 0.53 (0.08) | 0.51 (0.06) | 0.99 | 0.46 (0.06) | 0.32 (0.00) | 0.13 |
| | Leucine | 0.59 (0.09) | 0.60 (0.07) | 0.97 | 0.51 (0.06) | 0.33 (0.00) | 0.01 |
| | Valine | 0.71 (0.08) | 0.77 (0.06) | 0.66 | 0.56 (0.06) | 0.31 (0.00) | 0.003 |
| | Phenylalanine | 0.89 (0.07) | 0.53 (0.07) | ≤0.01 | 0.72 (0.06) | 0.39 (0.00) | ≤0.001 |
| | Tyrosine | 0.61 (0.09) | 0.51 (0.06) | 0.37 | 0.43 (0.05) | 0.39 (0.01) | 0.29 |
| **Apolipoproteins** | Apolipoprotein A1 | 0.70 (0.01) | 0.54 (0.01) | ≤0.01 | 0.73 (0.05) | 1.12 (0.00) | ≤0.001 |
| | Apolipoprotein B | 0.57 (0.05) | 0.37 (0.02) | ≤0.01 | 0.81 (0.06) | 0.54 (0.00) | 0.01 |
| **Cholesterols** | Total serum cholesterol | 1.04 (0.04) | 0.71 (0.03) | ≤0.01 | 1.33 (0.07) | 1.69 (0.01) | ≤0.001 |
| | VLDL cholesterol | 0.62 (0.02) | 0.48 (0.03) | ≤0.01 | 0.74 (0.04) | 0.80 (0.00) | 0.05 |
| | Remnant cholesterol | 0.75 (0.02) | 0.54 (0.03) | ≤0.01 | 0.92 (0.05) | 1.10 (0.00) | ≤0.001 |
| | LDL Cholesterol | 0.84 (0.02) | 0.64 (0.02) | ≤0.01 | 1.01 (0.05) | 1.17 (0.00) | 0.01 |
| | HDL Cholesterol | 0.80 (0.04) | 0.62 (0.02) | ≤0.01 | 0.81 (0.04) | 0.46 (0.01) | ≤0.001 |
| | HDL2 Cholesterol | 0.80 (0.03) | 0.62 (0.01) | ≤0.01 | 0.78 (0.04) | 0.43 (0.01) | ≤0.001 |
| | HDL3 Cholesterol | 0.56 (0.07) | 0.39 (0.04) | 0.36 | 0.56 (0.06) | 0.39 (0.00) | 0.03 |
| | Total esterified cholesterol | 0.99 (0.03) | 0.76 (0.04) | ≤0.01 | 1.18 (0.06) | 1.42 (0.01) | ≤0.001 |
| | Total free Cholesterol | 0.78 (0.05) | 0.66 (0.04) | 0.06 | 0.87 (0.05) | 0.98 (0.00) | 0.01 |
| **Fatty Acids** | Total fatty acids | 1.71 (0.09) | 1.13 (0.05) | ≤0.01 | 2.15 (0.12) | 0.83 (0.01) | ≤0.001 |
| | Docosahexaenoic acid | 0.75 (0.07) | 0.65 (0.05) | 0.18 | 0.79 (0.05) | 2.71 (0.00) | ≤0.001 |
| | 18:2 Linoleic acid | 1.34 (0.06) | 1.24 (0.03) | 0.11 | 1.38 (0.07) | 0.37 (0.01) | ≤0.001 |
| | n-3 fatty acids | 0.85 (0.08) | 0.73 (0.06) | 0.17 | 1.03 (0.06) | 1.26 (0.00) | ≤0.001 |
| | n-6 fatty acids | 1.08 (0.03) | 0.92 (0.03) | ≤0.01 | 1.28 (0.07) | 0.49 (0.01) | 0.13 |
| | PUFA | 1.08 (0.03) | 0.92 (0.03) | ≤0.01 | 1.33 (0.07) | 1.42 (0.01) | 0.12 |
| | MUFA | 1.60 (0.04) | 1.22 (0.05) | ≤0.01 | 1.49 (0.08) | 1.48 (0.00) | 0.02 |
| | SFA | 1.40 (0.06) | 1.00 (0.05) | ≤0.01 | 1.51 (0.08) | 1.52 (0.01) | 0.06 |
| **Glycerides and Phospholipids** | Total serum triglycerides | 0.88 (0.02) | 0.67 (0.01) | ≤0.01 | 0.98 (0.05) | 0.99 (0.00) | 0.11 |
| | VLDL triglycerides | 0.85 (0.01) | 0.65 (0.01) | ≤0.01 | 0.89 (0.05) | 0.81 (0.00) | ≤0.001 |
| | LDL triglycerides | 0.49 (0.03) | 0.39 (0.03) | 0.03 | 0.52 (0.03) | 0.47 (0.00) | 0.03 |
| | HDL triglycerides | 0.44 (0.04) | 0.30 (0.02) | 0.13 | 0.49 (0.05) | 0.41 (0.00) | 0.82 |
| | Total phosphoglycerides | 0.96 (0.04) | 0.72 (0.03) | ≤0.01 | 1.02 (0.06) | 1.01 (0.00) | 0.03 |
| | Phosphatidylcholine | 0.92 (0.04) | 0.70 (0.04) | ≤0.01 | 0.98 (0.05) | 1.12 (0.00) | 0.002 |
| | Sphingomyelins | 0.69 (0.06) | 0.56 (0.06) | 0.18 | 0.65 (0.05) | 0.54 (0.00) | 0.02 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Total cholines | 0.96 (0.04) | 0.74 (0.04) | ≤0.01 | 1.08 (0.06) | 1.12 (0.00) | 0.68 |
| **Glycolysis Related Metabolites** | Lactate | 1.86 (0.17) | 1.73 (0.05) | 0.19 | 1.60 (0.09) | 0.53 (0.00) | ≤0.001 |
| | Pyruvate | 0.79 (0.09) | 0.62 (0.05) | 0.06 | 0.66 (0.06) | 0.28 (0.00) | ≤0.001 |
| | Citrate | 1.12 (0.09) | 0.57 (0.06) | ≤0.01 | 0.91 (0.07) | 0.29 (0.00) | ≤0.001 |
| | Glycerol | 0.69 (0.07) | 0.60 (0.05) | 0.43 | 0.53 (0.08) | 0.34 (0.00) | ≤0.001 |
| **Ketone Bodies** | Acetate | 0.59 (0.09) | 0.59 (0.07) | 0.84 | 0.50 (0.05) | 0.28 (0.00) | ≤0.001 |
| | Beta- hydroxybutyrate | 0.85 (0.10) | 0.87 (0.04) | 0.29 | 0.72 (0.07) | 0.35 (0.00) | ≤0.001 |
| **Fluid Balance and Inflammation** | Creatinine | 0.64 (0.09) | 0.51 (0.06) | 0.51 | 0.66 (0.05) | 0.27 (0.00) | ≤0.001 |
| | Albumin | 0.65 (0.07) | 0.91 (0.05) | 0.01 | 0.85 (0.06) | 0.35 (0.00) | ≤0.001 |
| | Glycoprotein acetyls | 1.25 (0.08) | 1.09 (0.06) | 0.13 | 1.14 (0.07) | 0.53 (0.01) | ≤0.001 |

VIPS were mean averaged across 20 model iterations of PLSDA. Standard errors (SE) are shown in brackets. All model iterations were significant (p value $R^2$<0.05, p value $Q^2$ < 0.05).The distributions of VIP values for each metabolite measure were compared between populations using a MW test. Standard errors shown in brackets. MW: Mann-Whitney test. PLSDA: Partial Least Squares Discriminatory Analysis.

## Table B.5: VIPs for the prediction of ethnicity.

| Variable | Model 1 | Model 2 |
|---|---|---|
| **Age** | 5.49 | 5.34 |
| **Smoking Status** | 5.08 | 4.83 |
| **Parity** | 5.01 | 4.83 |
| **BMI** | 4.68 | 4.56 |
| **Total Fatty Acids** | 2.60 | 2.55 |
| **Serum Cholesterol** | 1.60 | 1.58 |
| **SFA** | 1.56 | 1.55 |
| **MUFA** | 1.42 | 1.42 |
| **PUFA** | 1.41 | 1.40 |
| **FAw6** | 1.35 | 1.34 |
| **GDM Status** | - | 1.33 |
| **Esterified Cholesterol** | 1.33 | 1.33 |
| **LA** | 1.18 | 1.19 |
| **LDL Cholesterol** | 1.10 | 1.09 |
| **Remnant Cholesterol** | 1.04 | 1.03 |
| **PC** | 1.03 | 1.05 |
| **Total Cholesterol** | 1.03 | 1.04 |

Table showing VIP scores from PLSDA models predicting ethnicity in the overall population (n=5339). Model 1: Included covariates of maternal age (years), smoking status, parity, and BMI (continuous). Model 2: Model one covariates + GDM status. Both models were statistically significant (p value R2 < 0.05 and p value Q2 <0.05).

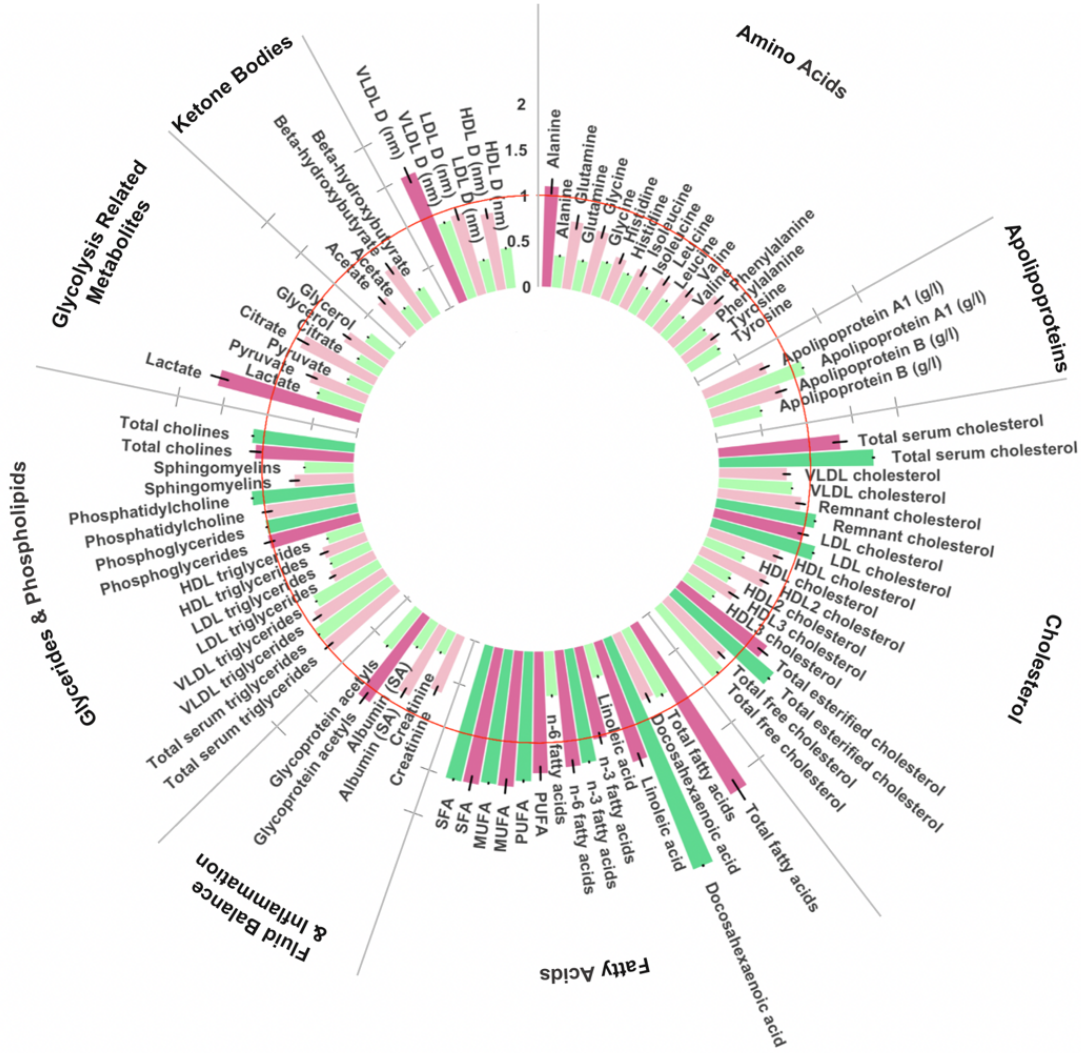



**Figure B.3: Important metabolites (VIP ≥1) in distinguishing WE and SA women in GDM and non-GDM women.**

Circular bar plot of VIPs from 20 iterations of PLSDA models predicting ethnicity in cases (n=256) and non-cases (n= 4770). Bars represent standard errors. PLSDA adjusted for maternal age (years), BMI (continuous), smoking status, parity, and multiple pregnancy status. Red line denotes VIP cut-off of 1. Dark pink: GDM cases, VIP ≥ 1, light pink: GDM cases, VIP <1, dark green: GDM non-cases, VIP ≥1, light green: GDM non-case, VIP <1. Units mmol/l unless stated. GRM: Glycolysis Related Metabolites. LPS: Lipoprotein Particle Size. MUFA: total monounsaturated fatty acids. SFA: total saturated fatty acids. VLDL_D: mean diameter of very-low density lipoproteins.

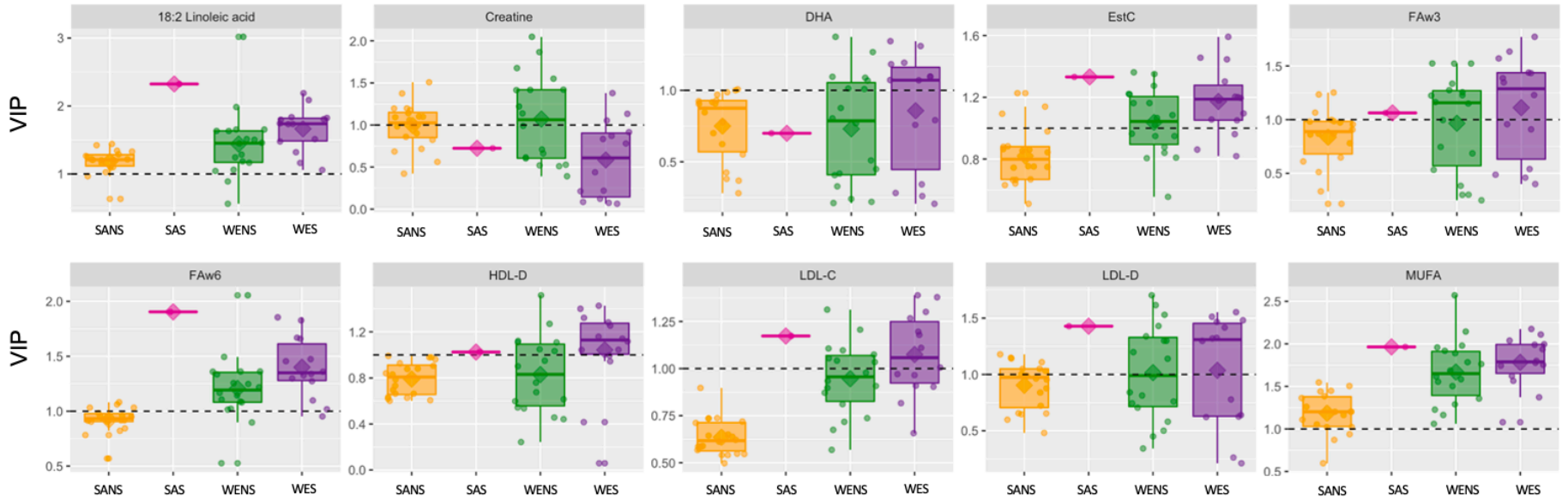



**Figure B.4: The impact of smoking on the metabolome.**

Boxplots of VIPs from significant model iterations following stratification by ethnicity and smoking status. Diamonds represent the mean VIP score. Dashed line illustrates the cut off for an important variable (VIP ≥ 1). Models were adjusted for age, BMI, parity and multiple pregnancy. **SANS**: South Asian Non-smokers. **SA**: South Asian Smokers. **WENS**: White European Non-Smokers. **WES**: WE Smokers.

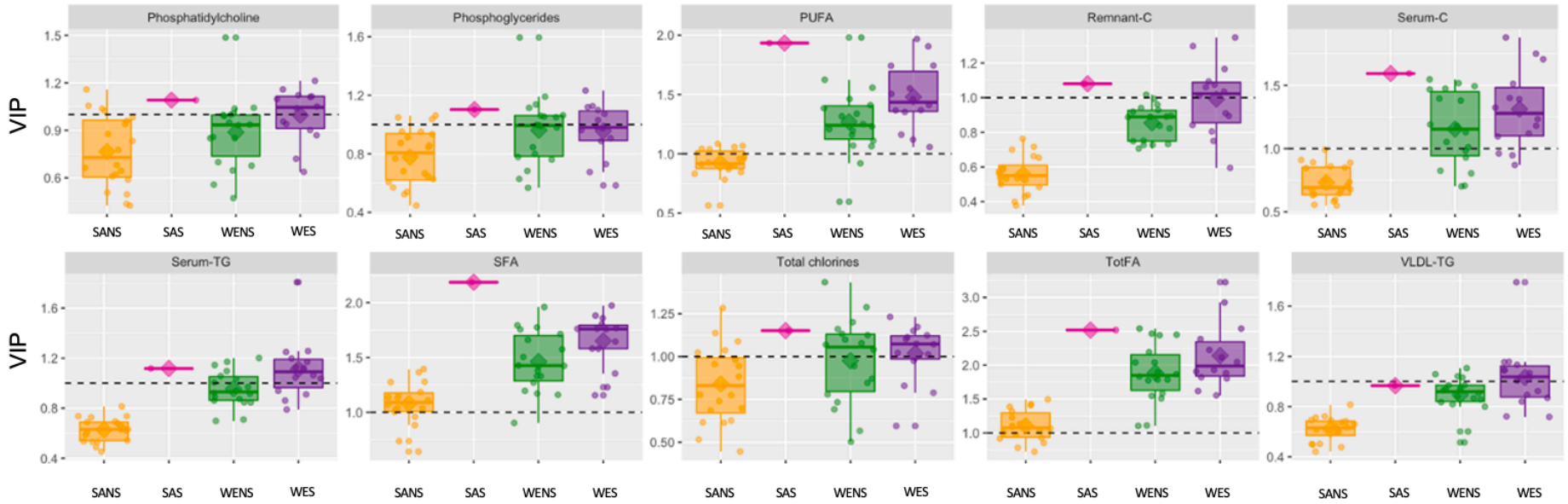



**Figure B.4 continued: The impact of smoking on the metabolome.**

Boxplots of VIPs from significant model iterations following stratification by ethnicity and smoking status. Diamonds represent the mean VIP score. Dashed line illustrates the cut off for an important variable (VIP ≥ 1). Models were adjusted for age, BMI, parity and multiple pregnancy. **SANS**: South Asian Non-smokers. **SA**: South Asian Smokers. **WENS**: White European Non-Smokers. **WES**: WE Smokers.

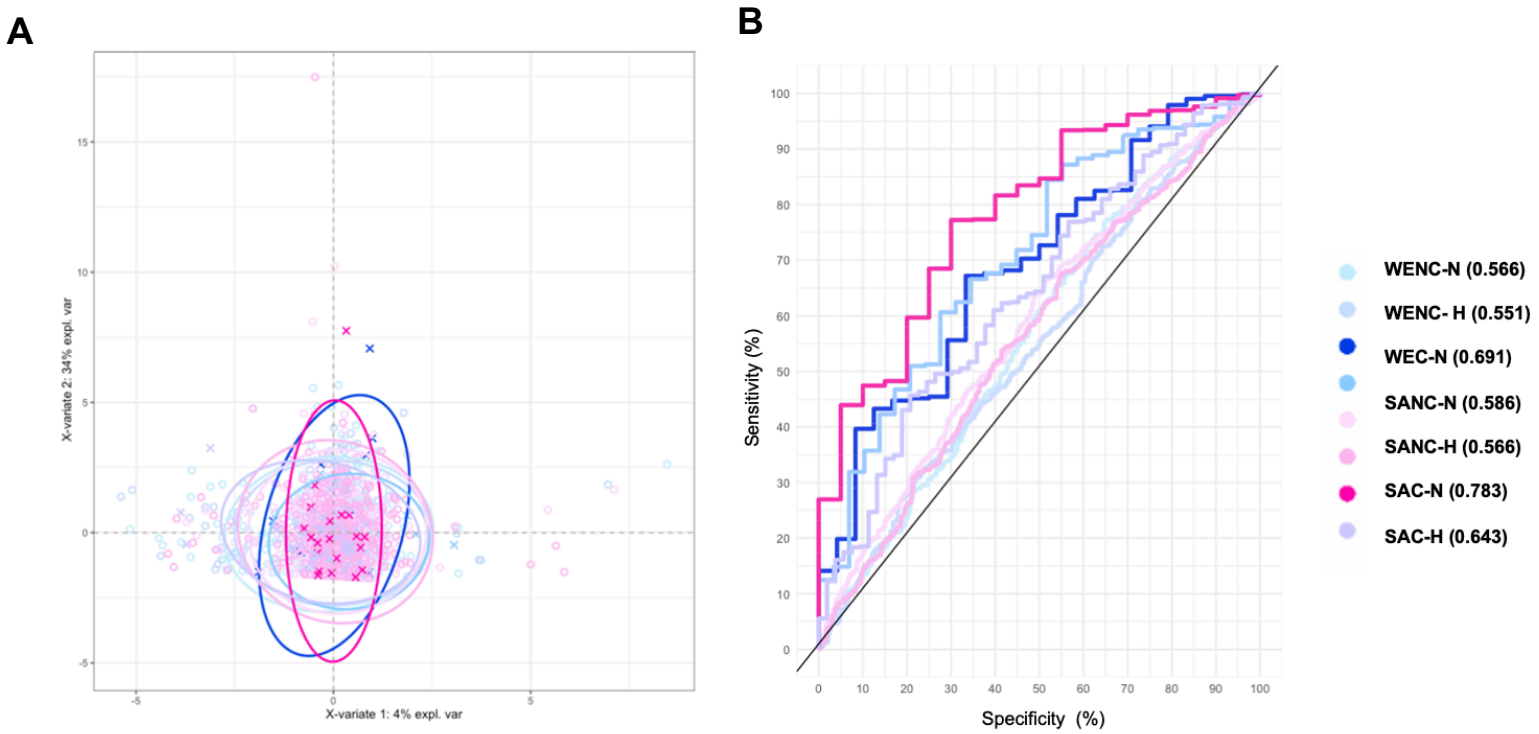
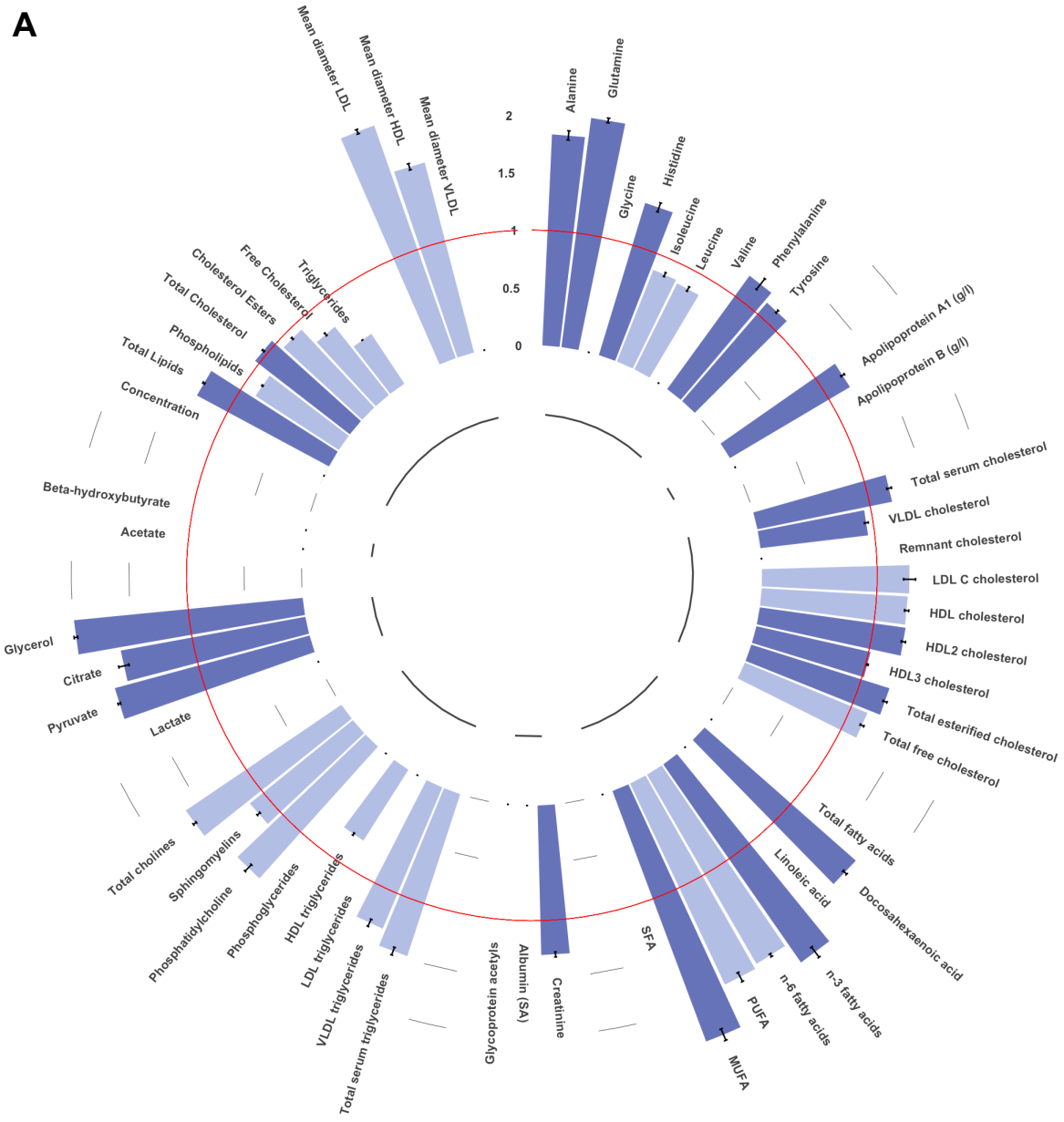
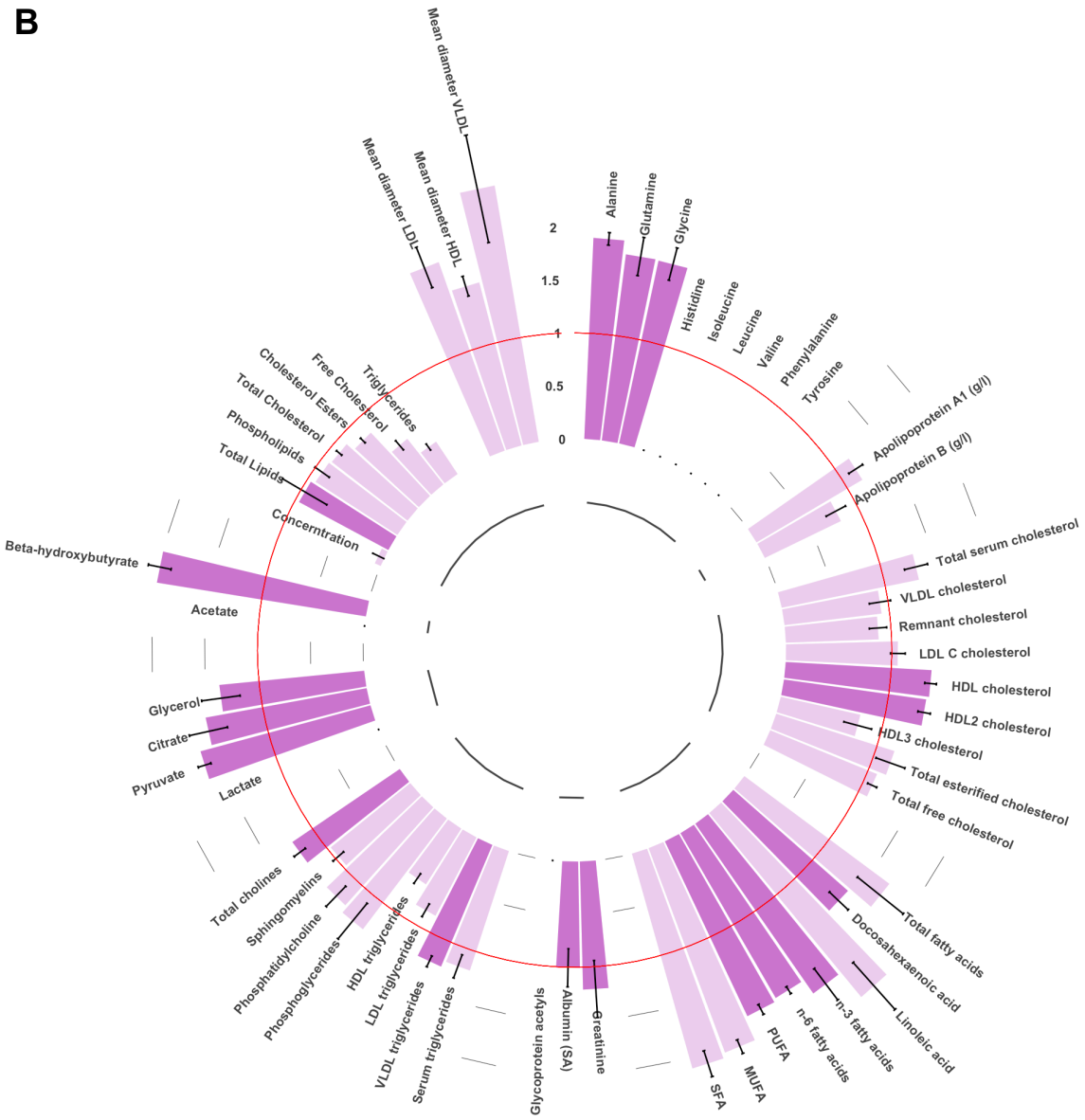



**Figure B.5: GDM in low-risk women.**

**A:** sPLSDA plot for the separation of low-risk mothers (n=1385) based upon their ethnicity, GDM status and BMI (normal vs high). **B:** Receiver Operator Curve (ROC) for sPLSDA model. SAC-H: high weight South Asian case (n=53) ; SAC-N: healthy weight South Asian case (n=20) ; SANC-H: high weight South Asian non-case (n=384); SANC-N health weight South Asian non case (n=407); WENC-H: high weight White European non-case (n=29) ; WEC-N: healthy weight White European Case (n=24); WENC-H: high weight White European non-case (n=374); WENC-N: healthy weight White European non-case (n=445).


A
Mean diameter LDL
Mean diameter HDL
Mean diameter VLDL
Alanine
Glutamine
Glycine
Histidine
Isoleucine
Leucine
Valine
Phenylalanine
Tyrosine
Apolipoprotein A1 (g/l)
Apolipoprotein B (g/l)
Total serum cholesterol
VLDL cholesterol
Remnant cholesterol
LDL C cholesterol
HDL cholesterol
HDL2 cholesterol
HDL3 cholesterol
Total esterified cholesterol
Total free cholesterol
Total fatty acids
Linoleic acid
Docosahexaenoic acid
n-3 fatty acids
n-6 fatty acids
PUFA
MUFA
SFA
Albumin (SA)
Creatinine
Glycoprotein acetyls
Total serum triglycerides
VLDL triglycerides
LDL triglycerides
HDL triglycerides
Phosphoglycerides
Phosphatidylcholine
Sphingomyelins
Total cholines
Lactate
Pyruvate
Citrate
Glycerol
Acetate
Beta-hydroxybutyrate
Concentration
Total Lipids
Phospholipids
Total Cholesterol
Cholesterol Esters
Free Cholesterol
Triglycerides
2
1.5
1
0.5
0

**B**


Mean diameter VLDL
Mean diameter LDL
Mean diameter HDL
Alanine
Glutamine
Glycine
Histidine
Isoleucine
Leucine
Valine
Phenylalanine
Tyrosine
Apolipoprotein A1 (g/l)
Apolipoprotein B (g/l)
Total serum cholesterol
VLDL cholesterol
Remnant cholesterol
LDL C cholesterol
HDL cholesterol
HDL2 cholesterol
HDL3 cholesterol
Total esterified cholesterol
Total free cholesterol
Total fatty acids
Docosahexaenoic acid
Linoleic acid
n-3 fatty acids
n-6 fatty acids
PUFA
MUFA
SFA
Creatinine
Albumin (SA)
Glycoprotein acetyls
Serum triglycerides
VLDL triglycerides
LDL triglycerides
HDL triglycerides
Phosphoglycerides
Phosphatidylcholine
Sphingomyelins
Total cholines
Lactate
Pyruvate
Citrate
Glycerol
Acetate
Beta-hydroxybutyrate
Concerntration
Total Lipids
Phospholipids
Total Cholesterol
Cholesterol Esters
Free Cholesterol
Triglycerides
0
0.5
1
1.5
2

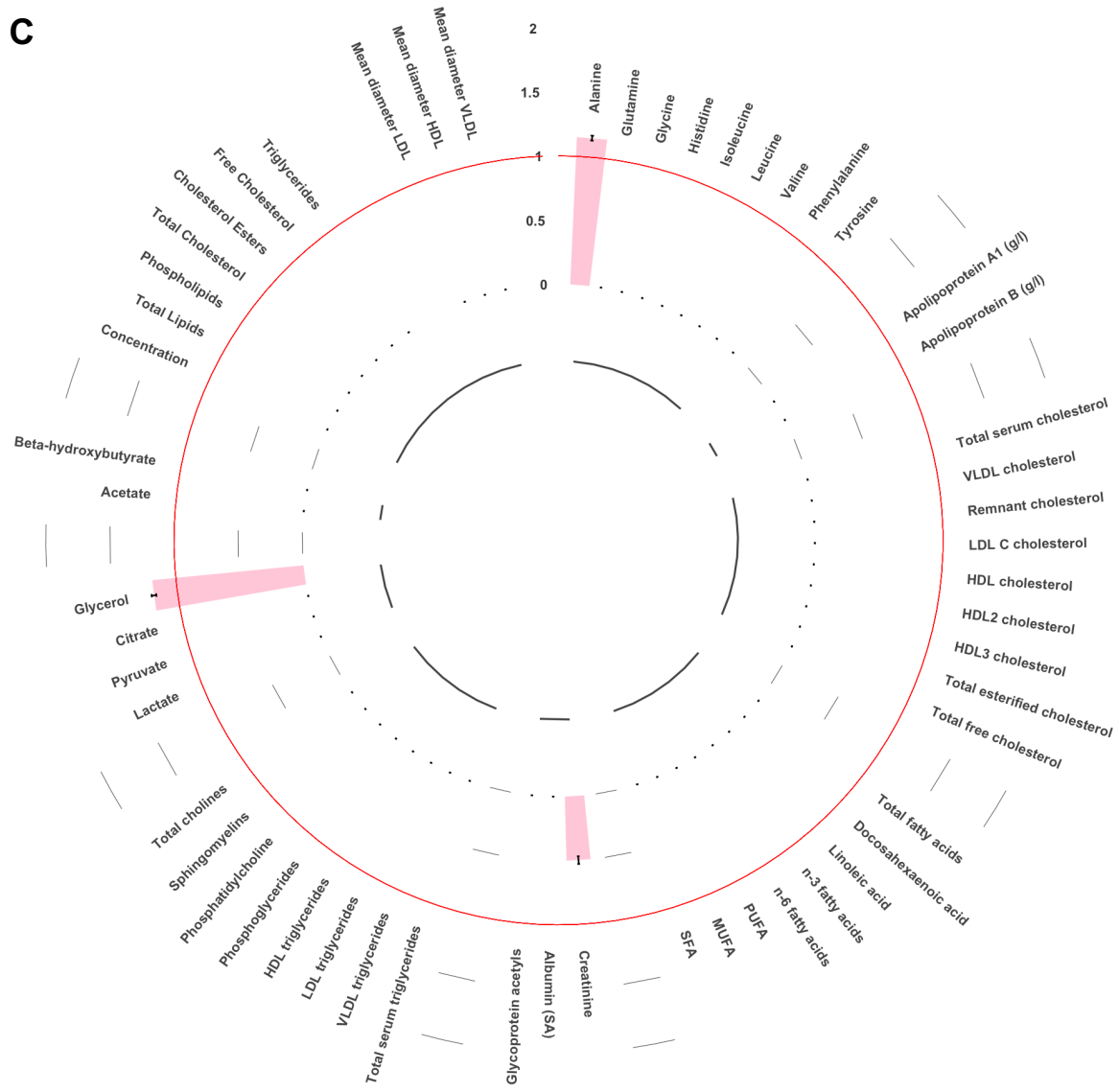



**Figure B.6. VIPs from PLSDA of metabolites identified in sPLSDA as important in distinguishing healthy-weight SA cases.**

Mean VIPs and standard errors of metabolites driving the distinction between comparisions with healthy-weight SA cases. **A**: healthy-weight SA cases vs healthy-weight WE cases. **B**: healthy-weight SA cases vs high-weight SA cases. **C**: healthy-weight SA cases from healthy-weight SA non-cases. Light Shaded colours illustrate metabolites selected based upon their high correlation (Pearson's correlation ≥ 0.90) with a metabolite identified in sPLSDA. Metabolite measures absent from the circle plot/ with a VIP of 0 indicate metabolite measures which were not identified as important in sPLSDA. All units mmol/L unless stated. SA: Surface area

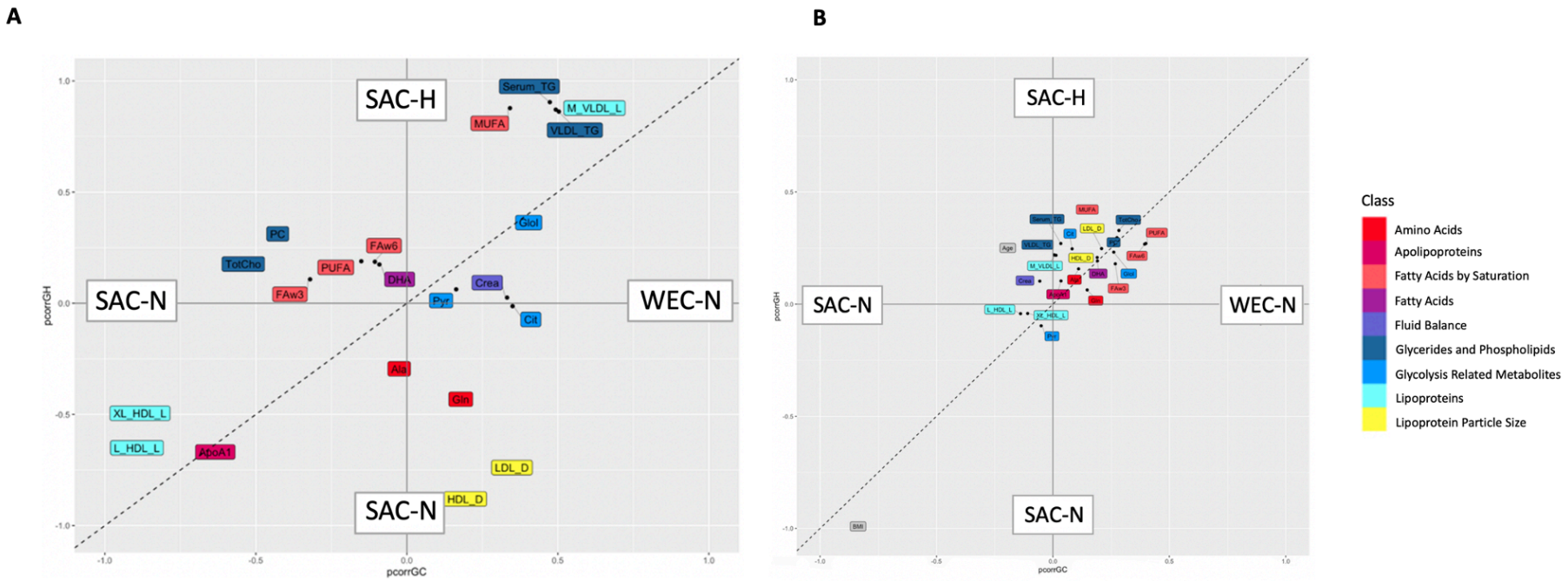
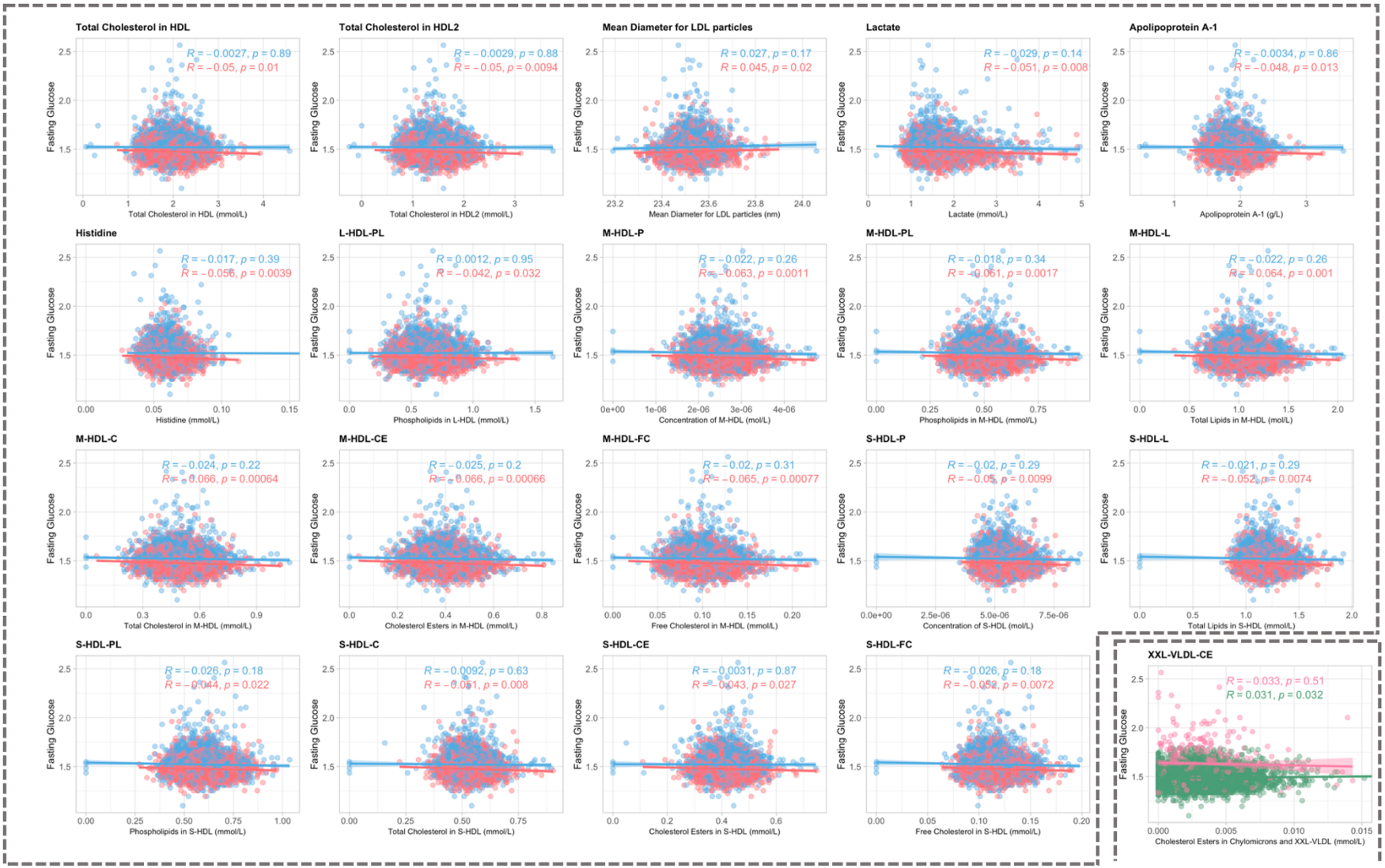


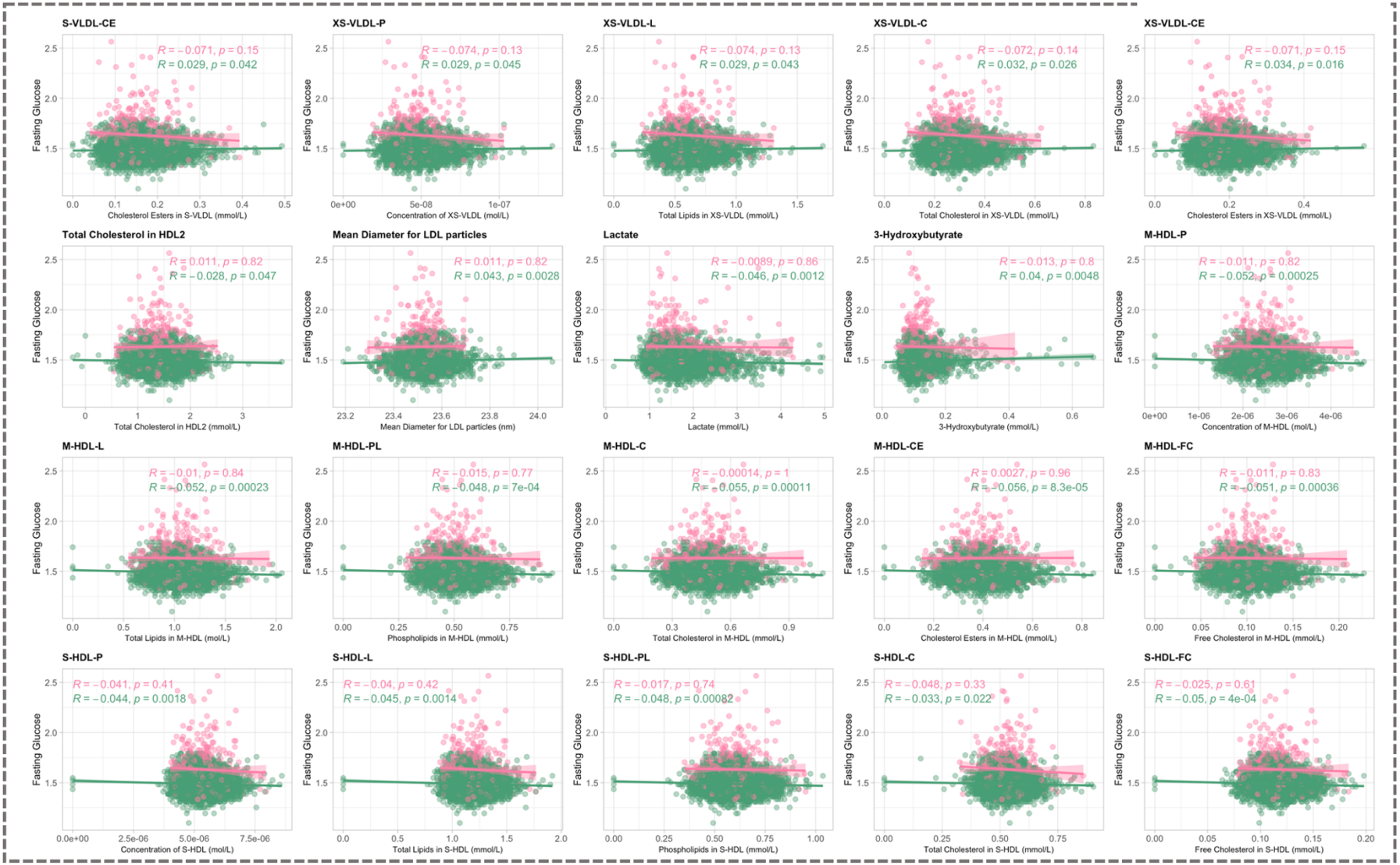

**Figure B.7: SUS plots of metabolites identified in sPLDSA to be important in characterising healthy-weight SA cases.**

Distance along the diagonals represents higher reliability. Distance along the horizontal represents higher magnitude. Top right/ bottom left corner represents metabolite values with the higher magnitude and higher reliability. **A**: Model including 34 metabolites identified in sPLSDA analysis. Orthogonal PLSDA (oPLSDA) models were non-significant. **B**: Model A + adjustment for maternal age and BMI (continuous). Centre metabolites represent lowest magnitude and lower reliability. OPLS-DA p value SAC-N vs SAC-H p value< 0.05, SAC-N vs WEC-N p value < 0.05. SAC-N: Healthy weight SA cases. SANC-N: Healthy weight SA non-cases. WEC-N: Healthy weight WE cases.

**A**


Total Cholesterol in HDL
R = −0.0027, p = 0.89
R = −0.05, p = 0.01
Total Cholesterol in HDL (mmol/L)
Total Cholesterol in HDL2
R = −0.0029, p = 0.88
R = −0.05, p = 0.0094
Total Cholesterol in HDL2 (mmol/L)
Mean Diameter for LDL particles
R = 0.027, p = 0.17
R = 0.045, p = 0.02
Mean Diameter for LDL particles (nm)
Lactate
R = −0.029, p = 0.14
R = −0.051, p = 0.008
Lactate (mmol/L)
Apolipoprotein A-1
R = −0.0034, p = 0.86
R = −0.048, p = 0.013
Apolipoprotein A-1 (g/L)
Histidine
R = −0.017, p = 0.39
R = −0.056, p = 0.0039
Histidine (mmol/L)
L-HDL-PL
R = 0.0012, p = 0.95
R = −0.042, p = 0.032
Phospholipids in L-HDL (mmol/L)
M-HDL-P
R = −0.022, p = 0.26
R = −0.063, p = 0.0011
Concentration of M-HDL (mol/L)
M-HDL-PL
R = −0.018, p = 0.34
R = −0.061, p = 0.0017
Phospholipids in M-HDL (mmol/L)
M-HDL-L
R = −0.022, p = 0.26
R = −0.064, p = 0.001
Total Lipids in M-HDL (mol/L)
M-HDL-C
R = −0.024, p = 0.22
R = −0.066, p = 0.00064
Total Cholesterol in M-HDL (mmol/L)
M-HDL-CE
R = −0.025, p = 0.2
R = −0.066, p = 0.00066
Cholesterol Esters in M-HDL (mmol/L)
M-HDL-FC
R = −0.02, p = 0.31
R = −0.065, p = 0.00077
Free Cholesterol in M-HDL (mmol/L)
S-HDL-P
R = −0.02, p = 0.29
R = −0.05, p = 0.0099
Concentration of S-HDL (mol/L)
S-HDL-L
R = −0.021, p = 0.29
R = −0.052, p = 0.0074
Total Lipids in S-HDL (mmol/L)
S-HDL-PL
R = −0.026, p = 0.18
R = −0.044, p = 0.022
Phospholipids in S-HDL (mmol/L)
S-HDL-C
R = −0.0092, p = 0.63
R = −0.051, p = 0.008
Total Cholesterol in S-HDL (mmol/L)
S-HDL-CE
R = −0.0031, p = 0.87
R = −0.043, p = 0.027
Cholesterol Esters in S-HDL (mmol/L)
S-HDL-FC
R = −0.026, p = 0.18
R = −0.052, p = 0.0072
Free Cholesterol in S-HDL (mmol/L)
XXL-VLDL-CE
R = −0.033, p = 0.51
R = 0.031, p = 0.032
Cholesterol Esters in Chylomicrons and XXL-VLDL (mmol/L)
Fasting Glucose

**B**


S-VLDL-CE
R = −0.071, p = 0.15
R = 0.029, p = 0.042
Cholesterol Esters in S-VLDL (mmol/L)
XS-VLDL-P
R = −0.074, p = 0.13
R = 0.029, p = 0.045
Concentration of XS-VLDL (mol/L)
XS-VLDL-L
R = −0.074, p = 0.13
R = 0.029, p = 0.043
Total Lipids in XS-VLDL (mmol/L)
XS-VLDL-C
R = −0.072, p = 0.14
R = 0.032, p = 0.026
Total Cholesterol in XS-VLDL (mmol/L)
XS-VLDL-CE
R = −0.071, p = 0.15
R = 0.034, p = 0.016
Cholesterol Esters in XS-VLDL (mmol/L)
Total Cholesterol in HDL2
R = 0.011, p = 0.82
R = −0.028, p = 0.047
Total Cholesterol in HDL2 (mmol/L)
Mean Diameter for LDL particles
R = 0.011, p = 0.82
R = 0.043, p = 0.0028
Mean Diameter for LDL particles (nm)
Lactate
R = −0.0089, p = 0.86
R = −0.046, p = 0.0012
Lactate (mmol/L)
3-Hydroxybutyrate
R = −0.013, p = 0.8
R = 0.04, p = 0.0048
3-Hydroxybutyrate (mmol/L)
M-HDL-P
R = −0.011, p = 0.82
R = −0.052, p = 0.00025
Concentration of M-HDL (mol/L)
M-HDL-L
R = −0.01, p = 0.84
R = −0.052, p = 0.00023
Total Lipids in M-HDL (mol/L)
M-HDL-PL
R = −0.015, p = 0.77
R = −0.048, p = 7e-04
Phospholipids in M-HDL (mmol/L)
M-HDL-C
R = −0.00014, p = 1
R = −0.055, p = 0.00011
Total Cholesterol in M-HDL (mmol/L)
M-HDL-CE
R = 0.0027, p = 0.96
R = −0.056, p = 8.3e-05
Cholesterol Esters in M-HDL (mmol/L)
M-HDL-FC
R = −0.011, p = 0.83
R = −0.051, p = 0.00036
Free Cholesterol in M-HDL (mmol/L)
S-HDL-P
R = −0.041, p = 0.41
R = −0.044, p = 0.0018
Concentration of S-HDL (mol/L)
S-HDL-L
R = −0.04, p = 0.42
R = −0.045, p = 0.0014
Total Lipids in S-HDL (mmol/L)
S-HDL-PL
R = −0.017, p = 0.74
R = −0.048, p = 0.00082
Phospholipids in S-HDL (mmol/L)
S-HDL-C
R = −0.048, p = 0.33
R = −0.033, p = 0.022
Total Cholesterol in S-HDL (mmol/L)
S-HDL-FC
R = −0.025, p = 0.61
R = −0.05, p = 4e-04
Free Cholesterol in S-HDL (mmol/L)
Fasting Glucose

C

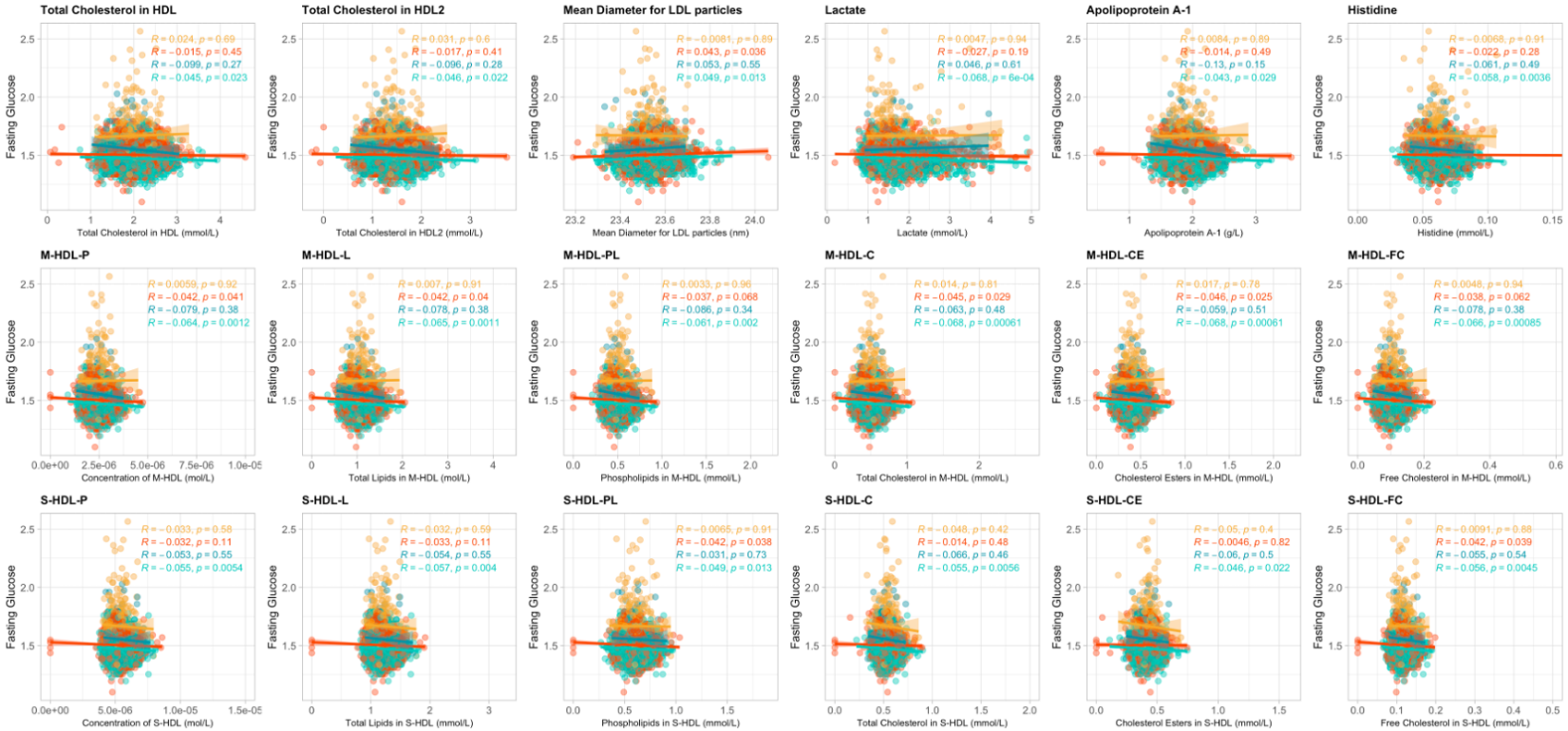



**Figure B.8: Metabolite values correlated with fasting glucose.**

A: Metabolite values with significant Pearson correlation coefficients with fasting glucose following stratification by ethnicity. Blue: South Asian. Red: White European. B: Metabolite values with a significant Pearson correlation coefficients with fasting glucose following stratification by case status. Pink: Cases. Green: Non-cases. C: Metabolite values with a significant Pearson correlation coefficients with fasting glucose following stratification by ethnicity and case status. Yellow: South Asian cases. Orange: South Asian non-cases. Blue: White European cases. Light blue: White European non-cases.

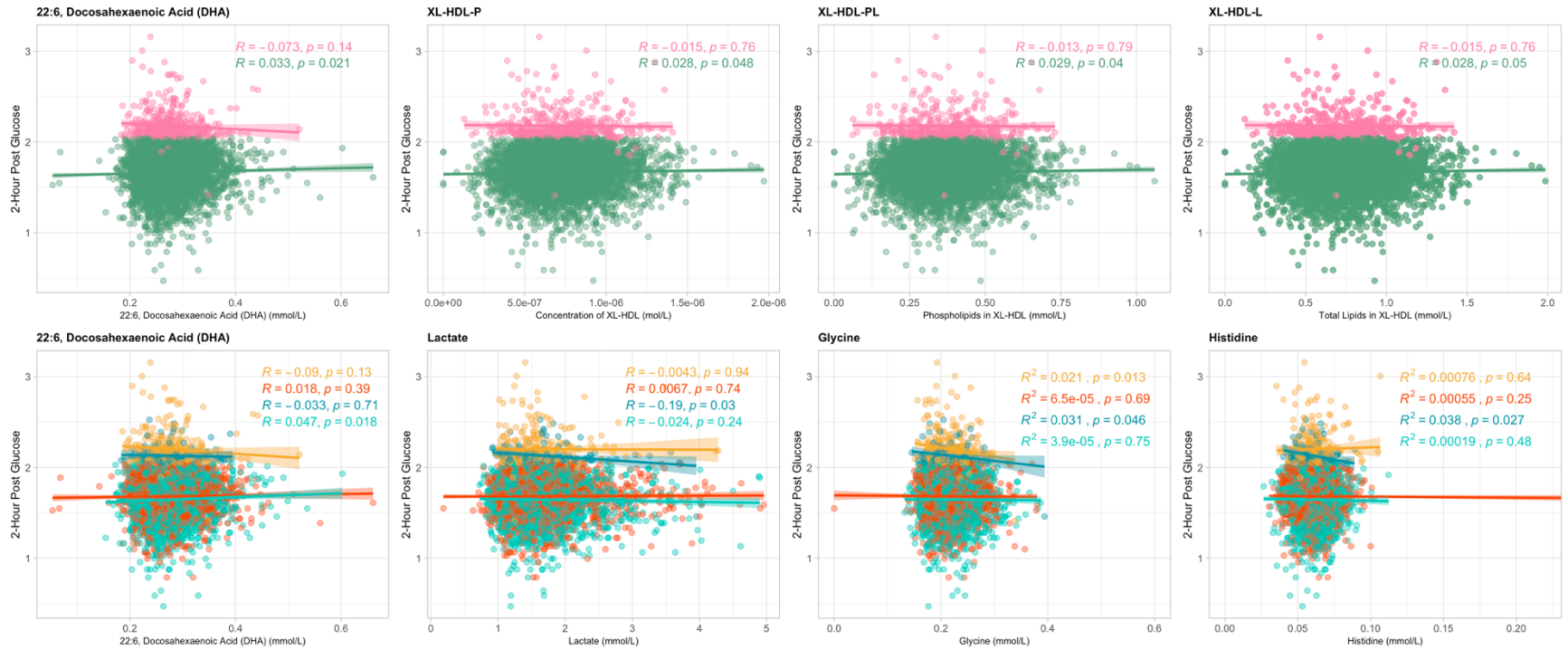



**Figure B.9: Metabolite values correlated with 2-hour post glucose.**

**Top row**: Metabolite values with a significantly correlated with 2-hour post glucose following stratification by case status. Pink: Cases. Green: Non-cases. **Bottom row**: Metabolite values with a significantly correlated with 2-hour post glucose following stratification by ethnicity and case status. Yellow: South Asian cases. Orange: South Asian non-cases. Blue: White European cases Light blue: White European non-cases. R= Pearson correlation coefficient

## Table B.6: Linear regression of fasting glucose.

| A | Overall (n=5538*) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Model 1 | | | Model 2 | | | Model 3 | | |
| Metabolite value | β | SE | P value | β | SE | P value | β | SE | P value |
| Lactate (mmol/l) | -0.008 | 0.003 | 0.003 | -0.009 | 0.003 | 0.001 | -0.009 | 0.003 | 0.001 |
| Mean Diameter for VLDL particles (nm) | 0.000 | 0.001 | 0.824 | | | | | | |
| Total Fatty Acids (mmol/l) | 0.000 | 0.000 | 0.718 | | | | | | |
| MUFA (mmol/l) | 0.001 | 0.001 | 0.507 | | | | | | |
| 18:2 Linoleic Acid (mmol/l) | 0.000 | 0.002 | 0.892 | | | | | | |
| SFA (mmol/l) | 0.000 | 0.001 | 0.770 | | | | | | |
| Esterified Cholesterol (mmol/l) | 0.001 | 0.002 | 0.706 | | | | | | |
| Analine (mmol/l) | -0.034 | 0.029 | 0.247 | | | | | | |
| Glutamine (mmol/l) | -0.036 | 0.031 | 0.243 | | | | | | |
| Total Serum Cholesterol | 0.000 | 0.001 | 0.725 | | | | | | |
| n-6 Fatty Acids (mmol/l) | 0.000 | 0.002 | 0.999 | | | | | | |
| PUFA (mmol/l) | 0.000 | 0.001 | 0.965 | | | | | | |
| Glycoprotein Acetyls (mmol/l) | 0.005 | 0.007 | 0.488 | | | | | | |
| Citrate (mmol/l) | -0.071 | 0.082 | 0.388 | | | | | | |
| Glycine (mmol/l) | -0.055 | 0.043 | 0.203 | | | | | | |
| Histidine (mmol/l) | -0.055 | 0.047 | 0.239 | | | | | | |
| Phenylanaline (mmol/l) | -0.016 | 0.032 | 0.624 | | | | | | |
| Tyrosine (mmol/l) | -0.030 | 0.041 | 0.461 | | | | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Apolipoprotein A1 (g/l)** | -0.009 | 0.006 | 0.166 | | | | | | |
| **LDL Cholesterol (mmol/l)** | 0.002 | 0.002 | 0.514 | | | | | | |
| **HDL Cholesterol (mmol/l)** | -0.006 | 0.004 | 0.142 | | | | | | |
| **HDL2 Cholesterol (mmol/l)** | -0.006 | 0.004 | 0.134 | | | | | | |
| **HDL3 Cholesterol (mmol/l)** | -0.030 | 0.032 | 0.343 | | | | | | |
| **n-3 Fatty Acids (mmol/l)** | 0.005 | 0.013 | 0.694 | | | | | | |
| **DHA (mmol/l)** | 0.031 | 0.031 | 0.230 | | | | | | |
| **Creatine (mmol/l)** | -0.368 | 0.226 | 0.103 | | | | | | |
| **Albumin (Signal Area)** | -0.599 | 0.238 | 0.012 | -0.596 | 0.231 | 0.010 | -0.611 | 0.233 | 0.009 |
| **Serum triglycerides (mmol/l)** | 0.002 | 0.002 | 0.445 | | | | | | |
| **VLDL triglycerides (mmol/l)** | 0.002 | 0.003 | 0.537 | | | | | | |
| **Phosphoglycerides (mmol/l)** | -0.001 | 0.003 | 0.751 | | | | | | |
| **Phosphatidylchlorine (mmol/l)** | 0.001 | 0.003 | 0.861 | | | | | | |
| **Sphingomyelins (mmol/l)** | -0.002 | 0.016 | 0.897 | | | | | | |
| **Total Cholines (mmol/l)** | 0.000 | 0.003 | 0.991 | | | | | | |
| **Pyruvate (mmol/l)** | -0.008 | 0.064 | 0.899 | | | | | | |
| **Glycerol (mmol/l)** | 0.000 | 0.086 | 0.996 | | | | | | |
| **Beta- hydroxybutyrate (mmol/l)** | 0.068 | 0.035 | 0.057 | | | | | | |
| **Mean Diameter for LDL particles (nm)** | 0.051 | 0.022 | 0.020 | 0.047 | 0.021 | 0.027 | 0.049 | 0.021 | 0.022 |
| **Mean Diameter for HDL particles (nm)** | 0.005 | 0.007 | 0.474 | | | | | | |

| B | White European (n=2267)* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Model 1 | | | Model 2 | | | Model 3 | | |
| **Metabolite Measure** | **β** | **SE** | **P value** | **β** | **SE** | **P value** | **β** | **SE** | **P value** |
| **Lactate (mmol/l)** | -0.009 | 0.003 | 0.006 | -0.011 | 0.003 | 0.001 | -0.010 | 0.003 | 0.003 |
| **Mean Diameter for VLDL particles (nm)** | 0.001 | 0.002 | 0.545 | | | | | | |
| **Total Fatty Acids (mmol/l)** | 0.000 | 0.001 | 0.926 | | | | | | |
| **MUFA (mmol/l)** | 0.000 | 0.002 | 0.917 | | | | | | |
| **18:2 Linoleic Acid (mmol/l)** | -0.001 | 0.002 | 0.784 | | | | | | |
| **SFA (mmol/l)** | 0.000 | 0.001 | 0.995 | | | | | | |
| **Esterified Cholesterol (mmol/l)** | -0.001 | 0.002 | 0.575 | | | | | | |
| **Analine (mmol/l)** | -0.014 | 0.034 | 0.674 | | | | | | |
| **Glutamine (mmol/l)** | -0.036 | 0.036 | 0.315 | | | | | | |
| **Total Serum Cholesterol** | -0.001 | 0.001 | 0.580 | | | | | | |
| **n-6 Fatty Acids (mmol/l)** | -0.001 | 0.002 | 0.652 | | | | | | |
| **PUFA (mmol/l)** | -0.001 | 0.002 | 0.698 | | | | | | |
| **Glycoprotein Acetyls (mmol/l)** | 0.011 | 0.009 | 0.211 | | | | | | |
| **Citrate (mmol/l)** | -0.087 | 0.095 | 0.362 | | | | | | |
| **Glycine (mmol/l)** | 0.060 | 0.050 | 0.225 | | | | | | |
| **Histidine (mmol/l)** | -0.428 | 0.173 | 0.014 | -0.459 | 0.167 | 0.006 | -0.449 | 0.170 | 0.008 |
| **Phenylanaline (mmol/l)** | -0.065 | 0.151 | 0.666 | | | | | | |
| **Tyrosine (mmol/l)** | -0.373 | 0.303 | 0.219 | | | | | | |
| **Apolipoprotein A1 (g/l)** | -0.018 | 0.007 | 0.014 | -0.017 | 0.007 | 0.019 | -0.018 | 0.007 | 0.014 |
| **LDL Cholesterol (mmol/l)** | 0.000 | 0.003 | 0.955 | | | | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **HDL Cholesterol (mmol/l)** | -0.011 | 0.004 | 0.012 | -0.010 | 0.004 | 0.019 | -0.011 | 0.004 | 0.011 |
| **HDL2 Cholesterol (mmol/l)** | -0.013 | 0.005 | 0.011 | -0.011 | 0.005 | 0.019 | -0.012 | 0.005 | 0.010 |
| **HDL3 Cholesterol (mmol/l)** | -0.066 | 0.037 | 0.077 | | | | | | |
| **n-3 Fatty Acids (mmol/l)** | 0.003 | 0.015 | 0.830 | | | | | | |
| **DHA (mmol/l)** | 0.042 | 0.036 | 0.249 | | | | | | |
| **Creatine (mmol/l)** | -0.166 | 0.277 | 0.548 | | | | | | |
| **Albumin (Signal Area)** | -0.387 | 0.276 | 0.160 | | | | | | |
| **Serum triglycerides (mmol/l)** | 0.002 | 0.003 | 0.433 | | | | | | |
| **VLDL triglycerides (mmol/l)** | 0.003 | 0.004 | 0.416 | | | | | | |
| **Phosphoglycerides (mmol/l)** | -0.004 | 0.004 | 0.313 | | | | | | |
| **Phosphatidylchlorine (mmol/l)** | -0.002 | 0.003 | 0.548 | | | | | | |
| **Sphingomyelins (mmol/l)** | -0.016 | 0.019 | 0.381 | | | | | | |
| **Total Cholines (mmol/l)** | -0.002 | 0.003 | 0.458 | | | | | | |
| **Pyruvate (mmol/l)** | 0.113 | 0.075 | 0.134 | | | | | | |
| **Glycerol (mmol/l)** | 0.075 | 0.101 | 0.457 | | | | | | |
| **Beta- hydroxybutyrate (mmol/l)** | 0.070 | 0.039 | 0.074 | | | | | | |
| **Mean Diameter for LDL particles (nm)** | 0.061 | 0.025 | 0.016 | 0.055 | 0.024 | 0.025 | 0.056 | 0.025 | 0.024 |
| **Mean Diameter for HDL particles (nm)** | -0.002 | 0.008 | 0.783 | | | | | | |

**C**

| | South Asian (n=2671) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Model 1 | | | Model 2 | | | Model 3 | | |
| **Metabolite Measure** | **β** | **SE** | **P value** | **β** | **SE** | **P value** | **β** | **SE** | **P value** |
| **Lactate (mmol/l)** | -0.007 | 0.004 | 0.105 | | | | | | |
| **Mean Diameter for VLDL particles (nm)** | -0.001 | 0.002 | 0.561 | | | | | | |
| **Total Fatty Acids (mmol/l)** | 0.000 | 0.001 | 0.605 | | | | | | |
| **MUFA (mmol/l)** | 0.002 | 0.002 | 0.435 | | | | | | |
| **18:2 Linoleic Acid (mmol/l)** | 0.001 | 0.003 | 0.683 | | | | | | |
| **SFA (mmol/l)** | 0.001 | 0.002 | 0.726 | | | | | | |
| **Esterified Cholesterol (mmol/l)** | 0.002 | 0.003 | 0.510 | | | | | | |
| **Analine (mmol/l)** | -0.042 | 0.046 | 0.360 | | | | | | |
| **Glutamine (mmol/l)** | -0.024 | 0.049 | 0.619 | | | | | | |
| **Total Serum Cholesterol** | 0.001 | 0.002 | 0.526 | | | | | | |
| **n-6 Fatty Acids (mmol/l)** | 0.001 | 0.003 | 0.752 | | | | | | |
| **PUFA (mmol/l)** | 0.001 | 0.002 | 0.745 | | | | | | |
| **Glycoprotein Acetyls (mmol/l)** | 0.003 | 0.011 | 0.772 | | | | | | |
| **Citrate (mmol/l)** | -0.031 | 0.131 | 0.813 | | | | | | |
| **Glycine (mmol/l)** | -0.029 | 0.069 | 0.676 | | | | | | |
| **Histidine (mmol/l)** | -0.044 | 0.054 | 0.417 | | | | | | |
| **Phenylanaline (mmol/l)** | -0.017 | 0.036 | 0.633 | | | | | | |
| **Tyrosine (mmol/l)** | -0.031 | 0.046 | 0.498 | | | | | | |
| **Apolipoprotein A1 (g/l)** | -0.002 | 0.010 | 0.875 | | | | | | |
| **LDL Cholesterol (mmol/l)** | 0.002 | 0.004 | 0.501 | | | | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **HDL Cholesterol (mmol/l)** | -0.001 | 0.006 | 0.824 | | | | | | |
| **HDL2 Cholesterol (mmol/l)** | -0.002 | 0.007 | 0.808 | | | | | | |
| **HDL3 Cholesterol (mmol/l)** | -0.001 | 0.051 | 0.983 | | | | | | |
| **n-3 Fatty Acids (mmol/l)** | 0.007 | 0.020 | 0.745 | | | | | | |
| **DHA (mmol/l)** | 0.035 | 0.049 | 0.474 | | | | | | |
| **Creatine (mmol/l)** | -0.619 | 0.343 | 0.071 | | | | | | |
| **Albumin (Signal Area)** | -0.866 | 0.382 | 0.024 | -0.818 | 0.366 | 0.025 | -0.858 | 0.375 | 0.022 |
| **Serum triglycerides (mmol/l)** | 0.002 | 0.004 | 0.660 | | | | | | |
| **VLDL triglycerides (mmol/l)** | 0.001 | 0.005 | 0.792 | | | | | | |
| **Phosphoglycerides (mmol/l)** | 0.002 | 0.005 | 0.766 | | | | | | |
| **Phosphatidylchlorine (mmol/l)** | 0.003 | 0.005 | 0.576 | | | | | | |
| **Sphingomyelins (mmol/l)** | 0.007 | 0.025 | 0.790 | | | | | | |
| **Total Cholines (mmol/l)** | 0.002 | 0.004 | 0.645 | | | | | | |
| **Pyruvate (mmol/l)** | -0.115 | 0.101 | 0.254 | | | | | | |
| **Glycerol (mmol/l)** | -0.080 | 0.137 | 0.559 | | | | | | |
| **Beta- hydroxybutyrate (mmol/l)** | 0.073 | 0.060 | 0.225 | | | | | | |
| **Mean Diameter for LDL particles (nm)** | 0.049 | 0.035 | 0.158 | | | | | | |
| **Mean Diameter for HDL particles (nm)** | 0.009 | 0.011 | 0.428 | | | | | | |

β coefficients of metabolite measures from linear regression models predicting fasting glucose (mmol/L). * 1 White European sample had a missing fasting glucose measure and was excluded from the analysis. Model 1 adjusted for maternal age, gestational age, parity and smoking status during pregnancy (yes/no). When β coefficients were significant (p value ≤ 0.05) models were additionally adjusted for BMI as a continuous variable (Model 2) and BMI as a dichotomous variable, grouping mothers based on whether they were above their ethnic specific cut off for overweight on the BMI scale. (Model 3) **A**: Overall sample results. **B**: White European. **C**: South Asian.

## Table B.7: Linear regression of 2-hour post glucose.

| A | Overall (n=5538*) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Model 1 | | | Model 2 | | | Model 3 | | |
| **Metabolite Measure** | **β** | **SE** | **P value** | **β** | **SE** | **P value** | **β** | **SE** | **P value** |
| **Lactate (mmol/l)** | -0.008 | 0.007 | 0.246 | | | | | | |
| **Mean Diameter for VLDL particles (nm)** | 0.000 | 0.003 | 0.996 | | | | | | |
| **Total Fatty Acids (mmol/l)** | 0.001 | 0.001 | 0.285 | | | | | | |
| **MUFA (mmol/l)** | 0.004 | 0.003 | 0.239 | | | | | | |
| **18:2 Linoleic Acid (mmol/l)** | 0.001 | 0.004 | 0.808 | | | | | | |
| **SFA (mmol/l)** | 0.004 | 0.003 | 0.221 | | | | | | |
| **Esterified Cholesterol (mmol/l)** | 0.005 | 0.004 | 0.243 | | | | | | |
| **Analine (mmol/l)** | 0.009 | 0.068 | 0.899 | | | | | | |
| **Glutamine (mmol/l)** | -0.052 | 0.072 | 0.471 | | | | | | |
| **Total Serum Cholesterol** | 0.003 | 0.003 | 0.267 | | | | | | |
| **n-6 Fatty Acids (mmol/l)** | 0.002 | 0.004 | 0.624 | | | | | | |
| **PUFA (mmol/l)** | 0.002 | 0.003 | 0.566 | | | | | | |
| **Glycoprotein Acetyls (mmol/l)** | 0.006 | 0.017 | 0.710 | | | | | | |
| **Citrate (mmol/l)** | 0.160 | 0.192 | 0.405 | | | | | | |
| **Glycine (mmol/l)** | -0.164 | 0.101 | 0.103 | | | | | | |
| **Histidine (mmol/l)** | -0.094 | 0.110 | 0.391 | | | | | | |
| **Phenylanaline (mmol/l)** | -0.048 | 0.074 | 0.521 | | | | | | |
| **Tyrosine (mmol/l)** | -0.074 | 0.095 | 0.437 | | | | | | |
| **Apolipoprotein A1 (g/l)** | 0.018 | 0.015 | 0.238 | | | | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **LDL Cholesterol (mmol/l)** | 0.004 | 0.005 | 0.479 | | | | | | |
| **HDL Cholesterol (mmol/l)** | 0.010 | 0.009 | 0.253 | | | | | | |
| **HDL2 Cholesterol (mmol/l)** | 0.010 | 0.010 | 0.289 | | | | | | |
| **HDL3 Cholesterol (mmol/l)** | 0.114 | 0.075 | 0.129 | | | | | | |
| **n-3 Fatty Acids (mmol/l)** | 0.033 | 0.030 | 0.276 | | | | | | |
| **DHA (mmol/l)** | 0.129 | 0.072 | 0.075 | | | | | | |
| **Creatine (mmol/l)** | 0.108 | 0.528 | 0.837 | | | | | | |
| **Albumin (Signal Area)** | -0.773 | 0.165 | 0.165 | | | | | | |
| **Serum triglycerides (mmol/l)** | 0.003 | 0.006 | 0.628 | | | | | | |
| **VLDL triglycerides (mmol/l)** | 0.001 | 0.007 | 0.931 | | | | | | |
| **Phosphoglycerides (mmol/l)** | 0.011 | 0.008 | 0.171 | | | | | | |
| **Phosphatidylchlorine (mmol/l)** | 0.009 | 0.007 | 0.199 | | | | | | |
| **Sphingomyelins (mmol/l)** | 0.031 | 0.037 | 0.408 | | | | | | |
| **Total Cholines (mmol/l)** | 0.009 | 0.007 | 0.164 | | | | | | |
| **Pyruvate (mmol/l)** | 0.110 | 0.150 | 0.463 | | | | | | |
| **Glycerol (mmol/l)** | 0.112 | 0.201 | 0.578 | | | | | | |
| **Beta- hydroxybutyrate (mmol/l)** | 0.006 | 0.083 | 0.945 | | | | | | |
| **Mean Diameter for LDL particles (nm)** | 0.058 | 0.051 | 0.253 | | | | | | |
| **Mean Diameter for HDL particles (nm)** | 0.022 | 0.016 | 0.158 | | | | | | |

**B**

| | White European (n=2267)* | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Model 1 | | | Model 2 | | | Model 3 | | |
| **Metabolite Measure** | **β** | **SE** | **P value** | **β** | **SE** | **P value** | **β** | **SE** | **P value** |
| **Lactate (mmol/l)** | -0.008 | 0.009 | 0.380 | | | | | | |
| **Mean Diameter for VLDL particles (nm)** | 0.001 | 0.004 | 0.842 | | | | | | |
| **Total Fatty Acids (mmol/l)** | 0.002 | 0.002 | 0.181 | | | | | | |
| **MUFA (mmol/l)** | 0.005 | 0.004 | 0.249 | | | | | | |
| **18:2 Linoleic Acid (mmol/l)** | 0.005 | 0.006 | 0.434 | | | | | | |
| **SFA (mmol/l)** | 0.006 | 0.004 | 0.118 | | | | | | |
| **Esterified Cholesterol (mmol/l)** | 0.010 | 0.006 | 0.075 | | | | | | |
| **Analine (mmol/l)** | 0.080 | 0.093 | 0.390 | | | | | | |
| **Glutamine (mmol/l)** | 0.021 | 0.098 | 0.833 | | | | | | |
| **Total Serum Cholesterol** | 0.007 | 0.004 | 0.091 | | | | | | |
| **n-6 Fatty Acids (mmol/l)** | 0.005 | 0.005 | 0.351 | | | | | | |
| **PUFA (mmol/l)** | 0.005 | 0.005 | 0.298 | | | | | | |
| **Glycoprotein Acetyls (mmol/l)** | 0.009 | 0.023 | 0.692 | | | | | | |
| **Citrate (mmol/l)** | -0.164 | 0.260 | 0.527 | | | | | | |
| **Glycine (mmol/l)** | -0.110 | 0.135 | 0.417 | | | | | | |
| **Histidine (mmol/l)** | -0.213 | 0.472 | 0.652 | | | | | | |
| **Phenylanaline (mmol/l)** | 0.011 | 0.411 | 0.979 | | | | | | |
| **Tyrosine (mmol/l)** | -0.598 | 0.824 | 0.468 | | | | | | |
| **Apolipoprotein A1 (g/l)** | 0.023 | 0.020 | 0.268 | | | | | | |
| **LDL Cholesterol (mmol/l)** | 0.011 | 0.007 | 0.126 | | | | | | |
| **HDL Cholesterol (mmol/l)** | 0.009 | 0.012 | 0.474 | | | | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **HDL2 Cholesterol (mmol/l)** | 0.008 | 0.013 | 0.534 | | | | | | |
| **HDL3 Cholesterol (mmol/l)** | 0.128 | 0.101 | 0.205 | | | | | | |
| **n-3 Fatty Acids (mmol/l)** | 0.069 | 0.041 | 0.097 | | | | | | |
| **DHA (mmol/l)** | 0.196 | 0.098 | 0.047 | 0.203 | 0.097 | 0.037 | 0.195 | 0.097 | 0.045 |
| **Creatine (mmol/l)** | -0.302 | 0.753 | 0.688 | | | | | | |
| **Albumin (Signal Area)** | 0.294 | 0.750 | 0.695 | | | | | | |
| **Serum triglycerides (mmol/l)** | 0.006 | 0.008 | 0.459 | | | | | | |
| **VLDL triglycerides (mmol/l)** | 0.005 | 0.010 | 0.630 | | | | | | |
| **Phosphoglycerides (mmol/l)** | 0.014 | 0.011 | 0.190 | | | | | | |
| **Phosphatidylchlorine (mmol/l)** | 0.013 | 0.009 | 0.167 | | | | | | |
| **Sphingomyelins (mmol/l)** | 0.074 | 0.051 | 0.151 | | | | | | |
| **Total Cholines (mmol/l)** | 0.013 | 0.009 | 0.147 | | | | | | |
| **Pyruvate (mmol/l)** | 0.193 | 0.205 | 0.348 | | | | | | |
| **Glycerol (mmol/l)** | 0.172 | 0.273 | 0.530 | | | | | | |
| **Beta- hydroxybutyrate (mmol/l)** | -0.077 | 0.106 | 0.468 | | | | | | |
| **Mean Diameter for LDL particles (nm)** | -0.077 | 0.106 | 0.468 | | | | | | |
| **Mean Diameter for HDL particles (nm)** | 0.021 | 0.021 | 0.328 | | | | | | |

**C**

| | South Asian (n=2671) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Model 1 | | | Model 2 | | | Model 3 | | |
| **Metabolite Measure** | **β** | **SE** | **P value** | **β** | **SE** | **P value** | **β** | **SE** | **P value** |
| **Lactate (mmol/l)** | -0.007 | 0.009 | 0.479 | | | | | | |
| **Mean Diameter for VLDL particles (nm)** | 0.000 | 0.005 | 0.979 | | | | | | |
| **Total Fatty Acids (mmol/l)** | 0.000 | 0.002 | 0.793 | | | | | | |
| **MUFA (mmol/l)** | 0.003 | 0.005 | 0.547 | | | | | | |
| **18:2 Linoleic Acid (mmol/l)** | -0.002 | 0.006 | 0.718 | | | | | | |
| **SFA (mmol/l)** | 0.001 | 0.004 | 0.790 | | | | | | |
| **Esterified Cholesterol (mmol/l)** | -0.001 | 0.006 | 0.862 | | | | | | |
| **Analine (mmol/l)** | -0.043 | 0.098 | 0.665 | | | | | | |
| **Glutamine (mmol/l)** | -0.111 | 0.104 | 0.285 | | | | | | |
| **Total Serum Cholesterol** | -0.001 | 0.004 | 0.875 | | | | | | |
| **n-6 Fatty Acids (mmol/l)** | -0.001 | 0.006 | 0.856 | | | | | | |
| **PUFA (mmol/l)** | -0.001 | 0.005 | 0.869 | | | | | | |
| **Glycoprotein Acetyls (mmol/l)** | 0.010 | 0.024 | 0.690 | | | | | | |
| **Citrate (mmol/l)** | 0.511 | 0.281 | 0.069 | | | | | | |
| **Glycine (mmol/l)** | -0.196 | 0.148 | 0.187 | | | | | | |
| **Histidine (mmol/l)** | -0.101 | 0.116 | 0.384 | | | | | | |
| **Phenylanaline (mmol/l)** | -0.055 | 0.077 | 0.481 | | | | | | |
| **Tyrosine (mmol/l)** | -0.076 | 0.098 | 0.438 | | | | | | |
| **Apolipoprotein A1 (g/l)** | 0.010 | 0.022 | 0.652 | | | | | | |
| **LDL Cholesterol (mmol/l)** | -0.004 | 0.008 | 0.586 | | | | | | |
| **HDL Cholesterol (mmol/l)** | 0.010 | 0.013 | 0.458 | | | | | | |
| **HDL2 Cholesterol (mmol/l)** | 0.010 | 0.014 | 0.479 | | | | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **HDL3 Cholesterol (mmol/l)** | 0.090 | 0.109 | 0.408 | | | | | | |
| **n-3 Fatty Acids (mmol/l)** | 0.000 | 0.044 | 0.992 | | | | | | |
| **DHA (mmol/l)** | 0.071 | 0.105 | 0.497 | | | | | | |
| **Creatine (mmol/l)** | 0.365 | 0.734 | 0.619 | | | | | | |
| **Albumin (Signal Area)** | -1.900 | 0.816 | 0.020 | -1.824 | 0.797 | 0.022 | -1.887 | 0.808 | 0.020 |
| **Serum triglycerides (mmol/l)** | 0.001 | 0.008 | 0.940 | | | | | | |
| **VLDL triglycerides (mmol/l)** | -0.002 | 0.010 | 0.837 | | | | | | |
| **Phosphoglycerides (mmol/l)** | 0.008 | 0.012 | 0.504 | | | | | | |
| **Phosphatidylchlorine (mmol/l)** | 0.004 | 0.010 | 0.661 | | | | | | |
| **Sphingomyelins (mmol/l)** | -0.016 | 0.054 | 0.767 | | | | | | |
| **Total Cholines (mmol/l)** | 0.005 | 0.010 | 0.583 | | | | | | |
| **Pyruvate (mmol/l)** | 0.046 | 0.216 | 0.832 | | | | | | |
| **Glycerol (mmol/l)** | 0.042 | 0.292 | 0.885 | | | | | | |
| **Beta- hydroxybutyrate (mmol/l)** | 0.118 | 0.128 | 0.355 | | | | | | |
| **Mean Diameter for LDL particles (nm)** | 0.097 | 0.075 | 0.194 | | | | | | |
| **Mean Diameter for HDL particles (nm)** | 0.019 | 0.023 | 0.416 | | | | | | |

β coefficients of metabolite measures from linear regression models of 2-hour post glucose (mmol/L). * 1 White European sample had a missing fasting glucose measure and was excluded from the analysis. Model 1 adjusted for maternal age, gestational age, parity and smoking status during pregnancy (yes/no). When β coefficients were significant (p value ≤ 0.05) models were additionally adjusted for BMI as a continuous variable (Model 2) and BMI as a dichotomous variable, grouping mothers based on whether they were above their ethnic specific cut off for overweight on the BMI scale (Model 3). **A**: Overall sample results. **B**: White European. **C:** South Asian.

# Appendix C
## Supplementary data for Chapter 5

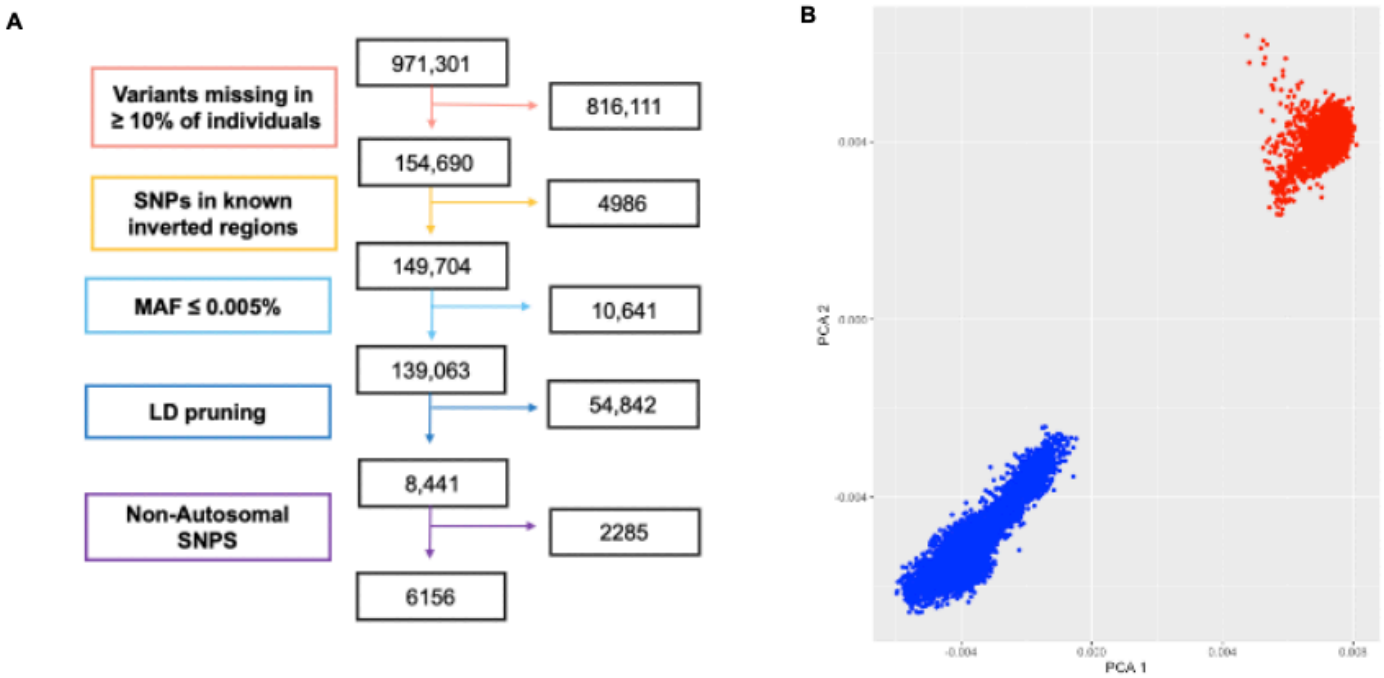



**Figure C.1: PCA to account for population stratification.**

**A**: Schematic of data processing steps taken prior to PCA analysis of genetic data. A R2 cut-off of 0.3 was utilised for LD pruning. During LD, pruning windows of 50 variants were checked for LD before moving the window 5bp. MAF: minor allele frequency. **B**: PCA plot of BiB genotype data calculated to account for population stratification. Blue: SAs. Red: WEs.

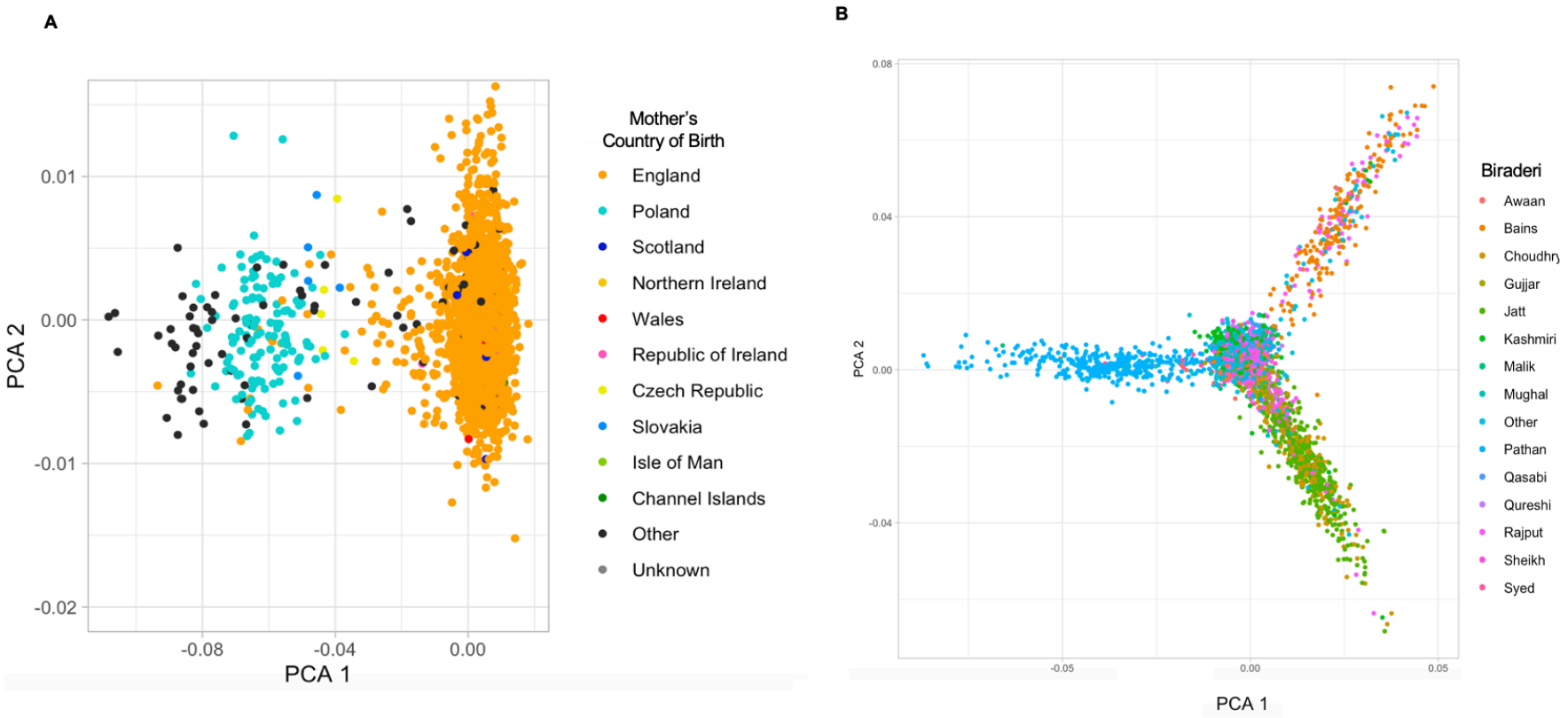



**Figure C.2: Population stratification within BiB.**

**A:** PCA of genetic data from retained WEs coloured based upon mother's country of birth. **B:** PCA of genetic data from retained SAs coloured based upon mother's Biraderi ('Brotherhood') membership.

**Table C.1: Regions of high 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. LD: linkage disequilibrium.

## Table C.2 : Proportion of outliers identified in each ethnicity.

| | South Asian | | White European | |
|---|---|---|---|---|
| **Metabolite** | **Number kept** | **% Lost** | **Number kept** | **% Lost** |
| **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 |

## Table C.3: Normalisation of metabolites.

| Transformation used | South Asian (%) | White European (%) |
|---|---|---|
| **None** | 2.1 | 1.4 |
| **Log** | 16.4 | 19.2 |
| **Square Root** | 69.2 | 74.6 |
| **Normal Score Transformation** | 12.3 | 4.8 |

Percentage of each metabolites normalised by each transformation method.

## Table C.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.

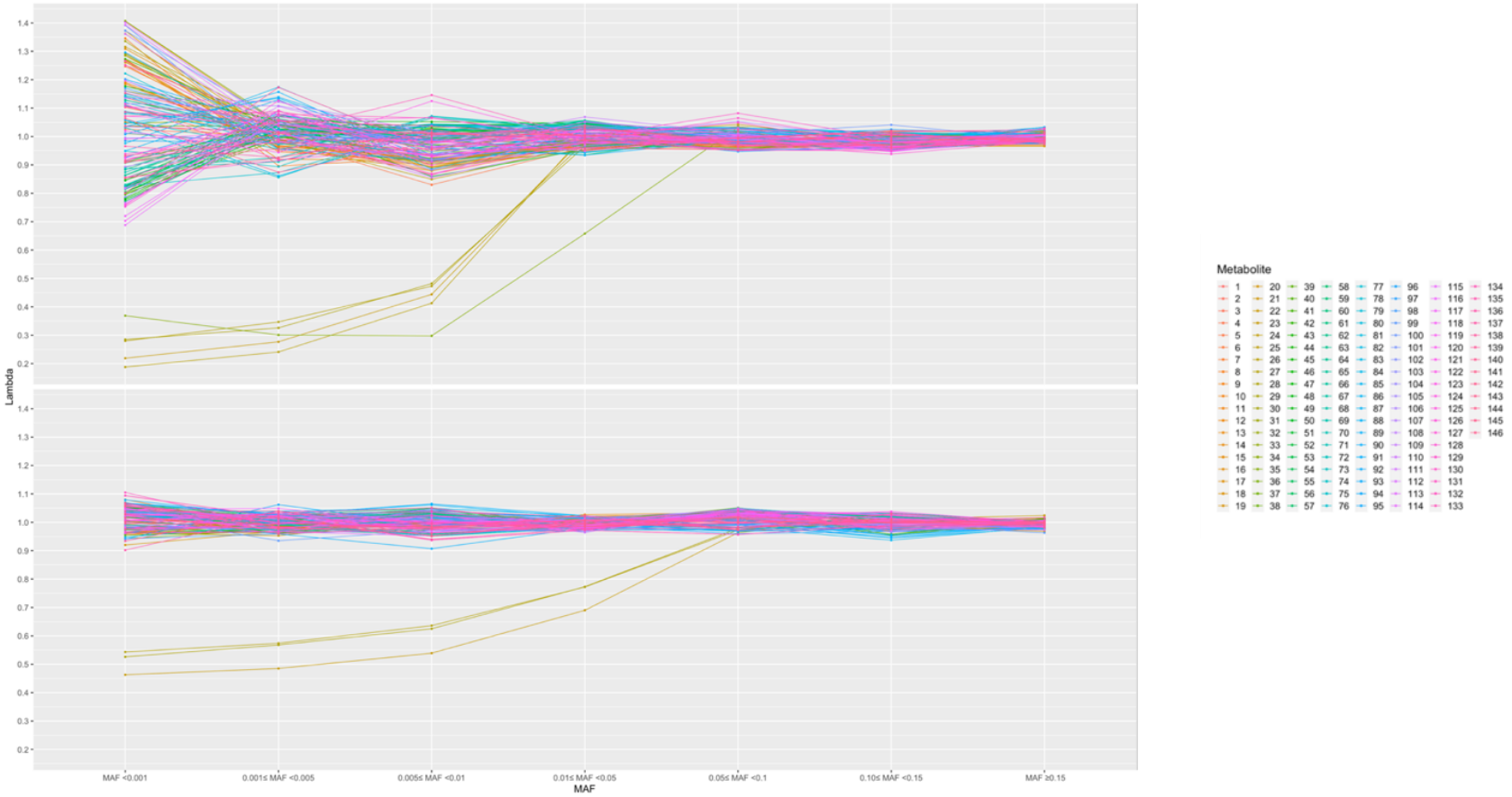



**Figure C.3: Plot of λ for each MAF cut-off by ethnicity.**

**Top**: λ values at varying MAF cut-offs in SAs. **Bottom**: λ values at varying MAF cut-offs in WEs. MAF: Minor allele frequency.

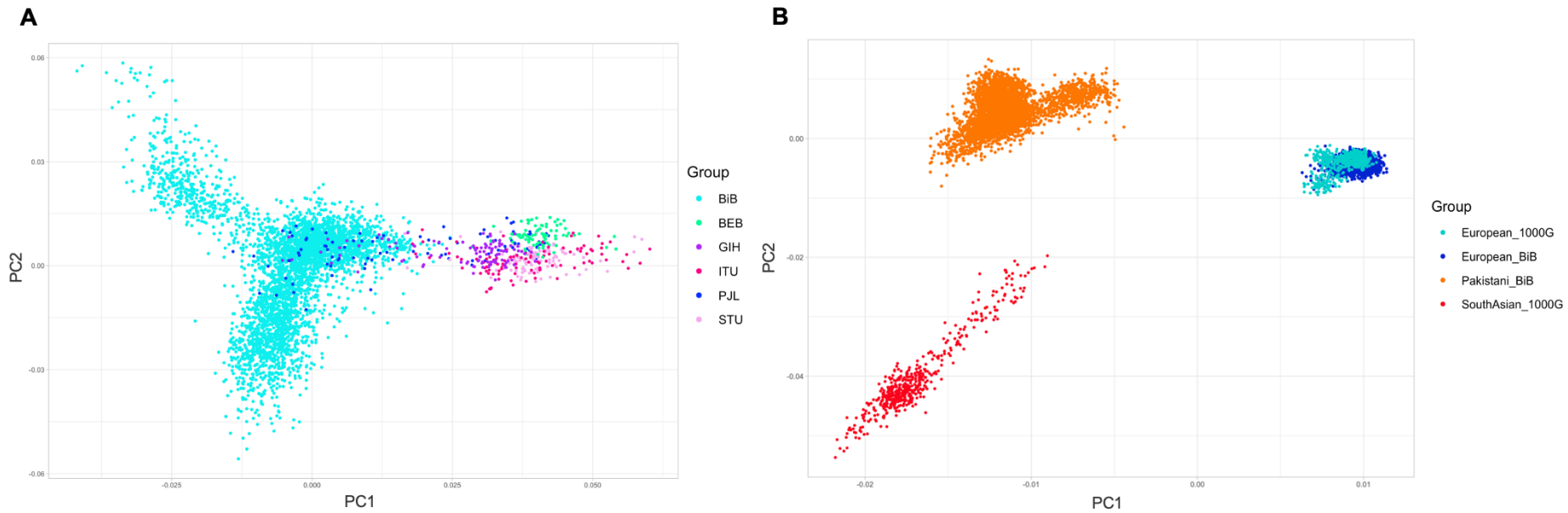



**Figure C.4: PCA comparing BiB data to all SA data from 1000G.**

**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.

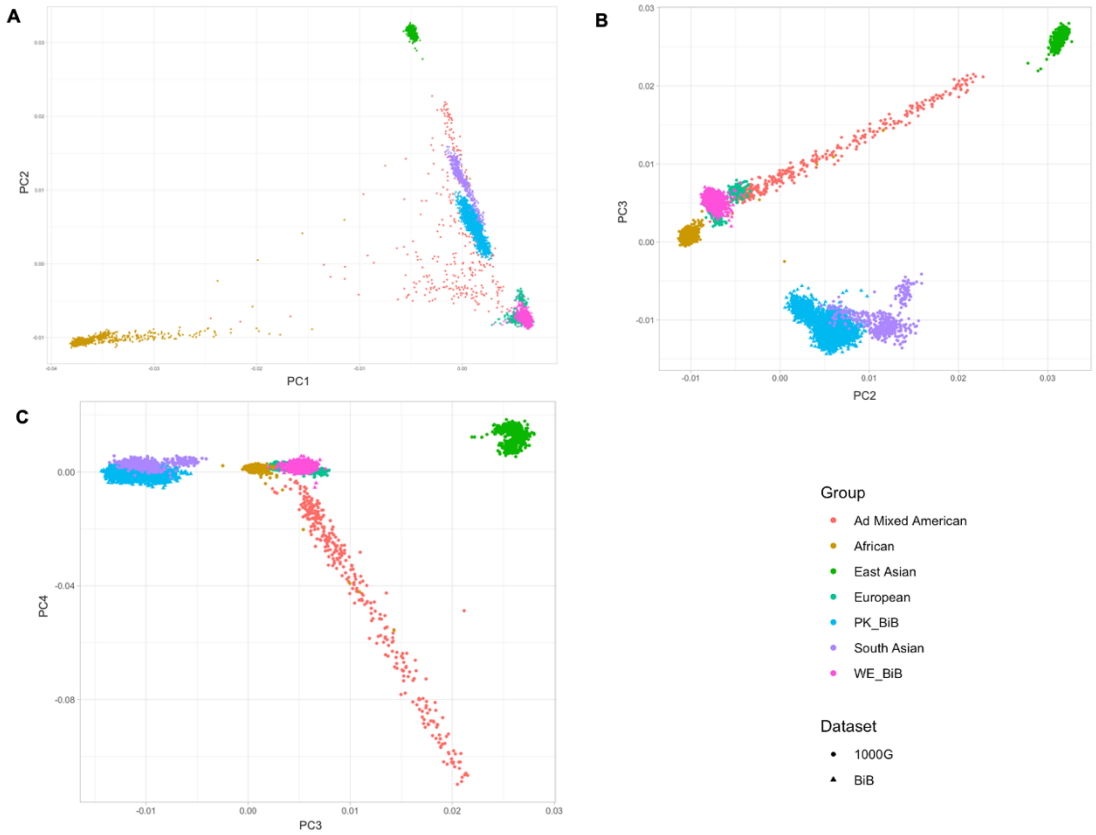



**Figure C.5: PCA comparing BiB data to all data from 1000G.**

**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.

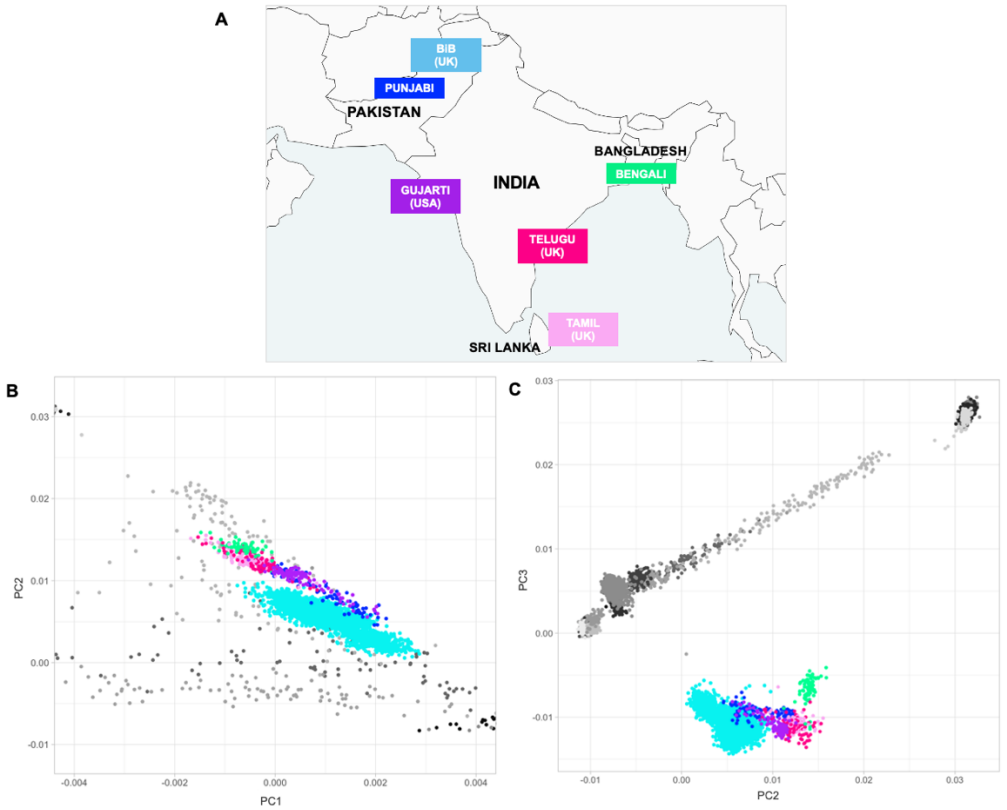



**Figure C.6: Modified Appendix Figure C.5 including the Pakistani BiB and SA 1000G populations.**

**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.

## Table C.5: Number of SNPs identified in the GWAS analyses before and after LD thinning.

| 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≤ $1 \times 10^{-5}$. Numbers in brackets represent the number of SNPs associated at the genome wide significant level (p value≤ $5 \times 10^{-8}$).

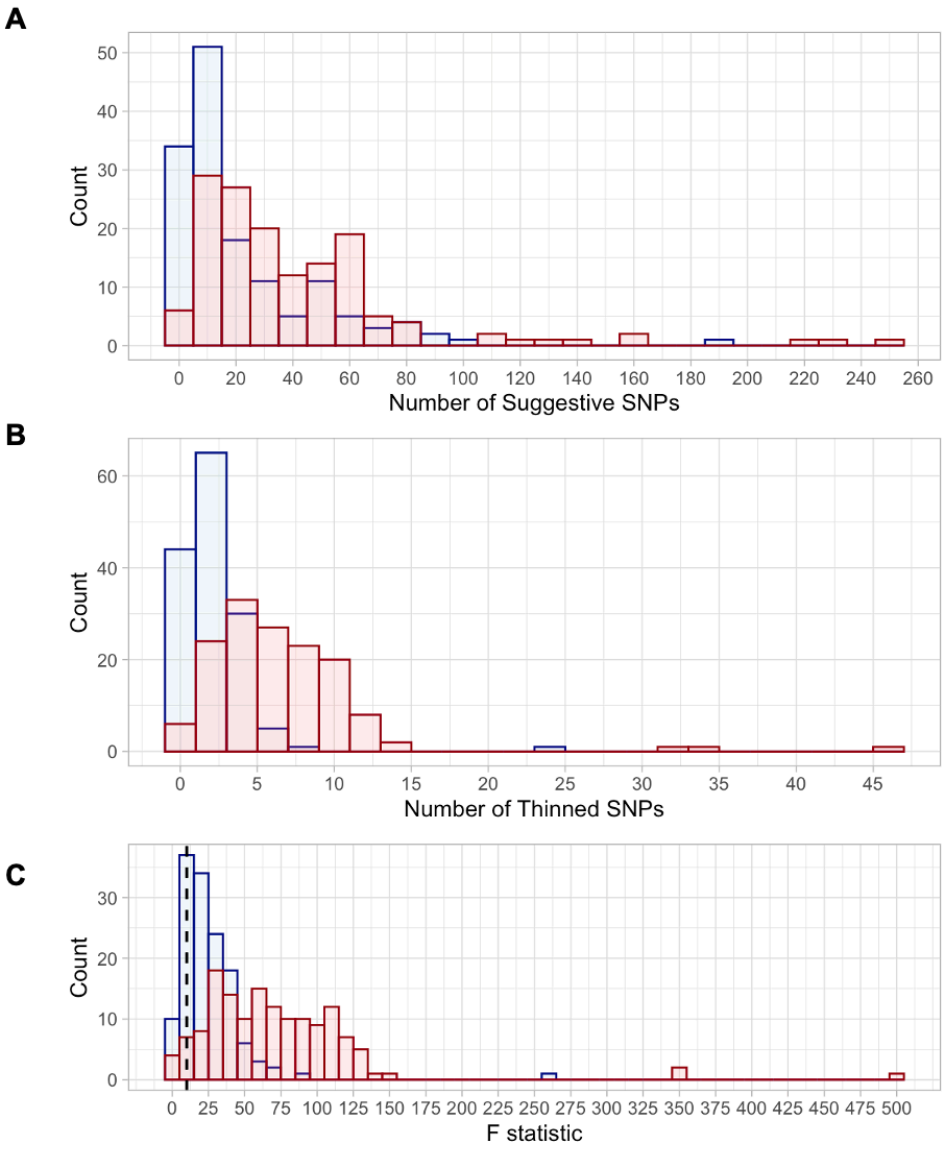



**Figure C.7: Histograms showing the number of SNPs and the strength of in each instrument.**

**A**: Histogram of number of SNPs identified for each metabolite at the suggestive level (p value 1 x $10^{-5}$). **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 F statistic of 10. An F statistic < 10 is an indicator of weak instrument bias. Blue: South Asians Red: White Europeans.

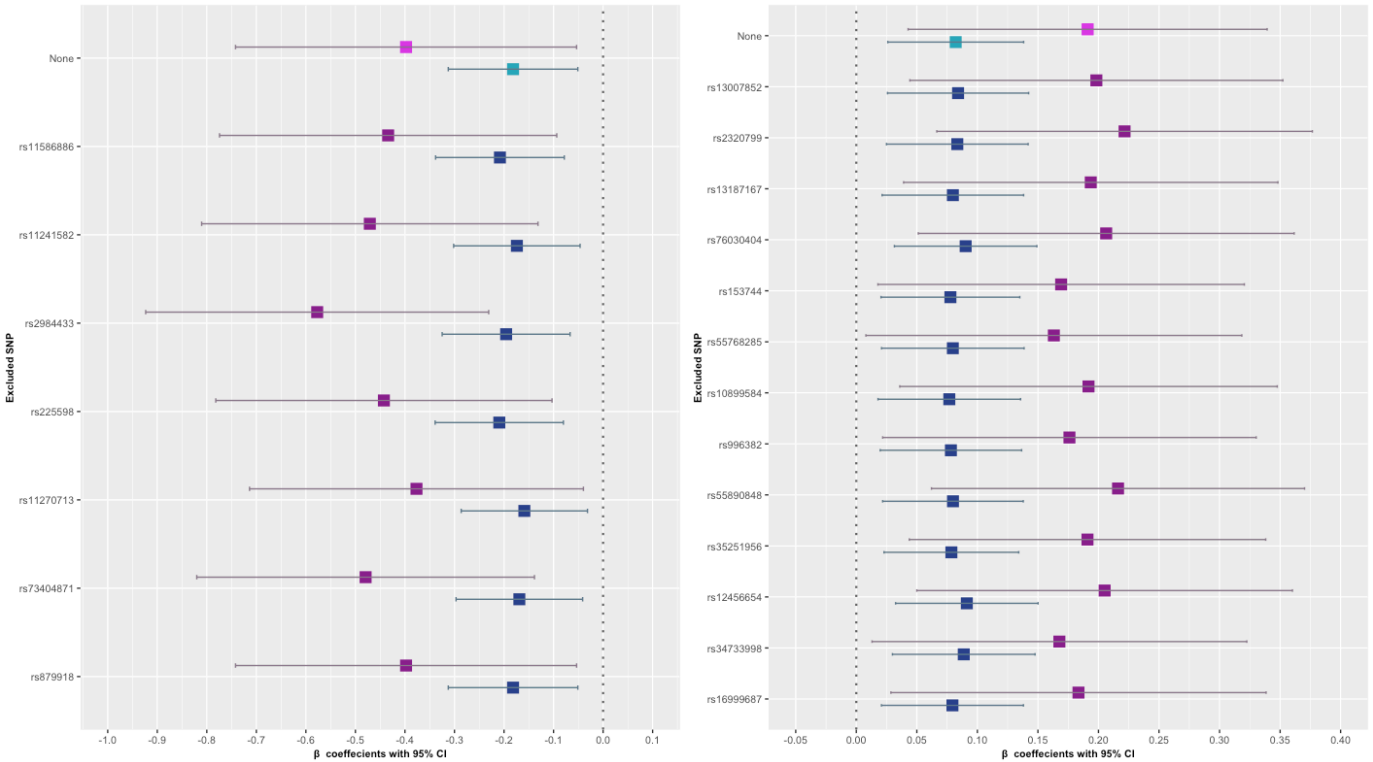



**Figure C.8: Forest plots of β values following leave-one-out analyses for identified associations in WEs.**

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.

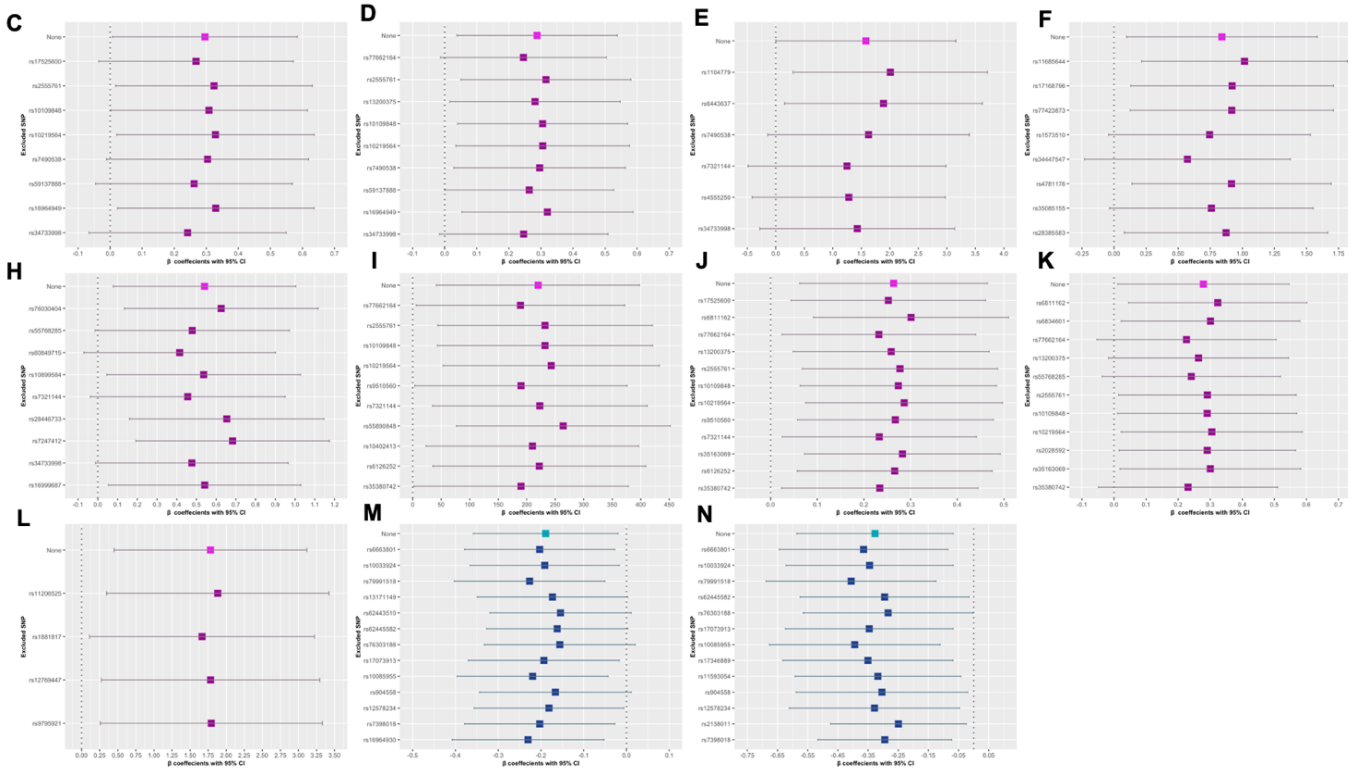


**Figure C.8 continued: Forest plots of β values following leave-one-out analyses for identified associations in WEs.**

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

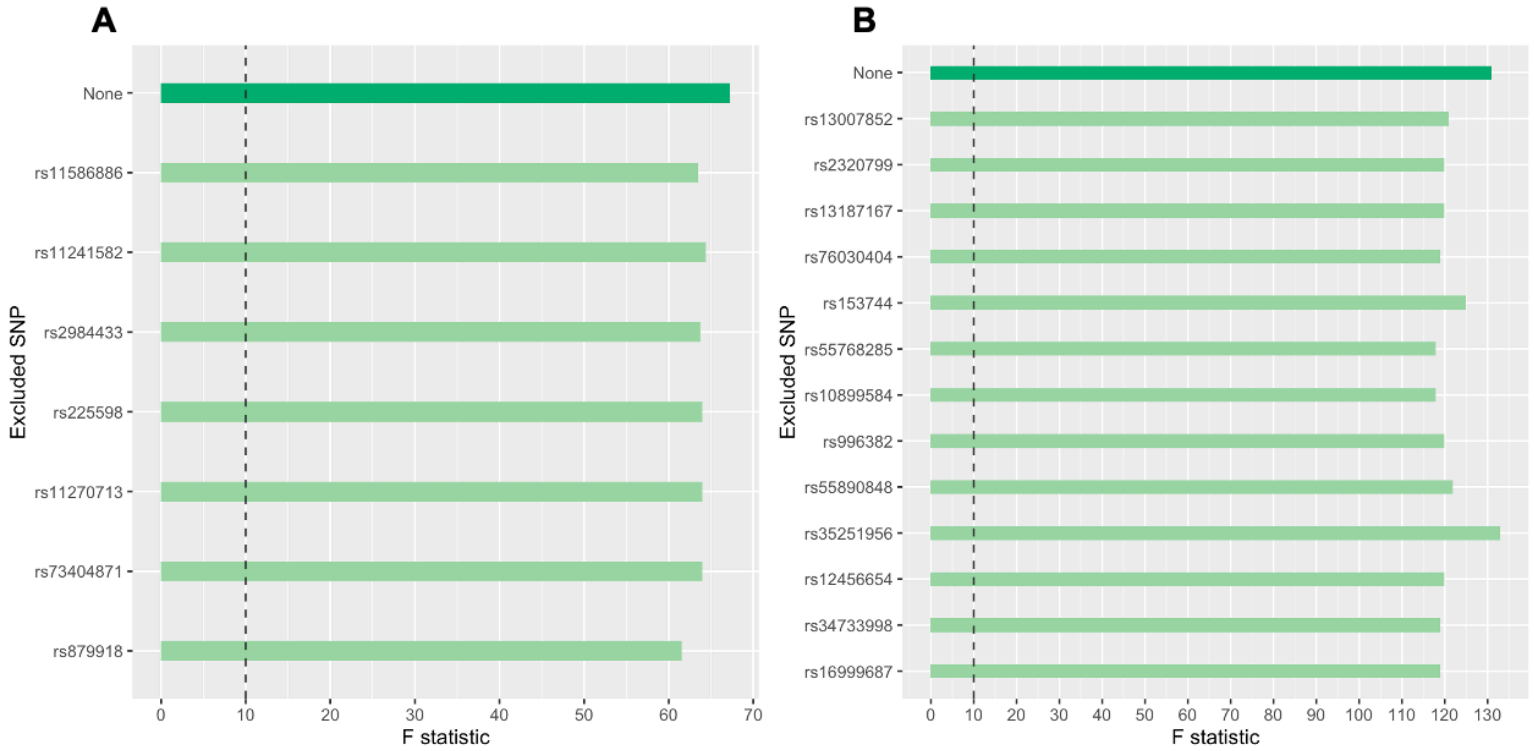



**Figure C.9: F statistics following leave-one-out analyses for identified associations in WEs.**

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.

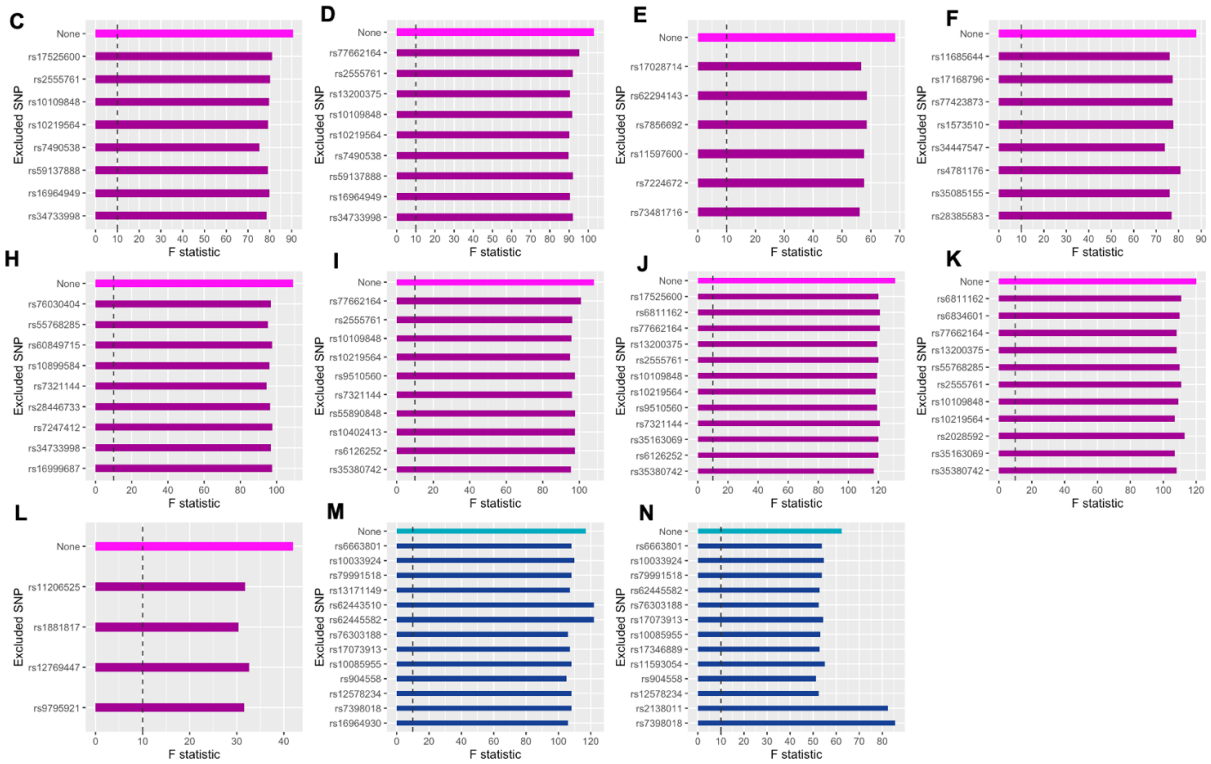



**Figure C.9 continued: F statistics following leave-one-out analyses for identified associations in WEs.**

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.

**Table C.6: Investigation of pleiotropy in MR analyses.**

**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.

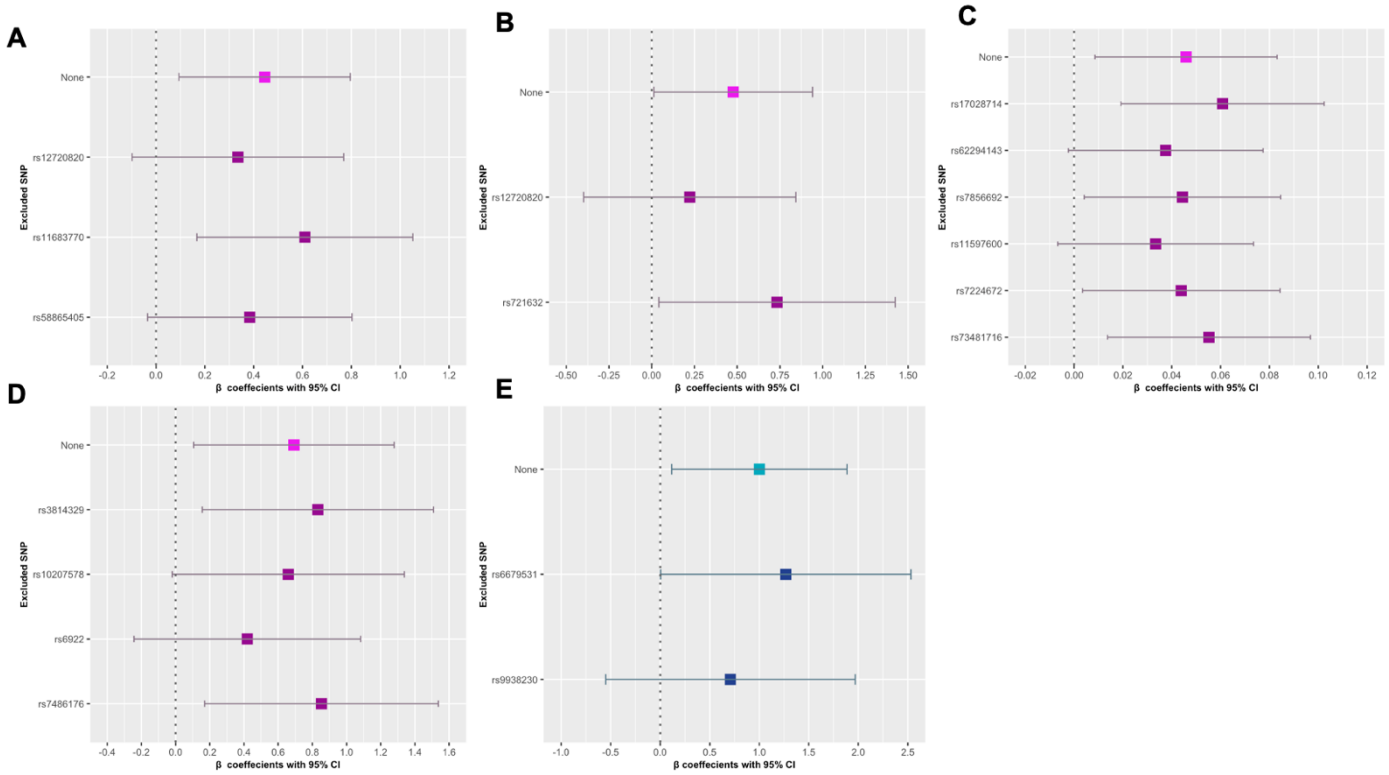



**Figure C.10: Forest plots of β values following leave-one-out analyses for identified associations in SAs.**

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**: FAw6. **B**: LA. **C**: M-VLDL-L. **D**: L-HDL-PL. **E**: S-HDL-C.

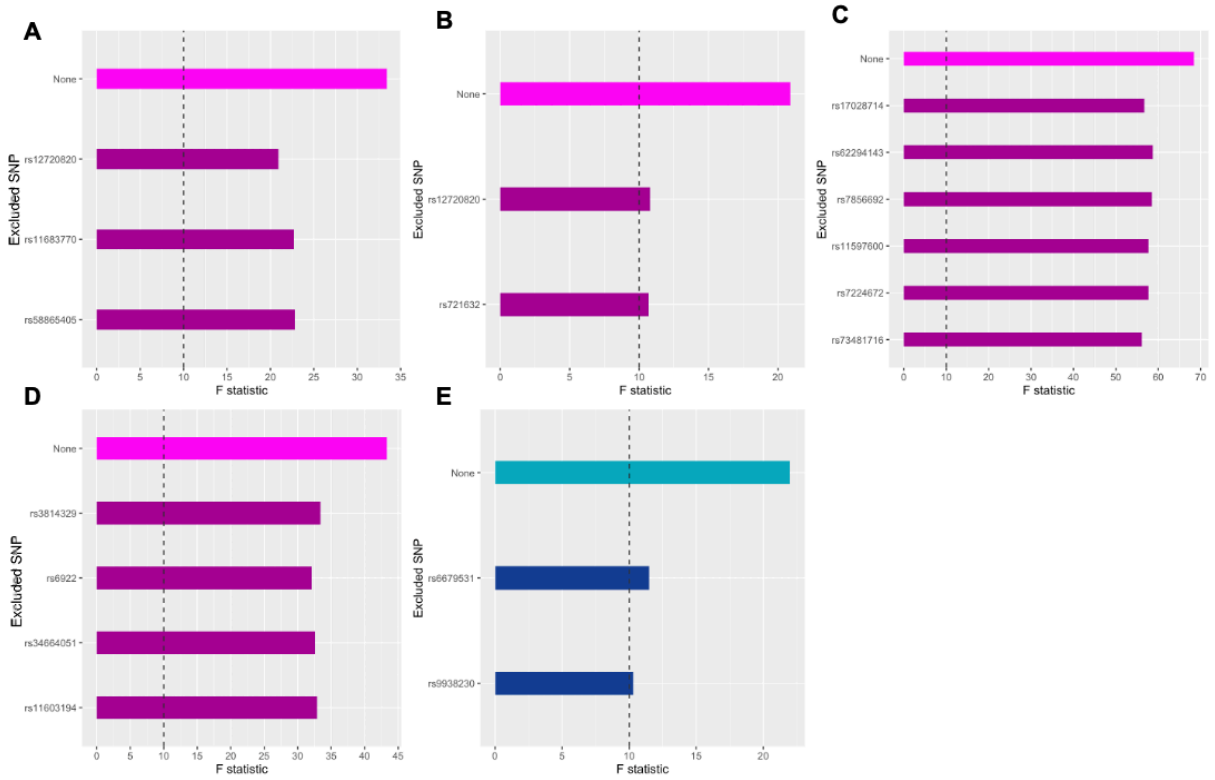
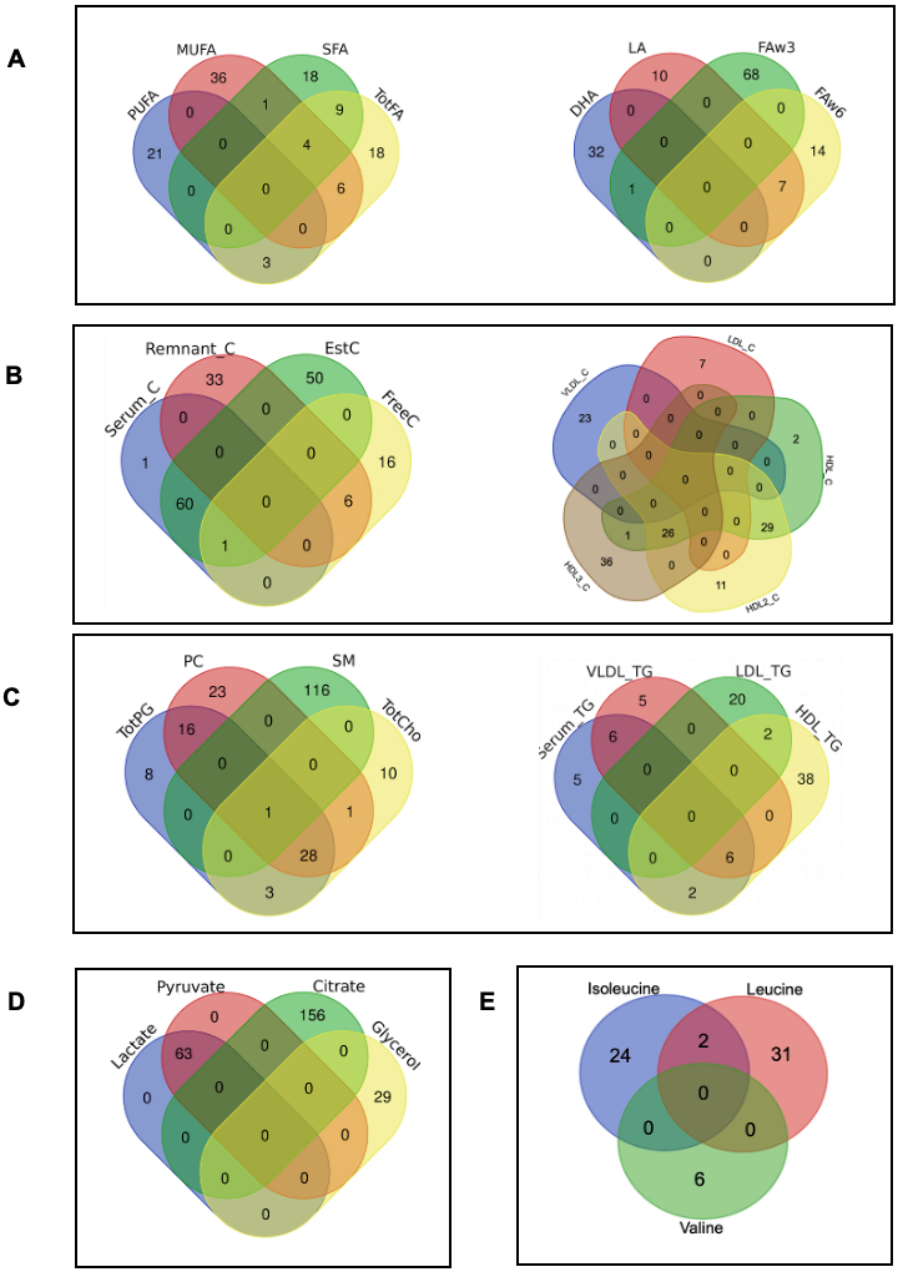


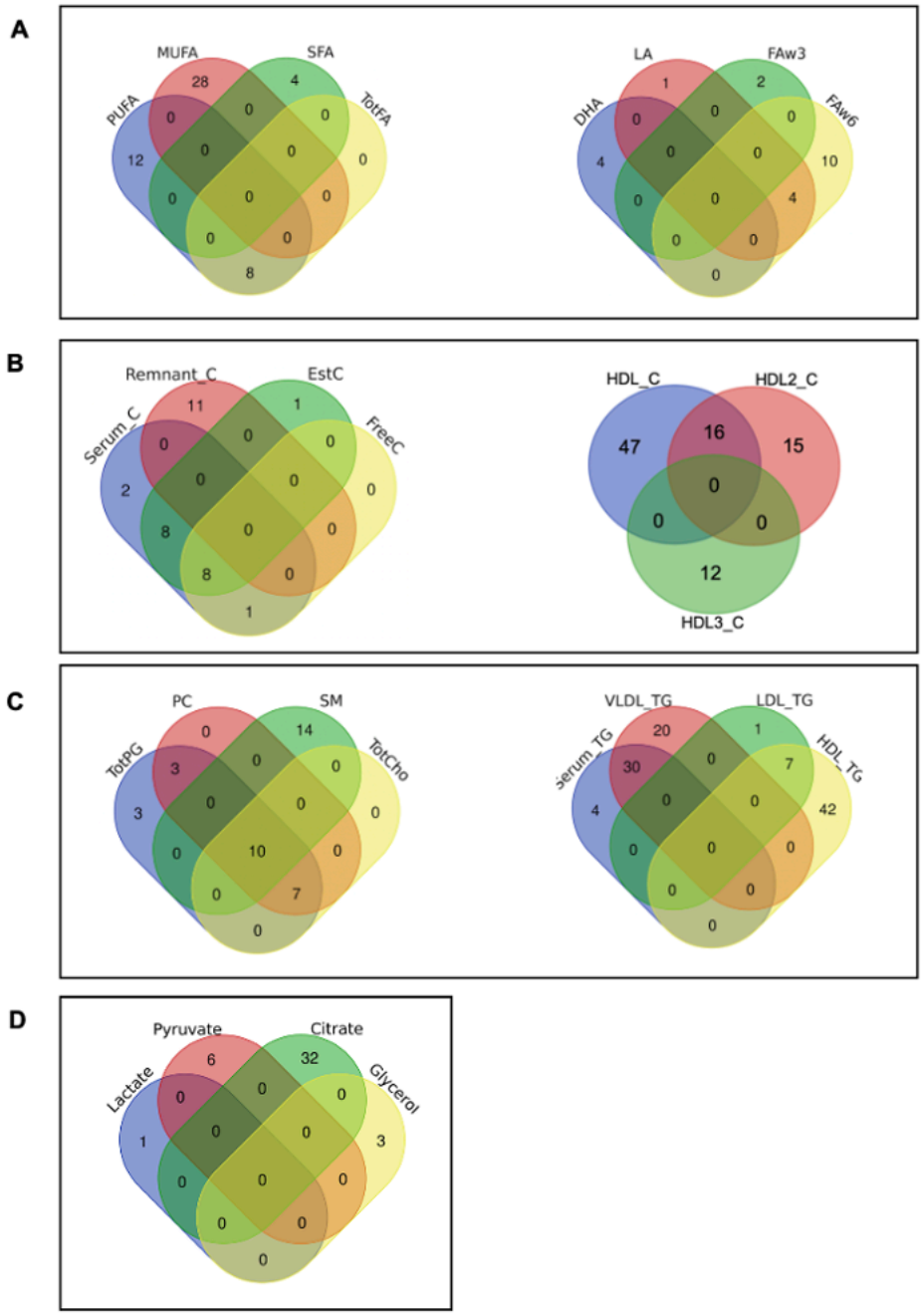

**Figure C.11: F statistics following leave-one- out analyses for identified associations in SAs.**

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.




## Figure C.12: Overlap of SNPs identified by class in WEs.

Venn diagram showing the overlap between suggestive (p value 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.




## Figure C.13: Overlap of SNPs identified by class in SAs.

Venn diagram highlighting the overlap between suggestive (p value 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.

## Table C.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.

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

| Metabolite Class | PC1 | | PC2 | |
|---|---|---|---|---|
| | 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.

**Table C.9: Number of SNPs identified in the GWAS analysis before and after LD thinning for the analysis in each class.**

| | White European | | South Asians | |
|---|---|---|---|---|
| **Metabolite class** | Suggestive | Thinned | Suggestive | Thinned |
| **XS-VLDL** | 18 | 7 | 1 | 1 |
| **S -VLDL** | 9 | 3 | 1 | 1 |
| **M-VLDL** | 9 | 4 | 1 | 1 |
| **L-VLDL** | 10 | 2 | 27 | 2 |
| **XL-VLDL** | 11 | 4 | 1 | 1 |
| **XXL-VLDL** | 34 | 5 | 1 | 1 |
| **All VLDL classes** | 21 | 10 | 16 | 6 |
| **IDL** | 41 | 12 | 8 | 2 |
| **S-LDL** | 31 | 4 | 5 | 3 |
| **M-LDL** | 18 | 7 | 1 | 1 |
| **L-LDL** | 24 | 10 | 2 | 3 |
| **All LDL classes** | 17 | 7 | 2 | 2 |
| **M-HDL** | - | - | 15 | 3 |
| **L-HDL** | 37 | 3 | 10 | 3 |
| **XL-HDL** | 38 | 2 | 4 | 2 |
| **All HDL classes** | - | - | 5 | 3 |
| **Fatty Acids** | 38 | 11 | 18 | 3 |
| **Lipoprotein Density** | 11 | 5 | 27 | 5 |
| **Cholesterol** | 46 | 8 | - | - |
| **Glycerides and Phospholipids** | 16 | 6 | 15 | 2 |
| **Branched Amino Acids** | 10 | 7 | 2 | 2 |
| **Unbranched Amino Acids** | - | - | 58 | 5 |

SNPs were classified as being in LD if their $R^2$ value exceeded 0.2. Suggestive level (p value $\leq 1 \times 10^{-5}$).

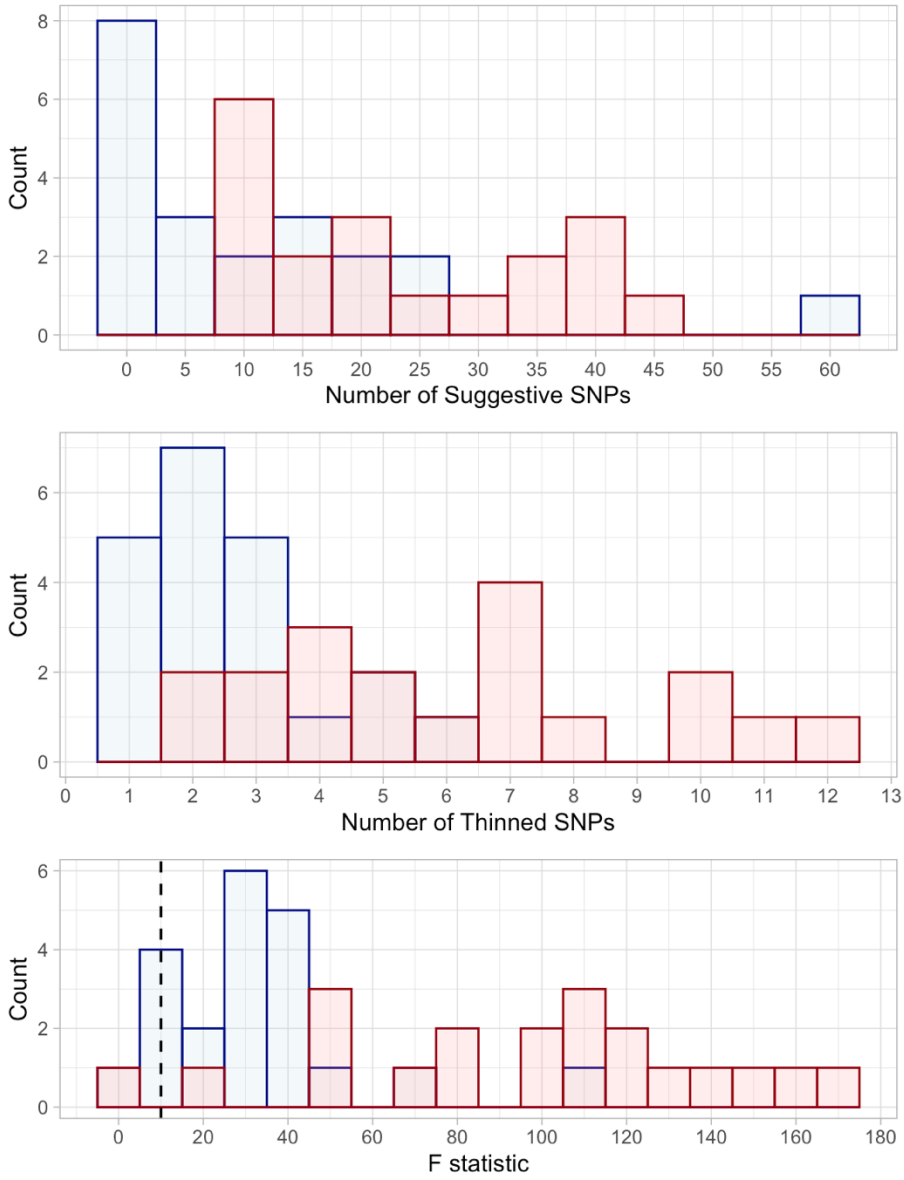



**Figure C.14: Histograms of the number of SNPs and strength of each instrument in the analysis of each metabolite class.**

**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: SAs. Red: WEs.

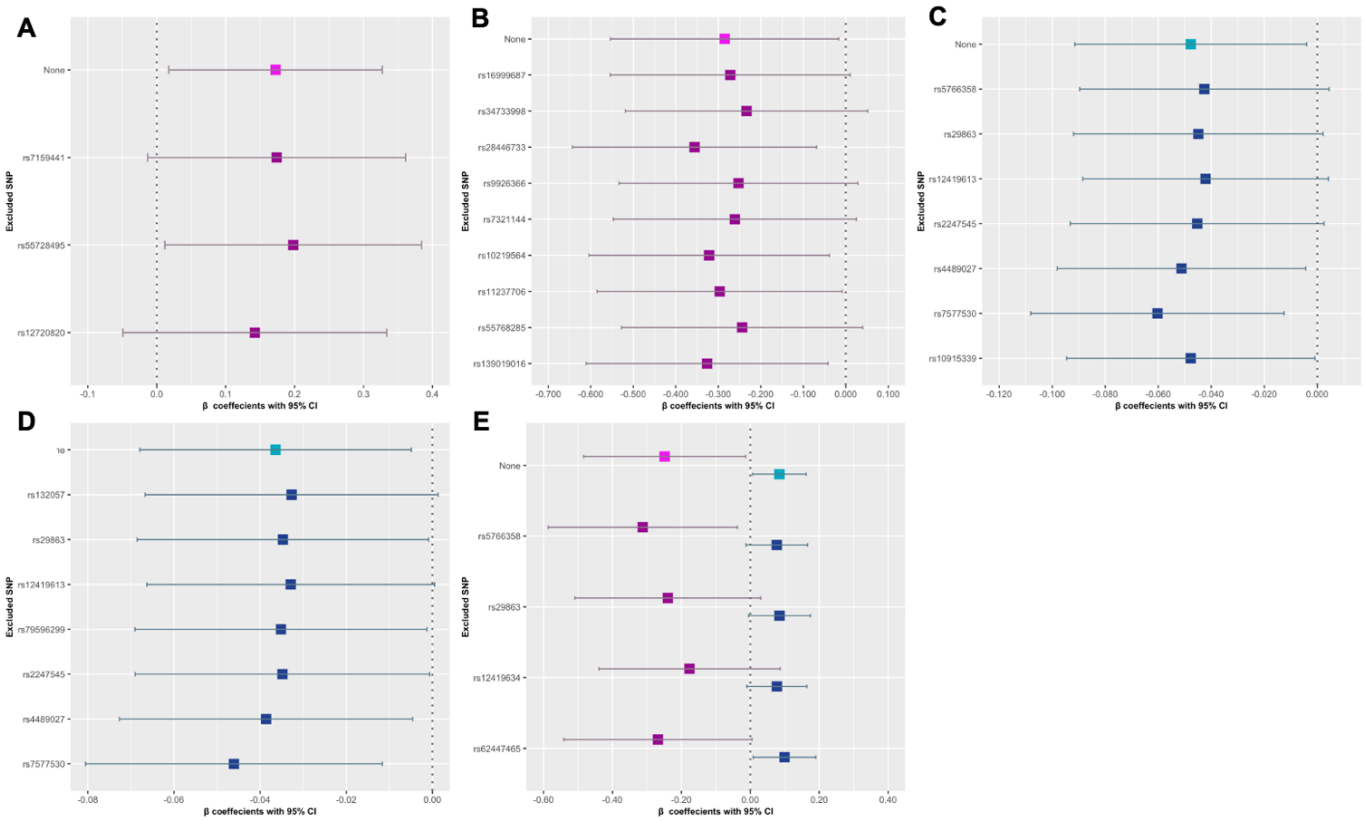



**Figure C.15: Forest plots of β values following leave-one-out analyses for identified associations in the analysis of metabolite class.**

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.

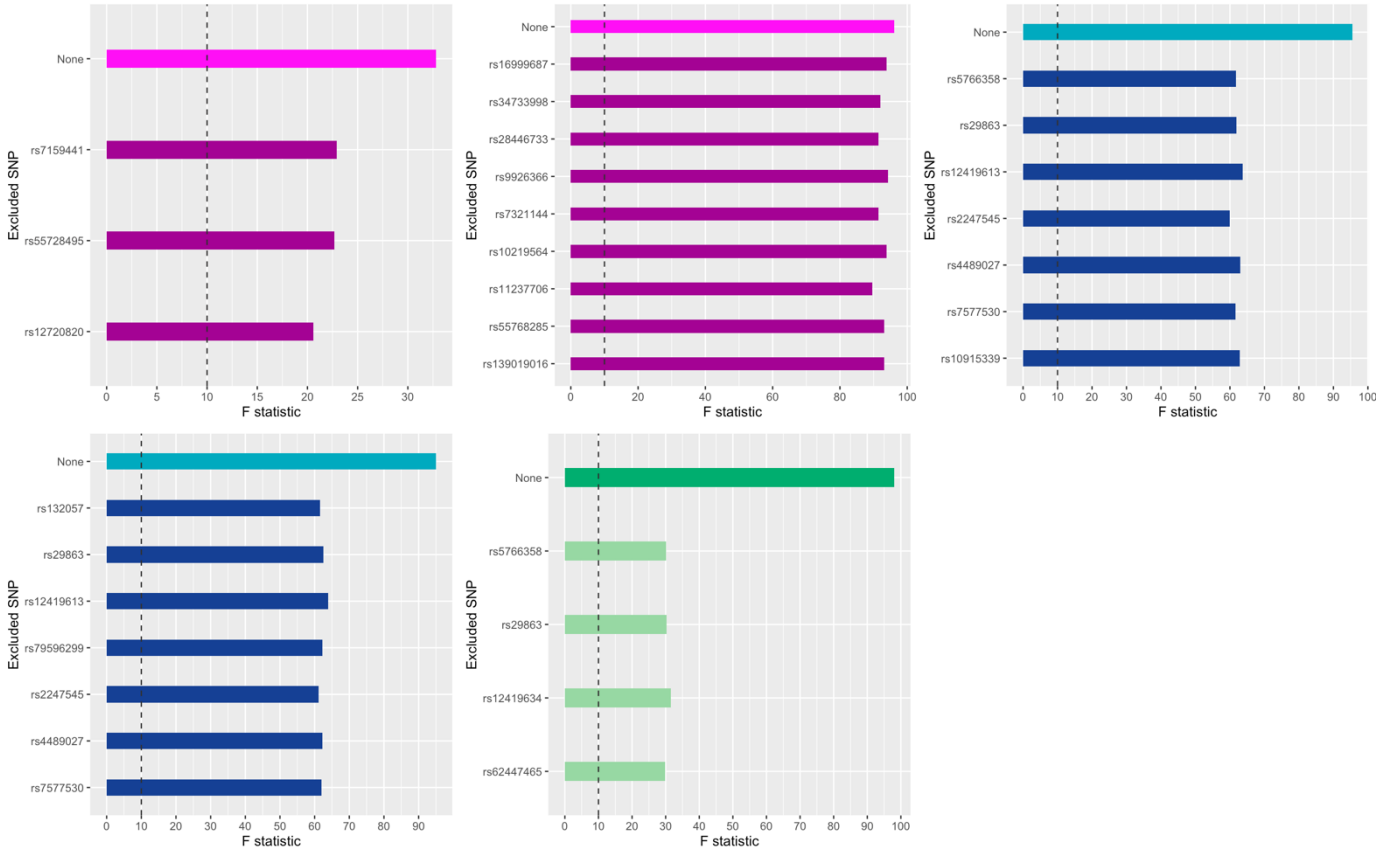



**Figure C.16: F statistics following leave-one-out analyses for identified associations in the analysis of metabolite classes.**

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.

## Table C.10: Investigation of pleiotropy in MR analyses of 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 |

## Table C.11: *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**

| Metabolite Class | Group | True β | Obs | $R^2$ | Variance of exposure | Variance of Outcome | Power (α = 0.05) | | | Power (α = 0.01) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 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** | WE | 0.000358 | -0.000274 | 0.0227 | 1.57 | 0.00814 | 0.05 | - | - | - | - |
| **Fasting** | 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.

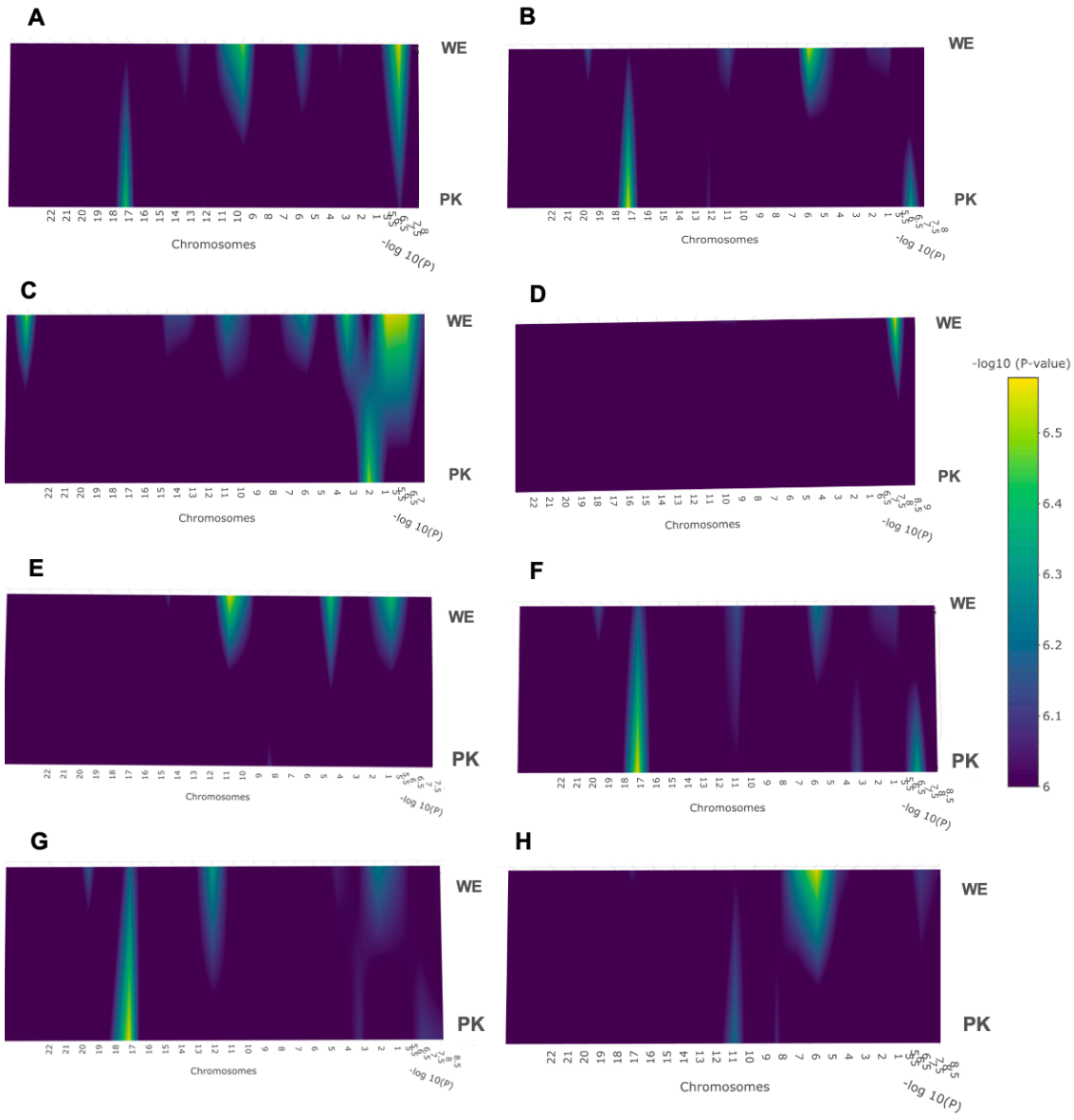



**Figure C.17: Heat maps highlighting p values above the suggestive level between ethnicities for fatty acids.**

**A:** TotFA, **B:** PUFA, **C:** MUFA, **D:** SFA, **E:** FAw3, **F:** FAw6, **G:** LA, **H:** DHA

# Appendix D
# Online supplementary data for Chapter 5

## Appendix D.1: GWAS results

Appendix D.1. GWAS results

## Appendix D.2: MR results

Appendix D.2. MR Results.xlsx

# List of References


Academy of Nutrition Dietetics. 2016. *Evidence analysis manual: Step in the academy evidence analysis process.*

Alesi, S, Ghelani, D, Rassie, K and Mousa, A (2021) Metabolomic biomarkers in gestational diabetes mellitus: A review of the evidence. *Int J Mol Sci.* 22,5512.

Alexander Kowarik, MT (2016) Imputation with the r package vim. *J. Stat. Softw.* 74, 1-16.

Alexopoulos, A-S, Qamar, A, Hutchins, K, Crowley, MJ, Batch, BC and Guyton, JR (2019) Triglycerides: Emerging targets in diabetes care? Review of moderate hypertriglyceridemia in diabetes. *Curr. Diabetes Rep.* 19, 13-13.

American Diabetes Association (ADA). 2020. *Gestational diabetes — gestational diabetes and a healthy baby?*

Ameur, A, Enroth, S, Johansson, A, Zaboli, G, Igl, W, Johansson, AC, Rivas, MA, Daly, MJ, Schmitz, G, Hicks, AA, Meitinger, T, Feuk, L, van Duijn, C, Oostra, B, Pramstaller, PP, Rudan, I, Wright, AF, Wilson, JF, Campbell, H and Gyllensten, U (2012) 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.* 90l, 809-820.

Arciero, E, Dogra, SA, Malawsky, DS, Mezzavilla, M, Tsismentzoglou, T, Huang, QQ, Hunt, KA, Mason, D, Sharif, SM, van Heel, DA, Sheridan, E, Wright, J, Small, N, Carmi, S, Iles, MM and Martin, HC (2021) Fine-scale population structure and demographic history of British Pakistanis. *Nat Commun.* 12, 7189.

Assaf-Balut, C, Garcia de la Torre, N, Duran, A, Fuentes, M, Bordiu, E, Del Valle, L, Familiar, C, Ortola, A, Jimenez, I, Herraiz, MA, Izquierdo, N, Perez, N, Torrejon, MJ, Ortega, MI, Illana, FJ, Runkle, I, de Miguel, MP, Montanez, C, Barabash, A, Cuesta, M, Rubio, MA and Calle-Pascual, AL (2017) A Mediterranean diet with additional extra virgin olive oil and pistachios reduces the incidence of gestational diabetes mellitus (gdm): A randomized controlled trial: The St. Carlos gdm prevention study. *PLoS One.* 12, e0185873.

Assaf-Balut, C, Garcia de le Toore, N, Fuentes, M, Duran, A, Bodio, E and del Valle, L (2018) A high adherance to six food targets of the mediterranean diet in the late first trimester is associated with reduction in the risk of materno-foetal outcomes: The St. Carlos gestational diabetes mellitus prevention study. *Nutrients.* 11,1.

Badon, SE, Enquobahrie, DA, Wartko, PD, Miller, RS, Qiu, C, Gelaye, B, Sorensen, TK and Williams, MA (2017) Healthy lifestyle during early pregnancy and risk of gestational diabetes mellitus. *American Journal of Epidemiology.* 186, 326-333.

Badoud, F, Lam, KP, DiBattista, A, Perreault, M, Zulyniak, MA, Cattrysse, B, Stephenson, S, Britz-McKibbin, P and Mutch, DM (2014) Serum and adipose tissue amino acid homeostasis in the metabolically healthy obese. *J. Proteome Res.* 13, 3455-3466.

Bai, Y, Tang, L, Li, L and Li, L (2020) The roles of adipoq rs266729 and mtnr1b rs10830963 polymorphisms in patients with gestational diabetes mellitus: A meta-analysis. *Gene.* 730, 144302.
Bandyopadhyay, M (2021) Gestational diabetes mellitus: A qualitative study of lived experiences of south asian immigrant women and perspectives of their health care providers in Melbourne, Australia. *BMC Pregnancy Childbirth.* 21, 500.
Bao, W, Bowers, K, Tobias, DK, Hu, FBH and Zhang, C (2013) Prepregnancy dietary protein intake, major dietary protein sources, and the risk of gestational diabetes mellitus. *Diabetes Care.* 36, 2001-2008.
Bao, W, Bowers, K, Tobias, DK, Olsen, SF, Chavarro, J, Vaag, A and Kiely, M (2014) Prepregnancy low-carbohydrate dietary pattern and risk of gestational diabetes mellitus: A prospective cohort study. *American Journal of Clinical Nutrition.* 99, 1378-1384.
Bao, W, Tobias, DK, Olsen, SF and Zhang, C (2014) Pre-pregnancy fried food consumption and the risk of gestational diabetes mellitus: A prospective cohort study. *NIH Public Access.* 57, 2485-2491.
Bar, N, Korem, T, Weissbrod, O, Zeevi, D, Rothschild, D, Leviatan, S, Kosower, N, Lotan-Pompan, M, Weinberger, A, Le Roy, CI, Menni, C, Visconti, A, Falchi, M, Spector, TD, consortium, ID, Adamski, J, Franks, PW, Pedersen, O and Segal, E (2020) A reference map of potential determinants for the human serum metabolome. *Nature.* 588, 135-140.
Barakat, S, Martinez, D, Thomas, M and Handley, M (2014) What do we know about gestational diabetes mellitus and risk for postpartum depression among ethnically diverse low-income women in the USA? *Arch Womens Ment Health.* 17l, 587-592.
Behboudi-Gandevani, S, Amiri, M, Bidhendi Yarandi, R and Ramezani Tehrani, F (2019) The impact of diagnostic criteria for gestational diabetes on its prevalence: A systematic review and meta-analysis. *Diabetol Metab Syndr.* 11, 11.
Benedetti, S, Al-Tannak, NF, Alzharani, M, Moir, HJ, Stensel, DJ, Thackray, AE, Naughton, DP, Dorak, MT, Spendiff, O and Hill, N (2019) Plasma free fatty acids metabolic profile with lc-ms and appetite-related hormones in south asian and white european men in relation to adiposity, physical activity and cardiorespiratory fitness: A cross-sectional study. *Metabolites.* 9, 71.
Bennett, CJ, Walker, RE, Blumfield, ML, Gwini, SM, Ma, J, Wang, F, Wan, Y, Dickinson, H and Truby, H (2018) Interventions designed to reduce excessive gestational weight gain can reduce the incidence of gestational diabetes mellitus: A systematic review and meta-analysis of randomised controlled trials. *Diabetes Res Clin Pract.* 141, 69-79.
Bird, PK, McEachan, RRC, Mon-Williams, M, Small, N, West, J, Whincup, P, Wright, J, Andrews, E, Barber, SE, Hill, LJB, Lennon, L, Mason, D, Shire, KA, Waiblinger, D, Waterman, AH, Lawlor, DA and Pickett, KE (2019) Growing up in bradford: Protocol for the age 7–11 follow up of the Born in Bradford birth cohort. *BMC Public Health.* 19, 939.
Blasco, H, Blaszczynski, J, Billaut, JC, Nadal-Desbarats, L, Pradat, PF, Devos, D, Moreau, C, Andres, CR, Emond, P, Corcia, P and Slowinski, R (2015) Comparative analysis of targeted metabolomics:

Dominance-based rough set approach versus orthogonal partial least square-discriminant analysis. *J Biomed Inform.* 53, 291-299.
Borenstein, M, Hedges, LV, Higgins, JPT and Rothstein, HR. 2009. *Introduction to meta-analysis*. London Wiley.
Bowers, K, Tobias, DK, Yeung, E, Hu, FB and Zhang, C (2012) A prospective study of prepregnancy dietary fat intake and risk of gestational diabetes. *American Journal of Clinical Nutrition.* 95, 446-453.
Bradford Metropolitan District Council. 2022. *Population.* [Online]. [Accessed 05/02/2022]. Available from: https://www.bradford.gov.uk/open-data/our-datasets/population/
Bragg, F, Kartsonaki, C, Guo, Y, Holmes, M, Du, H, Yu, C, Pei, P, Yang, L, Jin, D, Chen, Y, Schmidt, D, Avery, D, Lv, J, Chen, J, Clarke, R, Hill, M, Li, L, Millwood, I and Chen, Z (2021) Circulating metabolites and the development of type 2 diabetes in Chinese adults. *Diabetes Care.* 45, 477-480.
Brion, M-JA, Shakhbazov, K and Visscher, PM (2012) Calculating statistical power in mendelian randomization studies. *International Journal of Epidemiology.* 42, 1497-1501.
Brown, TJ, Brainard, J, Song, F, Wang, X, Abdelhamid, A, Hooper, L and Group, P (2019) Omega-3, omega-6, and total dietary polyunsaturated fat for prevention and treatment of type 2 diabetes mellitus: Systematic review and meta-analysis of randomised controlled trials. *BMJ.* 366, l4697.
Budu-Aggrey, A, Brumpton, B, Tyrrell, J, Watkins, S, Modalsli, EH, Celis-Morales, C, Ferguson, LD, Vie, GA, Palmer, T, Fritsche, LG, Loset, M, Nielsen, JB, Zhou, W, Tsoi, LC, Wood, AR, Jones, SE, Beaumont, R, Saunes, M, Romundstad, PR, Siebert, S, McInnes, IB, Elder, JT, Davey Smith, G, Frayling, TM, Asvold, BO, Brown, SJ, Sattar, N and Paternoster, L (2019) Evidence of a causal relationship between body mass index and psoriasis: A mendelian randomization study. *PLoS Med.* 16, e1002739.
Buniello, A, MacArthur, JAL, Cerezo, M, Harris, LW, Hayhurst, J, Malangone, C, McMahon, A, Morales, J, Mountjoy, E, Sollis, E, Suveges, D, Vrousgou, O, Whetzel, PL, Amode, R, Guillen, JA, Riat, HS, Trevanion, SJ, Hall, P, Junkins, H, Flicek, P, Burdett, T, Hindorff, LA, Cunningham, F and Parkinson, H (2019) The nhgri-ebi gwas catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research.* 47.
Burgess, S and Harshfield, E (2016) Mendelian randomization to assess causal effects of blood lipids on coronary heart disease: Lessons from the past and applications to the future. *Curr Opin Endocrinol Diabetes Obes.* 23, 124-130.
Burgess, S and Thompson, SG (2013) Use of allele scores as instrumental variables for mendelian randomization. *Int J Epidemiol.* 42, 1134-1144.
Burton, H, Jackson, C and Abubakar, I (2014) The impact of genomics on public health practice. *Br Med Bull.* 112, 37-46.
Carreras-Torres, R, Johansson, M, Gaborieau, V, Haycock, PC, Wade, KH, Relton, CL, Martin, RM, Davey Smith, G and Brennan, P (2017) The

role of obesity, type 2 diabetes, and metabolic factors in pancreatic cancer: A mendelian randomization study. *J Natl Cancer Inst.* 109l.
CDC. 2020. *LDL and HDL cholesterol: "Bad" and "good" cholesterol.* [Online]. [Accessed 01/04/2022]. Available from: https://www.cdc.gov/cholesterol/ldl_hdl.htm
Centre for Disease Control (CDC). 2019. *Gestational diabetes,*
Chai, JF, Raichur, S, Khor, IW, Torta, F, Chew, WS, Herr, DR, Ching, J, Kovalik, JP, Khoo, CM, Wenk, MR, Tai, ES and Sim, X (2020) Associations with metabolites in chinese suggest new metabolic roles in alzheimer's and parkinson's diseases. *Hum Mol Genet.* 29, 189-201.
Chang, CC, Chow, CC, Tellier, LC, Vattikuti, S, Purcell, SM and Lee, JJ (2015) Second-generation plink: Rising to the challenge of larger and richer datasets. *GigaScience.* 4,1.
Chen, L, Shi, L, Zhang, D and Chao, SM (2019a) Influence of acculturation on risk for gestational diabetes among Asian women. *Prev Chronic Dis.* 16, E158.
Chen, L, Yang, T, Chen, L, Wang, L, Wang, T, Zhao, L, Ye, Z, Zhang, S, Luo, L, Zheng, Z and Qin, J (2019b) Risk of congenital heart defects in offspring exposed to maternal diabetes mellitus: An updated systematic review and meta-analysis. *Arch Gynecol Obstet.* 300, 1491-1506.
Chen, Q, Francis, E, Hu, G and Chen, L (2018) Metabolomic profiling of women with gestational diabetes mellitus and their offspring: Review of metabolomics studies. *J Diabetes Complications.* 32, 512-523.
Chen, X, Stein, TP, Steer, RA and Scholl, TO (2019) Individual free fatty acids have unique associations with inflammatory biomarkers, insulin resistance and insulin secretion in healthy and gestational diabetic pregnant women. *BMJ Open Diabetes Res Care.* 7, e000632.
Chen, YD, Varasteh, BB and Reaven, GM (1993) Plasma lactate concentration in obesity and type 2 diabetes. *Diabete & Metabolisme.* 19, 348-354.
Chong, Y-S, Cai, S, Lin, H, Soh, SE, Lee, Y-S, Leow, MK-S, Chan, Y-H, Chen, L, Holbrook, JD, Tan, K-H, Rajadurai, VS, Yeo, GS-H, Kramer, MS, Saw, S-M, Gluckman, PD, Godfrey, KM and Kwek, K (2014) Ethnic differences translate to inadequacy of high-risk screening for gestational diabetes mellitus in an Asian population: A cohort study. *BMC Pregnancy Childbirth.* 14, 345.
Clish, CB (2015) Metabolomics: An emerging but powerful tool for precision medicine. *Cold Spring Harb Mol Case Stud.* 1l, a000588.
Cooke, BJN, Bush, WS and Crawford, DC (2020) Editorial: The importance of diversity in precision medicine research. *Front. Genet.* 11, 1664-8021.
Coustan, DR, Lowe, LP, Metzger, BE, Dyer, AR, International Association of, D and Pregnancy Study, G (2010) The hyperglycemia and adverse pregnancy outcome (hapo) study: Paving the way for new diagnostic criteria for gestational diabetes mellitus. *Am J Obstet Gynecol.* 202, 654 e651-656.
Crippa, A and Orsini, N (2016) Multivariate dose-response meta-analysis: The dosresmeta r package. *J. Stat. Softw., Code Snippets.* 72, 1 - 15.

Dalfrà, MG, Silvia Burlina, S, Del Vescovo, GG and Lapolla, A (2020) Genetics and epigenetics: New insight on gestational diabetes mellitus. *Front. Endocrinol.* 11, 602477.
Damm, P, Houshmand-Oeregaard, A, Kelstrup, L, Lauenborg, J, Mathiesen, ER and Clausen, TD (2016) Gestational diabetes mellitus and long-term consequences for mother and offspring: A view from Denmark. *Diabetologia.* 59, 1396-1399.
Darling, AL, Hart, KH, Macdonald, HM, Horton, K, Kang'ombe, AR, Berry, JL and Lanham-New, SA (2013) Vitamin d deficiency in UK South Asian women of childbearing age: A comparative longitudinal investigation with UK caucasian women. *Osteoporos Int.* 24, 477-488.
Davey Smith, G and Hemani, G (2014) Mendelian randomization: Genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* 23, R89-R98.
Davies, G, Jordan, S, Brooks, CJ, Thayer, D, Storey, M, Morgan, G, Allen, S, Garaiova, I, Plummer, S and Gravenor, M (2018) Long term extension of a randomised controlled trial of probiotics using electronic health records. *Sci Rep.* 8, 7668.
Davies, NM, Holmes, MV and Davey Smith, G (2018) Reading mendelian randomisation studies: A guide, glossary, and checklist for clinicians. *BMJ.* 362, k601.
de Souza, RJ, Shanmuganathan, M, Lamri, A, Atkinson, SA, Becker, A, Desai, D, Gupta, M, Mandhane, PJ, Moraes, TJ, Morrison, KM, Subbarao, P, Teo, KK, Turvey, SE, Williams, NC, Britz-McKibbin, P and Anand, SS (2020) Maternal diet and the serum metabolome in pregnancy: Robust dietary biomarkers generalizable to a multiethnic birth cohort. *CDN.* 4, nzaa144-nzaa144.
de Toro-Martin, J, Arsenault, BJ, Despres, JP and Vohl, MC (2017) Precision nutrition: A review of personalized nutritional approaches for the prevention and management of metabolic syndrome. *Nutrients.* 9.
Del Coco, L, Vergara, D, De Matteis, S, Mensa, E, Sabbatinelli, J, Prattichizzo, F, Bonfigli, AR, Storci, G, Bravaccini, S, Pirini, F, Ragusa, A, Casadei-Gardini, A, Bonafe, M, Maffia, M, Fanizzi, FP, Olivieri, F and Giudetti, AM (2019) NMR-based metabolomic approach tracks potential serum biomarkers of disease progression in patients with type 2 diabetes mellitus. *J Clin Med.* 8,23.
Diabetes Australia. 2020. *Gestational diabetes in australia.* diabetesaustralia.com.au: Diabetes Australia.
Diabetes Australia. 2021. *Managing gestational diabetes.* [Online]. [Accessed 09/02/2022]. Available from: https://www.diabetesaustralia.com.au/living-with-diabetes/managing-your-diabetes/managing-gestational-diabetes/
Diemer, EW, Labrecque, JA, Neumann, A, Tiemeier, H and Swanson, SA (2021) Mendelian randomisation approaches to the study of prenatal exposures: A systematic review. *Paediatr Perinat Epidemiol.* 35, 130-142.
Ding, M, Chavarro, J, Olsen, S, Lin, Y, Ley, SH, Bao, W, Rawal, S, Grunnet, LG, Thuesen, ACB, Mills, JL, Yeung, E, Hinkle, SN, Zhang, W, Vaag, A, Liu, A, Hu, FB and Zhang, C (2018) Genetic variants of gestational diabetes mellitus: A study of 112 snps among 8722 women in two independent populations. *Diabetologia.* 61, 1758-1768.

Domingues, LJ, Martínez-Gonzàlez, MA, Basterra-Gortari, FJ, Gea, A, Barbagallo, M and Bes-Restrollo, M (2014) Fast food consumption and gestational diabetes incidence in the SUN project. *PLoS One.* 9, 1-7.

Donazar-ezcurra, M, Burgo, CL-d, Martinez-gonzalez, MA, Basterra-gortari, FJ, Irala, JD and Bes-rastrollo, M (2017) Pre-pregnancy adherences to empirically derived dietary patterns and gestational diabetes risk in a mediterranean cohort: The Seguimiento Universidad de Navarra (SUN) project. *Br. J. Nutr.* 118, 715-721.

El-Chaar, D, Finkelstein, SA, Tu, X, Fell, DB, Gaudet, L, Sylvain, J, Tawagi, G, Wu Wen, S and Walker, M (2013) The impact of increasing obesity class on obstetrical outcomes. *JOGC.* 35, 224-233.

Emet, T, Üstüner, I, Güven, SG, Balık, G, Ural, ÜM, Tekin, YB, Şentürk, Ş, Şahin, FK and Avşar, AF (2013) Plasma lipids and lipoproteins during pregnancy and related pregnancy outcomes. *Arch. Gynecol. Obstet.* 288, 49-55.

Enquobahrie, DA, Denis, M, Tadesse, MG, Gelaye, B, Ressom, HW and Williams, MA (2015) Maternal early pregnancy serum metabolites and risk of gestational diabetes mellitus. *J. Clin. Endocr.* 100, 4348-4356.

Ensembl. 2021a. *Gene: Foxn1.* [Online]. [Accessed 01/04/2022]. Available from: https://www.ensembl.org/Homo_sapiens/Gene/Summary?g=ENSG00000108018;r=10:106573663-107164706

Ensembl. 2021b. *Gene: Slc13a2.* [Online]. [Accessed 01/04/2022]. Available from: https://www.ensembl.org/Homo_sapiens/Gene/Summary?g=ENSG00000108018;r=10:106573663-107164706

Ensembl. 2021c. *Gene: Sorcs1.* [Online]. [Accessed 01/04/2022]. Available from: https://www.ensembl.org/Homo_sapiens/Gene/Summary?g=ENSG00000108018;r=10:106573663-107164706

Farrar, D, Fairley, L, Santorelli, G, Tuffnell, D, Sheldon, TA, Wright, J, van Overveld, L and Lawlor, DA (2015) Association between hyperglycaemia and adverse perinatal outcomes in South Asian and white British women: Analysis of data from the Born in Bradford cohort. *Lancet Diabetes Endocrinol.* 3, 795-804.

Farrar, D, Simmonds, M and Griffin , S. 2016. *The identification and treatment of women with hyperglycaemia in pregnancy: An analysis of individual participant data, systematic reviews, meta-analyses and an economic evaluation*. Southampton, UK NIHR Journals Library.

Flynn, AC, Seed, PT, Patel, N, Barr, S, Bell, R, Briley, AL, Godfrey, KM, Nelson, SM, Oteng-ntim, E, Robinson, SM, Sanders, TA, Sattar, N and Wardle, J (2016) Dietary patterns in obese pregnant women: Influence of a behavioral intervention of diet and physical activity in the upbeat randomized controlled trial. *IJBNPA.* 13, 1-12.

Franzago, M, Fraticelli, F, Marchetti, D, Celentano, C, Liberati, M, Stuppia, L and Vitacolonna, E (2018) Nutrigenetic variants and cardio-metabolic risk in women with or without gestational diabetes. *Diabetes Res Clin Pract.* 137, 64-71.

Franzago, M, Santurbano, D, Vitacolonna, E and Stuppia, L (2020) Genes and diet in the prevention of chronic diseases in future generations. *Int J Mol Sci.* 21.
Gadve, SS, Chavanda, S, Mukherjee, AD, Aziz, S, Joshi, A and Patwardhan, M (2021) 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.* 25l, 176-181.
Gail, MH (2008) Estimation and interpretation of models of absolute risk from epidemiologic data, including family-based studies. *Lifetime Data Analysis.* 14, 18-36.
Gebauer, SK, Psota, TL, Harris, WS and Kris-Etherton, PM (2006) N−3 fatty acid dietary recommendations and food sources to achieve essentiality and cardiovascular benefits. *AJCN.* 83, 1526S-1535S.
GeneCards Suite. 2022a. *Peli2 gene - pellino e3 ubiquitin protein ligase family member 2.* [Online]. [Accessed 01/04/2022]. Available from: https://www.genecards.org/cgi-bin/carddisp.pl?gene=PELI2&keywords=rs7159441
GeneCards Suite. 2022b. *Rgl1 gene - ral guanine nucleotide dissociation stimulator like 1.* [Online]. [Accessed 01/04/2022]. Available from: https://www.genecards.org/cgi-bin/carddisp.pl?gene=PELI2&keywords=rs7159441
Genomes Project Consortium (2015) A global reference for human genetic variation. *Nature.* 526, 68-74.
George, G, Gan, S, Huang, Y, Appleby, P, Nar, AS, Venkatesan, R, Mohan, V, Palmer, CNA and Doney, ASF (2020) Phegwas: A new dimension to visualize gwas across multiple phenotypes. *Bioinformatics.* 36, 2500-2505.
Gicevic, S, Gaskins, AJ, Fung, TT, Rosner, B, Tobias, K, Isanaka, S and Willett, WC (2018) Evaluating pre-pregnancy dietary diversity vs . Dietary quality scores as predictors of gestational diabetes and hypertensive disorders of pregnancy. *PLoS ONE.* 1-14.
Gillespie, M, Jassal, B, Stephan, R, Milacic, M, Rothfels, K, Senff-Ribeiro, A, Griss, J, Sevilla, C, Matthews, L, Gong, C, Deng, C, Varusai, T, Ragueneau, E, Haider, Y, May, B, Shamovsky, V, Weiser, J, Brunson, T, Sanati, N, Beckman, L, Shao, X, Fabregat, A, Sidiropoulos, K, Murillo, J, Viteri, G, Cook, J, Shorser, S, Bader, G, Demir, E, Sander, C, Haw, R, Wu, G, Stein, L, Hermjakob, H and D'Eustachio, P (2021) The reactome pathway knowledgebase 2022. *Nucleic Acids Res.* 50, D687-D692.
Gorgal, R, Gonçalves, E, Barros, M, Namora, G, Magalhães, A, Rodrigues, T and Montenegro, N (2012) Gestational diabetes mellitus: A risk factor for non-elective cesarean section. *J Obstet Gynaecol Res.* 38, 154-159.
Goyal, S, Tanigawa, Y, Zhang, W, Chai, JF, Almeida, M, Sim, X, Lerner, M, Chainakul, J, Ramiu, JG, Seraphin, C, Apple, B, Vaughan, A, Muniu, J, Peralta, J, Lehman, DM, Ralhan, S, Wander, GS, Singh, JR, Mehra, NK, Sidorov, E, Peyton, MD, Blackett, PR, Curran, JE, Tai, ES, van Dam, R, Cheng, CY, Duggirala, R, Blangero, J, Chambers, JC, Sabanayagam, C, Kooner, JS, Rivas, MA, Aston, CE and Sanghera, DK (2021) APOC3 genetic variation, serum triglycerides,

and risk of coronary artery disease in asian indians, europeans, and other ethnic groups. *Lipids Health Dis.* 20, 113.
Gray, RG, Kousta, E, McCarthy, MI, Godsland, IF, Venkatesan, S, Anyaoku, V and Johnston, DG (2013) Ethnic variation in the activity of lipid desaturases and their relationships with cardiovascular risk factors in control women and an at-risk group with previous gestational diabetes mellitus: A cross-sectional study. *Lipids Health Dis.* 12, 25-25.
Greenhalgh, T, Clinch, M, Afsar, N, Choudhury, Y, Sudra, R, Campbell-Richards, D, Claydon, A, Hitman, GA, Hanson, P and Finer, S (2015) Socio-cultural influences on the behaviour of South Asian women with diabetes in pregnancy: Qualitative study using a multi-level theoretical approach. *BMC Med.* 13, 120.
Griffith, RJ, Alsweiler, J, Moore, AE, Brown, S, Middleton, P, Shepherd, E and Crowther, CA (2020) Interventions to prevent women from developing gestational diabetes mellitus: An overview of cochrane reviews. *Cochrane Database Syst Rev.* 6l, CD012394.
Gulbins, E, Szabo, I, Baltzer, K and Lang, F (1997) Ceramide-induced inhibition of t lymphocyte voltage-gated potassium channel is mediated by tyrosine kinases. *Proc Natl Acad Sci U S A.* 94, 7661-7666.
Guo, F, Long, W, Zhou, W, Zhang, B, Liu, J and Yu, B (2018) Fto, gckr, cdkal1 and cdkn2a/b gene polymorphisms and the risk of gestational diabetes mellitus: A meta-analysis. *Arch Gynecol Obstet.* 298, 705-715.
Guo, XY, Shu, J, Fu, XH, Chen, XP, Zhang, L, Ji, MX, Liu, XM, Yu, TT, Sheng, JZ and Huang, HF (2019) Improving the effectiveness of lifestyle interventions for gestational diabetes prevention: A meta-analysis and meta-regression. *BJOG.* 126l, 311-320.
Haber, EP, Procopio, J, Carvalho, CR, Carpinelli, AR, Newsholme, P and Curi, R (2006) New insights into fatty acid modulation of pancreatic beta-cell function. *Int Rev Cytol.* 248, 1-41.
Hagenbeek, FA, Pool, R, van Dongen, J, Draisma, HHM, Jan Hottenga, J, Willemsen, G, Abdellaoui, A, Fedko, IO, den Braber, A, Visser, PJ, de Geus, E, Willems van Dijk, K, Verhoeven, A, Suchiman, HE, Beekman, M, Slagboom, PE, van Duijn, CM, Consortium, BM, Harms, AC, Hankemeier, T, Bartels, M, Nivard, MG and Boomsma, DI (2020) Heritability estimates for 361 blood metabolites across 40 genome-wide association studies. *Nat Commun.* 11, 39.
Hariton, E and Locascio, JJ (2018) Randomised controlled trials - the gold standard for effectiveness research: Study design: Randomised controlled trials. *BJOG.* 125, 1716.
Hassani, ZS, Boffetta, P and Hosseinzadeh, M (2020) Dietary patterns and risk of gestational diabetes mellitus: A systematic review and meta-analysis of cohort studies. *Clin Nutr ESPEN.* 36, 1-9.
He, JR, Yuan, MY, Chen, NN, Lu, JH, Hu, CY, Mai, WB, Zhang, RF, Pan, YH, Qiu, L, Wu, YF, Xiao, WQ, Liu, Y, Xia, HM and Qiu, X (2015) Maternal dietary patterns and gestational diabetes mellitus: A large prospective cohort study in China. *Br J Nutr.* 113, 1292-1300.
Hernandez, TL, Mande, A and Barbour, LA (2018) Nutrition therapy within and beyond gestational diabetes. *Diabetes Res. Clin. Pract.*145, 39-50.

Herring, SJ, Nelson, DB, Davey, A, Klotz, AA, Dibble, LV, Oken, E and Foster, GD (2012) Determinants of excessive gestational weight gain in urban, low-income women. *WHI.* 22, e439-446.
Higgins, J, Churchill, R, Chandler, J, Cumpston, M, Li, T, Page, M and Welch, V. 2019. *Cochrane handbook for systematic reviews of interventions*. Chichester (UK) John Wiley & Sons.
Hosseinkhani, S, Dehghanbanadaki, H, Aazami, H, Pasalar, P, Asadi, M and Razi, F (2021) Association of circulating omega 3, 6 and 9 fatty acids with gestational diabetes mellitus: A systematic review. *BMC Endocr. Disord.* 21, 120.
Howe, KL, Achuthan, P, Allen, J, Allen, J, Alvarez-Jarreta, J, Amode, MR, Armean, IM, Azov, AG, Bennett, R, Bhai, J, Billis, K, Boddu, S, Charkhchi, M, Cummins, C, Da Rin Fioretto, L, Davidson, C, Dodiya, K, El Houdaigui, B, Fatima, R, Gall, A, Garcia Giron, C, Grego, T, Guijarro-Clarke, C, Haggerty, L, Hemrom, A, Hourlier, T, Izuogu, OG, Juettemann, T, Kaikala, V, Kay, M, Lavidas, I, Le, T, Lemos, D, Gonzalez Martinez, J, Marugán, JC, Maurel, T, McMahon, AC, Mohanan, S, Moore, B, Muffato, M, Oheh, DN, Paraschas, D, Parker, A, Parton, A, Prosovetskaia, I, Sakthivel, MP, Salam, Ahamed I A, Schmitt, BM, Schuilenburg, H, Sheppard, D, Steed, E, Szpak, M, Szuba, M, Taylor, K, Thormann, A, Threadgold, G, Walts, B, Winterbottom, A, Chakiachvili, M, Chaubal, A, De Silva, N, Flint, B, Frankish, A, Hunt, SE, IIsley, GR, Langridge, N, Loveland, JE, Martin, FJ, Mudge, JM, Morales, J, Perry, E, Ruffier, M, Tate, J, Thybert, D, Trevanion, SJ, Cunningham, F, Yates, AD, Zerbino, DR and Flicek, P (2020) Ensembl 2021. *Nucleic Acids Res.* 49, D884-D891.
Hrolfsdottir, L, Gunnarsdottir, I, Birgisdottir, BE, Hreidarsdottir, IT and Kr.Smarason, A (2019) Can a simple dietary screening in early pregnancy identify dietary habits associated with gesational diabetes. *Nutrients.* 11, 1-12.
Hu, J, Gillies, CL, Lin, S, Stewart, ZA, Melford, SE, Abrams, KR, Baker, PN, Khunti, K and Tan, BK (2021) Association of maternal lipid profile and gestational diabetes mellitus: A systematic review and meta-analysis of 292 studies and 97,880 women. *EClinicalMedicine.* 34, 100830.
Hu, J, Oken, E, Izzuddin, MA, Pi-Id, L, Gao, M, Wei, X and Wen, D (2019) Dietary patterns during pregnancy are associated with the risk of gestational diabetes mellitus: Cohort study. *Nutrients.* 11, 1-15.
International Diabetes Federation. 2019. *IDF diabetes atlas.* 9th ed. Brussels, Belgium.
Inthout, J, Ioannidis, JPA and Borm, GF (2014) The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard dersimonian-laird method. *BMC Med. Res. Methodol.* 1-12.
Jackson, D, Law, M, Rücker, G and Schwarzer, G (2017) The Hartung-Knapp modification for random-effects meta-analysis: A useful refinement but are there any residual concerns ? *Stat Med.* 3923-3934.
Jacob, S, Nodzenski, M, Reisetter, AC, Bain, JR, Muehlbauer, MJ, Stevens, RD, Ilkayeva, OR, Lowe, LP, Metzger, BE, Newgard, CB, Scholtens, DM, Lowe, WL, Jr. and Group, HSCR (2017) Targeted metabolomics demonstrates distinct and overlapping maternal metabolites

associated with BMI, glucose, and insulin sensitivity during pregnancy across four ancestry groups. *Diabetes Care.* 40, 911-919.
Jacobson, TA, Glickstein, SB, Rowe, JD and Soni, PN (2012) Effects of eicosapentaenoic acid and docosahexaenoic acid on low-density lipoprotein cholesterol and other lipids: A review. *J Clin Lipidol.* 6, 5-18.
Jamilian, M, Samimi, M, Ebrahimi, FA, Hashemi, T, Taghizadeh, M, Razavi, M, Sanami, M and Asemi, Z (2017) The effects of vitamin d and omega-3 fatty acid co-supplementation on glycemic control and lipid concentrations in patients with gestational diabetes. *J Clin Lipidol.* 11, 459-468.
Jiang, R, Wu, S, Fang, C, Wang, C, Yang, Y, Liu, C, Hu, J and Huang, Y (2020) Amino acids levels in early pregnancy predict subsequent gestational diabetes. *J Diabetes.* 12, 503-511.
Johnston, R, Jones, K and Manley, D (2018) Confounding and collinearity in regression analysis: A cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. *Qual Quant.* 52, 1957-1976.
Jolliffe, IT and Cadima, J (2016) Principal component analysis: A review and recent developments. *Philos Trans A Math Phys Eng Sci.* 374, 20150202.
Jones, TE, Pories, WJ, Houmard, JA, Tanner, CJ, Zheng, D, Zou, K, Coen, PM, Goodpaster, BH, Kraus, WE and Dohm, GL (2019) Plasma lactate as a marker of metabolic health: Implications of elevated lactate for impairment of aerobic metabolism in the metabolic syndrome. *Surgery.* 166, 861-866.
Josse, J and Husson, F (2016) Missmda: A package for handling missing values in multivariate data analysis. *J. Stat. Softw.* 70,1.
Kamat, MA, Blackshaw, JA, Young, R, Surendran, P, Burgess, S, Danesh, J, Butterworth, AS and Staley, JR (2019) Phenoscanner v2: An expanded tool for searching human genotype-phenotype associations. *Bioinformatics.* 35, 4851-4853.
Kampmann, U, Knorr, S, Fuglsang, J and Ovesen, P (2019) Determinants of maternal insulin resistance during pregnancy: An updated overview. *J. Diabetes Res.* 2019, 5320156.
Kandasamy, S, Nguyen, L, Desai, D, Anand, SS, Sherifali, D and de Souza, RJ (2021) Barriers to, and facilitators of, lifestyle changes to prevent gestational diabetes: An interpretive description of South Asian women and health-care providers living and working in Southern Ontario, Canada. *Can J Diabetes.* 45, 144-154.
Kanguru, L, Bezawada, N, Hussein, J and Bell, J (2014) The burden of diabetes mellitus during pregnancy in low- and middle-income countries: A systematic review. *Glob Health Action.* 7, 23987.
Karamanos, B, Thanopoulou, A, Anastasiou, E, Assaad-Khalil, S, Albache, N, Bachaoui, M, Slama, CB, El Ghomari, H, Jotic, A, Lalic, N, Lapolla, A, Saab, C, Marre, M, Vassallo, J, Savona-Ventura, C and Group, M-GS (2014) Relation of the mediterranean diet with the incidence of gestational diabetes. *Eur J Clin Nutr.* 68, 8-13.
Kawasaki, M, Arata, N, Miyazaki, C, Mori, R, Kikuchi, T, Ogawa, Y and Ota, E (2018) Obesity and abnormal glucose tolerance in offspring of

diabetic mothers: A systematic review and meta-analysis. *PLoS One.* 13, e0190676.
Kc, K, Shakya, S and Zhang, H (2015) Gestational diabetes mellitus and macrosomia: A literature review. *Ann. Nutr. Metab.* 66(suppl 2), 14-20.
Kettunen, J, Tukiainen, T, Sarin, AP, Ortega-Alonso, A, Tikkanen, E, Lyytikainen, LP, Kangas, AJ, Soininen, P, Wurtz, P, Silander, K, Dick, DM, Rose, RJ, Savolainen, MJ, Viikari, J, Kahonen, M, Lehtimaki, T, Pietilainen, KH, Inouye, M, McCarthy, MI, Jula, A, Eriksson, J, Raitakari, OT, Salomaa, V, Kaprio, J, Jarvelin, MR, Peltonen, L, Perola, M, Freimer, NB, Ala-Korpela, M, Palotie, A and Ripatti, S (2012) Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nat Genet.* 44, 269-276.
Koletzko, B, Reischl, E, Tanjung, C, Gonzalez-Casanova, I, Ramakrishnan, U, Meldrum, S, Simmer, K, Heinrich, J and Demmelmair, H (2019) FADS1 and FADS2 polymorphisms modulate fatty acid metabolism and dietary impact on health. *Annu Rev Nutr.* 39, 21-44.
Kouhkan, A, Najafi, L, Malek, M, Baradaran, HR, Hosseini, R, Khajavi, A and Khamseh, ME (2021) Gestational diabetes mellitus: Major risk factors and pregnancy-related outcomes: A cohort study. *Int. J. Reprod. BioMed.* 19, 827-836.
Kramer, CK, Campbell, S and Retnakaran, R (2019) Gestational diabetes and the risk of cardiovascular disease in women: A systematic review and meta-analysis. *Diabetologia.* 62, 905-914.
Krauss, RM (2004) Lipids and lipoproteins in patients with type 2 diabetes. *Diabetes Care.* 27, 1496-1504.
Kuller, LH and Catov, J (2017) Invited commentary: Gestational hypertension and diabetes-a major public health concern. *Am J Epidemiol.* 186l, 1125-1128.
Kunasegaran, T, Balasubramaniam, V, Arasoo, VJT, Palanisamy, UD and Ramadas, A (2021) Gestational diabetes mellitus in Southeast Asia: A scoping review. *Int J Environ Res Public Health.* 18.
Kwak, SH, Kim, SH, Cho, YM, Go, MJ, Cho, YS, Choi, SH, Moon, MK, Jung, HS, Shin, HD, Kang, HM, Cho, NH, Lee, IK, Kim, SY, Han, BG, Jang, HC and Park, KS (2012) A genome-wide association study of gestational diabetes mellitus in Korean women. *Diabetes.* 61, 531-541.
Lagunova, Z, Porojnicu, AC, Lindberg, F, Hexeberg, S and Moan, J (2009) The dependency of vitamin d status on body mass index, gender, age and season. *Anticancer Res.* 29, 3713-3720.
Lamyian, M, Hosseinpour-Niazi, S, Mirmiran, P, M., BL, Goshtasebil, A and Azizi, F (2017) Pre-pregnancy fast food consumption is associated with gestational diabetes mellitus among Tehranian women. *Nutrients.* 9, 3.
Laufs, U, Parhofer, KG, Ginsberg, HN and Hegele, RA (2020) Clinical review on triglycerides. *Eur. Heart J.* 41, 99-109c.
Law, KP and Zhang, H (2017) The pathogenesis and pathophysiology of gestational diabetes mellitus: Deductions from a three-part longitudinal metabolomics study in China. *Clin Chim Acta.* 468, 60-70.
Lawlor, DA, West, J, Fairley, L, Nelson, SM, Bhopal, RS, Tuffnell, D, Freeman, DJ, Wright, J, Whitelaw, DC and Sattar, N (2014)

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.* 57, 2492-2500.
Layton, J, Powe, C, Allard, C, Battista, MC, Doyon, M, Bouchard, L, Perron, P, Wessel, J and Hivert, MF (2019) Maternal lipid profile differs by gestational diabetes physiologic subtype. *Metabolism.* 91, 39-42.
Lê Cao, K, Boitard, S and Besse, P (2011) Sparse pls discriminant analysis: Biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics.* 12, 253.
Le Donne, M, Alibrandi, A, Vita, R, Zanghı, D, Triolo, O and Benvenga, S (2016) Does eating oily fish improve gestational and neonatal outcomes ? Findings from a sicilian study. *Women and Birth.* 29, 50-57.
Leenders, M and van Loon, LJ (2011) Leucine as a pharmaconutrient to prevent and treat sarcopenia and type 2 diabetes. *Nutr. Rev.* 69, 675-689.
Li, J, Guasch-Ferre, M, Chung, W, Ruiz-Canela, M, Toledo, E, Corella, D, Bhupathiraju, SN, Tobias, DK, Tabung, FK, Hu, J, Zhao, T, Turman, C, Feng, YA, Clish, CB, Mucci, L, Eliassen, AH, Costenbader, KH, Karlson, EW, Wolpin, BM, Ascherio, A, Rimm, EB, Manson, JE, Qi, L, Martinez-Gonzalez, MA, Salas-Salvado, J, Hu, FB and Liang, L (2020) The mediterranean diet, plasma metabolome, and cardiovascular disease risk. *Eur Heart J.* 41, 2645-2656.
Liang, Y, Gong, Y, Zhang, X, Yang, D, Zhao, D and Quan, L (2018) Dietary protein intake, meat consumption, and dairy consumption in the year preceding pregnancy and during pregnancy and their associations with the risk of gestational diabetes mellitus: A prospective cohort study in southwest china. *Front. Endocrinol.* 9, 1-9.
Liu, J, van Klinken, JB, Semiz, S, van Dijk, KW, Verhoeven, A, Hankemeier, T, Harms, AC, Sijbrands, E, Sheehan, NA, van Duijn, CM and Demirkan, A (2017) A mendelian randomization study of metabolite profiles, fasting glucose, and type 2 diabetes. *Diabetes.* 66, 2915-2926.
Liu, T, Li, J, Xu, F, Wang, M, Ding, S, Xu, H and Dong, F (2016) Comprehensive analysis of serum metabolites in gestational diabetes mellitus by uplc/q-tof-ms. *Anal. Bioanal. Chem.* 408, 1125-1135.
Liu, Y, Kuang, A, Talbot, O, Bain, JR, Muehlbauer, MJ, Hayes, MG, Ilkayeva, OR, Lowe, LP, Metzger, BE, Newgard, CB, Scholtens, DM and Lowe, WL, Jr. (2020) Metabolomic and genetic associations with insulin resistance in pregnancy. *Diabetologia.* 63, 1783-1795.
Looman, M, Schoenaker, DAJM, Soedamah-muthu, SS, Geelen, A, Feskens, EJM and Mishra, GD (2018) Pre-pregnancy dietary carbohydrate quantity and quality, and risk of developing gestational diabetes: The australian longitudinal study on women's health. *Br. J. Nutr.* 435-444.
Lorenzo-Almorós, A, Hang, T, Peiró, C, Soriano-Guillén, L, Egido, J, Tuñón, J and Lorenzo, Ó (2019) Predictive and diagnostic biomarkers for gestational diabetes and its associated metabolic and cardiovascular diseases. *Cardiovasc, Diabetol.* 18.
Machiela, MJ and Chanock, SJ (2015) Ldlink: A web-based application for exploring population-specific haplotype structure and linking

correlated alleles of possible functional variants. *Bioinformatics.* 31, 3555-3557.
Macotela, Y, Emanuelli, B, Bång, AM, Espinoza, DO, Boucher, J, Beebe, K, Gall, W and Kahn, CR (2011) Dietary leucine--an environmental modifier of insulin resistance acting on multiple levels of metabolism. *PLoS One.* 6, e21187.
Mak, JKL, Pham, NM, Lee, AH, Tang, L, Pan, X-f, Binns, CW and Sun, X (2018) Dietary patterns during pregnancy and risk of gestational diabetes : A prospective cohort study in estern China. *Nutr. J.* 17, 1-11.
Makshood, M, Post, WS and Kanaya, AM (2019) Lipids in south asians: Epidemiology and management. *Curr. Cardiovasc. Risk Rep.* 13l, 24.
*Rcompanion: Functions to support extension education program evaluation.* 2021.
Mao, X, Chen, X, Chen, C, Zhang, H and Law, KP (2017) Metabolomics in gestational diabetes. *Clinica Chimica Acta.* 475, 116-127.
Marees, AT, de Kluiver, H, Stringer, S, Vorspan, F, Curis, E, Marie-Claire, C and Derks, EM (2018) A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *Int J Methods Psychiatr Res.* 27, e1608.
Markovic, TP, Muirhead, R, Overs, S, Ross, GP, Chun, J, Louie, Y, Kizirian, N, Denyer, G, Petocz, P and Hyett, J (2016) Randomized controlled trial investigating the effects of a low – glycemic index diet on pregnancy outcomes in women at high risk of gestational diabetes mellitus : The GI baby 3 study. *Diabetes Care.* 39, 31-38.
Mauro, AK, Rengarajan, A, Albright, C and D.S, B (2022) Fatty acids in normal and pathological pregnancies. *Mol. Cell. Endocrinol.* 539l.
Mayo Clinic. 2022. *Hdl cholesterol: How to boost your 'good' cholesterol.* [Online]. [Accessed 11/04/2022]. Available from: https://www.mayoclinic.org/diseases-conditions/high-blood-cholesterol/in-depth/hdl-cholesterol/art-20046388
McCabe, CF and Perng, W (2017) Metabolomics of diabetes in pregnancy. *Curr Diab Rep.* 17, 57.
McDonald, R, Karahalios, A, Le, T and Said, J (2015) A retrospective analysis of the relationship between ethnicity, body mass index, and the diagnosis of gestational diabetes in women attending an australian antenatal clinic. *Int J Endocrinol.* 2015, 297420.
Mchiza, ZJ, Goedecke, JH and Lambert, EV (2010) Accuracy of reporting food energy intake: Influence of ethnicity and body weight status in south african women. *SAJCN.* 23, 84-89.
McIntyre, HD, Catalano, P, Zhang, C, Desoye, G, Mathiesen, ER and Damm, P (2019) Gestational diabetes mellitus. *Nat Rev Dis Primers.* 5, 47.
McIntyre, HD, Kapur, A, Divakar, H and Hod, M (2020) Gestational diabetes mellitus-innovative approach to prediction, diagnosis, management, and prevention of future ncd-mother and offspring. *Front Endocrinol (Lausanne).* 11, 614533.
MedlinePlus. 2021. *Apob gene.* [Online]. [Accessed 01/04/2022]. Available from: https://medlineplus.gov/genetics/gene/apob/
Menard, V, Sotunde, OF and Weiler, HA (2020) 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.* 44, 139-145 e131.
Mensink, RP, Zock, PL, Kester, AD and Katan, MB (2003) Effects of dietary fatty acids and carbohydrates on the ratio of serum total to hdl cholesterol and on serum lipids and apolipoproteins: A meta-analysis of 60 controlled trials. *AJCN.* 77, 1146-1155.
Mijatovic-Vukas, J, Capling, L, Cheng, S, Stamatakis, E, Louie, J, Cheung, NW, Markovic, T, Ross, G, Senior, A, Brand-Miller, JC and Flood, VM (2018) Associations of diet and physical activity with risk for gestational diabetes mellitus: A systematic review and meta-analysis. *Nutrients.* 10, 698.
Mills, HL, Patel, N, White, SL, Pasupathy, D, Briley, AL, Santos Ferreira, DL, Seed, PT, Nelson, SM, Sattar, N, Tilling, K, Poston, L, Lawlor, DA and Consortium, U (2019) 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.* 17, 15.
Mohanty, AF, Siscovick, DS, Williams, MA, Thompson, ML, Burbacher, TM and Enquobahrie, DA (2015) Periconceptional seafood intake and pregnancy complications. *Public Health Nutr.* 19l, 1795-1803.
Mokkala, K, Vahlberg, T, Pellonperä, O, Houttu, N, Koivuniemi, E and Laitinen, K (2020) Distinct metabolic profile in early pregnancy of overweight and obese women developing gestational diabetes. *J Nutr.* 150, 31-37.
Monni, G, Atzori, L, Corda, V, Dessolis, F, Iuculano, A, Hurt, KJ and Murgia, F (2021) Metabolomics in prenatal medicine: A review. *Front. Med.* 8l.
Moon, JH, Kwak, SH and Jang, HC (2017) Prevention of type 2 diabetes mellitus in women with previous gestational diabetes mellitus. *Korean J. Intern. Med.* 32, 26-41.
Myers, TA, Chanock, SJ and Machiela, MJ (2020) Ldlinkr: An r package for rapidly calculating linkage disequilibrium statistics in diverse populations. *Front Genet.* 11, 157.
National Health Service (NHS). 2019. *Treatment - gestational diabetes*. Unpublished.
Newman, C and Dunne, FP (2021) Metformin for pregnancy and beyond: The pros and cons. *Diabet Med.* e14700.
Newsholme, P, Brennan, L, Rubi, B and Maechler, P (2005) New insights into amino acid metabolism, beta-cell function and diabetes. *Clin Sci (Lond).* 108, 185-194.
NHS. 2019. *Treatment: Gestational diabetes.* [Online]. [Accessed 09/02/22]. Available from: https://www.nhs.uk/conditions/gestational-diabetes/treatment/
NHS. 2020. *Have a healthy diet in pregnancy.* [Online]. [Accessed 05/04/2022]. Available from: https://www.nhs.uk/pregnancy/keeping-well/have-a-healthy-diet/
Nie, C, He, T, Zhang, W, Zhang, G and Ma, X (2018) Branched chain amino acids: Beyond nutrition metabolism. *Int J Mol Sci.* 19,l.
Nielsen, KK, Damm, P, Bygbjerg, IC and Kapur, A (2018) Barriers and facilitators for implementing programmes and services to address hyperglycaemia in pregnancy in low and middle income countries: A systematic review. *Diabetes Res Clin Pract.* 145, 102-118.

NIH. 2020a. *A breif guide to genomics.* [Online]. [Accessed 10/02/2022]. Available from: https://www.genome.gov/about-genomics/fact-sheets/A-Brief-Guide-to-Genomics
NIH. 2020b. *Human genome project faq.* [Online]. [Accessed 05/04/2022]. Available from: https://www.genome.gov/human-genome-project/Completion-FAQhttps://www.genome.gov/human-genome-project/Completion-FAQ
Nishikawa, E, Oakley, L, Seed, PT, Doyle, P and Oteng-Ntim, E (2017) Maternal bmi and diabetes in pregnancy: Investigating variations between ethnic groups using routine maternity data from London, UK. *PLoS One.* 12, e0179332.
O'Malley, EG, Reynolds, CME, Killalea, A, O'Kelly, R, Sheehan, SR and Turner, MJ (2020) Maternal obesity and dyslipidemia associated with gestational diabetes mellitus (gdm). *Eur J Obstet Gynecol Reprod Biol.* 246, 67-71.
Odar, E, Wandabwa, J and Kiondo, P (2004) Maternal and fetal outcome of gestational diabetes mellitus in Mulago Hospital, Uganda. *Afr. Health Sci.* 4l, 9-14.
Ooi, EM, Ng, TW, Watts, GF and Barrett, PH (2013) Dietary fatty acids and lipoprotein metabolism: New insights and updates. *Curr Opin Lipidol.* 24, 192-197.
Opie, RS, Neff, M and Tierney, AC (2016) A behavioural nutrition intervention for obese pregnant women : Effects on diet quality , weight gain and the incidence of gestational diabetes. *ANZJOG.* 56, 364-373.
Orsini, N, Li, R, Wolk, A, Khudyakov, P and Spiegelman, D (2012) Meta-analysis for linear and nonlinear dose-response relations: Examples, an evaluation of approximations, and software. *Am J Epidemiol.* 175, 66-73.
Osorio-Yáñez, C, Gelaye, B, Qiu, C, Bao, W, Cardenas, A, Enquobahrie, DE and Williams, MA (2017) Maternal intake of fried foods and risk of gestational diabetes mellitus. *Ann. Epidemiol.* 27, 384-390.
Östman, E, Samigullin, A, Heyman-Lindén, L, Andersson, K, Björck, I, Öste, R and Humpert, PM (2020) A novel nutritional supplement containing amino acids and chromium decreases postprandial glucose response in a randomized, double-blind, placebo-controlled study. *PLoS One.* 15, e0234237.
Pan, X-F, Huang, Y, Li, X, Wang, Y, Ye, Y, Chen, H, Marklund, M, Wen, Y, Liu, Y, Zeng, H, Qi, X, Yang, X, Yang, C-X, Liu, G, Gibson, RA, Xu, S, Yu, D, Chen, D, Li, Y, Mei, Z, Pan, A and Wu, JHY (2021) Circulating fatty acids and risk of gestational diabetes mellitus: Prospective analyses in China. *Eur. J. Endocrinol.* 185, 87-97.
Pang, WW, Colega, M, Cai, S, Chan, YH, Chen, L-w, Soh, S-e, Han, WM, Tan, KH, Lee, YS and Saw, S-m (2017) Higher maternal dietary protein intake is associated with a higher risk of gestational diabetes mellitus in a multiethnic Asian cohort. *J Nutr.* 653-660.
Pathirana, MM, Lassi, ZS, Roberts, CT and Andraweera, PH (2020) Cardiovascular risk factors in offspring exposed to gestational diabetes mellitus in utero: systematic review and meta-analysis. *J. DOHaD.* 11, 599-616.

Perperoglou, A, Sauerbrei, W, Abrahamowicz, M and Schmid, M (2019) A review of spline function procedures in r. *BMC Med. Res. Methodol.* 19, 46.
Perreault, M, Zulyniak, MA, Badoud, F, Stephenson, S, Badawi, A, Buchholz, A and Mutch, DM (2014) A distinct fatty acid profile underlies the reduced inflammatory state of metabolically healthy obese individuals. *PloS one.* 9, e88539-e88539.
Phonyiam, R and Berry, DC (2021) Racial and ethnic disparities in health care and health outcomes for pregnant women with diabetes. *NWH.* 25, 437-449.
Plows, JF, Stanley, JL, Baker, PN, Reynolds, CM and Vickers, MH (2018) The pathophysiology of gestational diabetes mellitus. *Int J Mol Sci.* 19,11.
Popejoy, AB and Fullerton, SM (2016) Genomics is failing on diversity *Nature.* 538, 161-164.
Popova, PV, Pustozerov, EA, Tkachuk, AS and Grineva, EN (2021) Improving nutrition for the prevention of gestational diabetes: Current status and perspectives. *World J Diabetes.* 12, 1494-1506.
Qiao, Y, Tinker, L, Olendzki, BC, Hébert, JR, Balasubramanian, R, Rosal, MC, Hingle, M, Song, Y, Schneider, KL, Liu, S, Sims, S, Ockene, JK, Sepavich, DM, Shikany, JM, Persuitte, G and Ma, Y (2014) Racial/ethnic disparities in association between dietary quality and incident diabetes in postmenopausal women in the united states: The women's health initiative 1993–2005. *Ethn. Health.* 19, 328-347.
*R: A language and environment for statistical computing.* 2020. Vienna, Austria: R Foundation for Statistical Computing
Rai, AS, Sletner, L, Jenum, AK, Overby, NC, Stafne, SN, Lekva, T, Pripp, AH and Sagedal, LR (2021) Identifying women with gestational diabetes based on maternal characteristics: An analysis of four Norwegian prospective studies. *BMC Pregnancy Childbirth.* 21, 615.
Ralston, JC, Zulyniak, MA, Nielsen, DE, Clarke, S, Badawi, A, El-Sohemy, A, Ma, DW and Mutch, DM (2013) Ethnic-and sex-specific associations between plasma fatty acids and markers of insulin resistance in healthy young adults. *Nutr. Metab.* 10, 42.
Raynor, P and Born in Bradford Collaborative, G (2008) Born in Bradford, a cohort study of babies Born in Bradford, and their parents: Protocol for the recruitment phase. *BMC Public Health.* 8, 327.
Reactome. 2022. *Rgl1 [cytosol].* [Online]. [Accessed 01/04/2022]. Available from: https://reactome.org/content/detail/R-HSA-561160
Read, SH, Rosella, LC, Berger, H, Feig, DS, Fleming, K, Ray, JG, Shah, BR and Lipscombe, LL (2021) BMI and risk of gestational diabetes among women of South Asian and Chinese ethnicity: A population-based study. *Diabetologia.* 64, 805-813.
Read, SH, Wu, W, Ray, JG, Lowe, J, Feig, DS and Lipscombe, LL (2019) Characteristics of women with gestational diabetes from non-caucasian compared with Caucasian ethnic groups. *Can J Diabetes.* 43, 600-605.
Rhee, EP. 2020. Chapter 6 - genetic influence on the metabolome. In: Adamski, J ed. *Metabolomics for biomedical research.* Academic Press, pp.105-121.

Rivadeneira, F and Uitterlinden, AG. 2021. Chapter 18 - genetics of osteoporosis. In: Dempster, DW, et al. eds. *Marcus and Feldman's osteoporosis (fifth edition).* Academic Press, pp.405-451.
Rondanelli, M, Nichetti, M, Peroni, G, Faliva, MA, Naso, M, Gasparri, C, Perna, S, Oberto, L, Di Paolo, E, Riva, A, Petrangolini, G, Guerreschi, G and Tartara, A (2021) 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.* 7, 622391-622391.
Rönö, K, Elisabeth, N, Miira, G, Klemetti, M, Stach-lempinen, B, Huvinen, E, Meinilä, J, Valkama, A, Tiitinen, A and Paavo, R (2018) Effect of a lifestyle intervention during pregnancy — findings from the Finnish gestational diabetes prevention trial (RADIEL). *J. Perinatol.* 38, 1157-1164.
Rosenstein, MG, Cheng, YW, Snowden, JM, Nicholson, JM, Doss, AE and Caughey, AB (2012) The risk of stillbirth and infant death stratified by gestational age in women with gestational diabetes. *AJOG.* 206, 309.e301-309.e3097.
Ryckman, KK, Spracklen, CN, Smith, CJ, Robinson, JG and Saftlas, AF (2015) Maternal lipid levels during pregnancy and gestational diabetes: A systematic review and meta-analysis. *BJOG.* 122, 643-651.
Saeedi, M, Cao, Y, Fadl, H, Gustafson, H and Simmons, D (2021) Increasing prevalence of gestational diabetes mellitus when implementing the iadpsg criteria: A systematic review and meta-analysis. *Diabetes Res Clin Pract.* 172, 108642.
Sahariah, SA, Potdar, RD, Gandhi, M, Kehoe, SH, Brown, N, Sane, H, Coakley, PJ, Marley-Zagar, E, Chopra, H, Shivshankaran, D, Cox, VA, Jackson, AA, Margetts, BM and Fall, CH (2016) A daily snack containing leafy green vegetables, fruit, and milk before and during pregnancy prevents gestational diabetes in a randomized, controlled trial in Mumbai, India. *J. Nutr.* 146, 1453S-1460S.
Sanchalika, A and Teresa, J (2015) Risk of gestational diabetes among south asian immigrants living in new jersey--a retrospective data review. *J Racial Ethn Health Disparities.* 2, 510-516.
Sanchis, AM, Díaz, G, Javier, JF, Gortari, B, De, C, Arrillaga, F and González, MAM (2018) Association between pre-pregnancy consumption of meat , iron intake , and the risk of gestational diabetes : The sun project. *Eur. J. Nutr.* 57, 939-949.
Santos Ferreira, DL, Williams, DM, Kangas, AJ, Soininen, P, Ala-Korpela, M, Smith, GD, Jarvelin, M-R and Lawlor, DA (2017) Association of pre-pregnancy body mass index with offspring metabolic profile: Analyses of 3 european prospective birth cohorts. *PLOS Medicine.* 14, e1002376.
Schaefer-Graf, U, Napoli, A, Nolan, CJ and Diabetic Pregnancy Study, G (2018) Diabetes in pregnancy: A new decade of challenges ahead. *Diabetologia.* 61, 1012-1021.
Schiavone, M, Putoto, G, Laterza, F and Pizzol, D (2016) Gestational diabetes: An overview with attention for developing countries. *Endocr Regul.* 50, 62-71.
Schoenaker, DA, Mishra, GD, Callaway, LK and Soedamah-Muthu, SS (2016) The role of energy, nutrients, foods, and dietary patterns in the

development of gestational diabetes mellitus: A systematic review of observational studies. *Diabetes Care.* 39, 16-23.
Schoenaker, DA, Soedamah-muthu, SS, Callaway, LK and Mishra, GD (2015) Pre-pregnancy dietary patterns and risk of gestational diabetes mellitus: Results from an Australian population-based prospective cohort study. *Diabetologia.* 58, 2726-2735.
Schwartz, N, Nachum, Z and Green, MS (2015) The prevalence of gestational diabetes mellitus recurrence--effect of ethnicity and parity: A meta analysis. *Am J Obstet Gynecol.* 213, 310-317.
Sedhain, A, Bhattarai, GR, Yadav, SR, Pandey, BR and Pant, TP (2020) Geographic and seasonal variation of vitamin D: A retrospective study in two centers of Nepal. *J Nepal Health Res Counc.* 18l, 103-107.
Seldin, MF, Shigeta, R, Villoslada, P, Selmi, C, Tuomilehto, J, Silva, G, Belmont, JW, Klareskog, L and Gregersen, PK (2006) European population substructure: Clustering of northern and southern populations. *PloS Genetics.* 2,5.
Shah, NS, Wang, MC, Freaney, PM, Perak, AM, Carnethon, MR, Kandula, NR, Gunderson, EP, Bullard, KM, Grobman, WA, O'Brien, MJ and Khan, SS (2021) Trends in gestational diabetes at first live birth by race and ethnicity in the US, 2011-2019. *JAMA.* 326, 660-669.
Sheehan, NA, Meng, S and Didelez, V. 2011. Mendelian randomisation: A tool for assessing causality in observational epidemiology. In: Teare, MD ed. *Genetic epidemiology.* Totowa, NJ: Humana Press, pp.153-166.
Sheridan, E, Wright, J, Small, N, Corry, PC, Oddie, S, Whibley, C, Petherick, ES, Malik, T, Pawson, N, McKinney, PA and Parslow, RC (2013) Risk factors for congenital anomaly in a multiethnic birth cohort: An analysis of the Born in Bradford study. *Lancet.* 382, 1350-1359.
Simmons, D, Devlieger, R, Assche, AV, Jans, G, Galjaard, S, Corcoy, R, Adelantado, JM, Dunne, F, Desoye, G, Harreiter, J and Kautzky-willer, A (2017) Effect of physical activity and/or healthy eating on GDM risk: The Dali Lifestyle study. *J. Clin. Endocr.* 102, 903-913.
Skrivankova, VW, Richmond, RC, Woolf, BAR, Davies, NM, Swanson, SA, VanderWeele, TJ, Timpson, NJ, Higgins, JPT, Dimou, N, Langenberg, C, Loder, EW, Golub, RM, Egger, M, Davey Smith, G and Richards, JB (2021) Strengthening the reporting of observational studies in epidemiology using mendelian randomisation (strobe-mr): Explanation and elaboration. *BMJ.* 375l, n2233.
Smith, JM (2010) Maternal-fetal nutrition during pregnancy and lactation. *J. Obstet. Gynaecol.* 30l, 766-766.
Sonagra, AD, Biradar, SM, K, D and Murthy D S, J (2014) Normal pregnancy- a state of insulin resistance. *J. Clin. Diagnostic Res.* 8l, CC01-CC03.
Sotos-Prieto, M and Mattei, J (2018) Mediterranean diet and cardiometabolic diseases in racial/ethnic minority populations in the United States. *Nutrients.* 10.
Staley, JR, Blackshaw, J, Kamat, MA, Ellis, S, Surendran, P, Sun, BB, Paul, DS, Freitag, D, Burgess, S, Danesh, J, Young, R and Butterworth, AS (2016) Phenoscanner: A database of human genotype-phenotype associations. *Bioinformatics.* 32, 3207-3209.

*Stata statistical software: Release 17.* 2017. College Station, TX: StataCorp LLC.
Sterne, JAC, Egger, M, Moher, D and Boutron, I. 2017. Chapter 10: Addressing reporting biases. In: Higgins Jpt, CR, Chandler J, Cumpston Ms ed. *Cochrane handbook for systematic reviews of interventions* Cochrane.
Strandberg, RB, Iversen, MM, Jenum, AK, Sorbye, LM, Vik, ES, Schytt, E, Aasheim, V and Nilsen, RM (2021) Gestational diabetes mellitus by maternal country of birth and length of residence in immigrant women in Norway. *Diabet Med.* 38, e14493.
Super, S and Wagemakers, A (2021) Understanding empowerment for a healthy dietary intake during pregnancy. *Int J Qual Stud Health Well-being.* 16, 1857550.
Tajima, R, Yachi, Y, Tanaka, Y and Kawasaki, YA (2017) Carbohydrate intake during early pregnancy is inverselyassociated with abnormal glucose challenge test results in Japanese pregnant women. *Diabtetes Metabolism Research and Reviews.* 33, 1.
Tam, V, Patel, N, Turcotte, M, Bosse, Y, Pare, G and Meyre, D (2019) Benefits and limitations of genome-wide association studies. *Nat Rev Genet.* 20, 467-484.
Tarry-Adkins, JL, Ozanne, SE and Aiken, CE (2021) Impact of metformin treatment during pregnancy on maternal outcomes: A systematic review/meta-analysis. *Sci Rep.* 11, 9240.
Taylor, K, Ferreira, DLS, West, J, Yang, T, Caputo, M and Lawlor, DA (2019) Differences in pregnancy metabolic profiles and their determinants between white European and South Asian women: Findings from the born in bradford cohort. *Metabolites.* 9.
Taylor, K, McBride, N, Goulding, NJ, Burrows, K, Mason, D, Pembrey, L, Yang, T, Azad, R, Wright, J and Lawlor, DA (2021) Metabolomics datasets in the Born in Bradford cohort. *Wellcome Open Research.* 5l.
Tebani, A and Bekri, S (2019) Paving the way to precision nutrition through metabolomics. *Front Nutr.* 6, 41.
Teede, HJ, Bailey, C, Moran, LJ, Bahri Khomami, M, Enticott, J, Ranasinha, S, Rogozinska, E, Skouteris, H, Boyle, JA, Thangaratinam, S and Harrison, CL (2022) Association of antenatal diet and physical activity-based interventions with gestational weight gain and pregnancy outcomes: A systematic review and meta-analysis. *JAMA Intern Med.* 182, 106-114.
Thanawala, U, Divakar, H, Jain, R and Agarwal, MM (2021) Negotiating gestational diabetes mellitus in india: A national approach. *Medicina (Kaunas).* 57.
*Review Manager (Revman)* Cochrane Collaboration. 2020.
Thevenot, EA, Roux, A, Xu, Y, Ezan, E and Junot, C (2015) Analysis of the human adult urinary metabolome variations with age, body mass index, and gender by implementing a comprehensive workflow for univariate and opls statistical analyses. *J Proteome Res.* 14, 3322-3335.
Tobias, DK, Zhang, C, Chavarro, J, Bowers, K, Rich-edwards, J, Rosner, B, Mozaffarian, D and Hu, FB (2012) Prepregnancy adherence to dietary patterns and lower risk of gestational diabetes mellitus. *AJCN.* 96, 289-295.

Tryggvadottir, EA, Medek, H, Birgisdottir, BE, Geirsson, RT and Gunnarsdottir, I (2015) Association between healthy maternal dietary pattern and risk for gestational diabetes mellitus. *Eur. J. Clin. Nutr.* 70, 237-242.

UCSF Health. 2022. *Dietary recommendations for gestational diabetes.* [Online]. [Accessed 09/02/2022]. Available from: https://www.ucsfhealth.org/education/dietary-recommendations-for-gestational-diabetes

Uffelmann, E, Huang, QQ, Munung, NS, de Vries, J, Okada, Y, Martin, AR, Martin, HC, Lappalainen, T and Posthuma, D (2021) Genome-wide association studies. *Nat. Rev Methods.* 1, 59.

Unnikrishnan, R, Anjana, RM and Mohan, V (2016) Diabetes mellitus and its complications in India. *Nat Rev Endocrinol.* 12l, 357-370.

Valentini, R, Dalfra, MG, Masin, M, Barison, A, Marialisa, M, Pegoraro, E and Lapolla, A (2012) A pilot study on dietary approaches in multiethnicity: Two methods compared. *Int J Endocrinol.* 2012, 985136.

van den Berg, S, Vandenplas, J, van Eeuwijk, FA, Lopes, MS and Veerkamp, RF (2019) Significance testing and genomic inflation factor using high-density genotypes or whole-genome sequence data. *J Anim Breed Genet.* 136, 418-429.

Villafan-Bernal, JR, Acevedo-Alba, M, Reyes-Pavon, R, Diaz-Parra, GA, Lip-Sosa, DL, Vazquez-Delfin, HI, Hernandez-Muñoz, M, Bravo-Aguirre, DE, Figueras, F and Martinez-Portilla, RJ (2019) Plasma levels of free fatty acids in women with gestational diabetes and its intrinsic and extrinsic determinants: Systematic review and meta-analysis. *J. Diabetes Res.* 2019, 7098470-7098470.

Visscher, PM, Hill, WG and Wray, NR (2008) Heritability in the genomics era — concepts and misconceptions. *Nat. Rev. Genet.* 9, 255-266.

Viswanathan, M, Berkman, ND, Dryden, DM and Hartling, L. 2013. Ahrq methods for effective health care. *Assessing risk of bias and confounding in observational studies of interventions or exposures: Further development of the rti item bank.* Rockville (MD): Agency for Healthcare Research and Quality (US).

Vounzoulaki, E, Khunti, K, Abner, SC, Tan, BK, Davies, MJ and Gillies, CL (2020) Progression to type 2 diabetes in women with a known history of gestational diabetes: Systematic review and meta-analysis. *BMJ.* 369, m1361.

Wang, DD and Hu, FB (2018) Precision nutrition for prevention and management of type 2 diabetes. *Lancet Diabetes Endocrinol.* 6, 416-426.

Wang, J, Li, Z and Lin, L (2019) Maternal lipid profiles in women with and without gestational diabetes mellitus. *Medicine (Baltimore).* 98, e15320.

Wang, Q, Würtz, P, Auro, K, Mäkinen, VP, Kangas, AJ, Soininen, P, Tiainen, M, Tynkkynen, T, Jokelainen, J, Santalahti, K, Salmi, M, Blankenberg, S, Zeller, T, Viikari, J, Kähönen, M, Lehtimäki, T, Salomaa, V, Perola, M, Jalkanen, S, Järvelin, MR, Raitakari, OT, Kettunen, J, Lawlor, D, A. and Ala-Korpela, M (2016) Metabolic profiling of pregnancy: Cross-sectional and longitudinal evidence. *BMC Medicine.* 14, 205.

Wang, QY, You, LH, Xiang, LL, Zhu, YT and Zeng, Y (2021) Current progress in metabolomics of gestational diabetes mellitus. *World J Diabetes.* 12, 1164-1186.
Weissgerber, TL and Mudd, LM (2015) Preeclampsia and diabetes. *Curr. Diabetes Rep.* 15, 9-9.
WHO Expert Consultation (2004) Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Lancet.* 363, 157-163.
Wiesner, P and Watson, KE (2017) Triglycerides: A reappraisal. *Trends Cardiovasc. Med.* 27, 428-432.
Willer, CJ, Li, Y and Abecasis, GR (2010) Metal: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 26, 2190-2191.
Wong, NKP, Nicholls, SJ, Tan, JTM and Bursill, CA (2018) The role of high-density lipoproteins in diabetes and its vascular complications. *Int. J. Mol. Sci.* 19, 1680.
World Health Organisation (WHO). 2020. *Diabetes*. Unpublished.
World Health Organization Consultation (2014) Diagnostic criteria and classification of hyperglycaemia first detected in pregnancy: A world health organization guideline. *Diabetes Res Clin Pract.* 103l, 341-363.
Worley, B and Powers, R (2013) Multivariate analysis in metabolomics. *Curr Metabolomics.* 1, 92-107.
Wright, J, Small, N, Raynor, P, Tuffnell, D, Bhopal, R, Cameron, N, Fairley, L, Lawlor, DA, Parslow, R, Petherick, ES, Pickett, KE, Waiblinger, D, West, J and Born in Bradford Scientific Collaborators, G (2013) Cohort profile: The Born in Bradford multi-ethnic family cohort study. *Int J Epidemiol.* 42, 978-991.
Wu, NN, Zhao, D, Ma, W, Lang, JN, Liu, SM, Fu, Y, Wang, X, Wang, ZW and Li, Q (2021) A genome-wide association study of gestational diabetes mellitus in Chinese women. *J Matern Fetal Neonatal Med.* 34, 1557-1564.
Wu, Y, Dong, Y, Atefi, M, Liu, Y, Elshimali, Y and Vadgama, JV (2016) Lactate, a neglected factor for diabetes and cancer interaction. *Mediators Inflamm.* 2016, 6456018.
Wurtz, P, Havulinna, AS, Soininen, P, Tynkkynen, T, Prieto-Merino, D, Tillin, T, Ghorbani, A, Artati, A, Wang, Q, Tiainen, M, Kangas, AJ, Kettunen, J, Kaikkonen, J, Mikkila, V, Jula, A, Kahonen, M, Lehtimaki, T, Lawlor, DA, Gaunt, TR, Hughes, AD, Sattar, N, Illig, T, Adamski, J, Wang, TJ, Perola, M, Ripatti, S, Vasan, RS, Raitakari, OT, Gerszten, RE, Casas, JP, Chaturvedi, N, Ala-Korpela, M and Salomaa, V (2015) Metabolite profiling and cardiovascular event risk: A prospective study of 3 population-based cohorts. *Circulation.* 131, 774-785.
Xia, F, He, C, Ren, M, Xu, F-G and Wan, J-B (2020) Quantitative profiling of eicosanoids derived from n-6 and n-3 polyunsaturated fatty acids by twin derivatization strategy combined with lc-ms/ms in patients with type 2 diabetes mellitus. *Analytica Chimica Acta.* 1120, 24-35.
Xu, T, Dainelli, L, Yu, K, Ma, L, Silva Zolezzi, I, Detzel, P and Fang, H (2017) The short-term health and economic burden of gestational diabetes mellitus in China: A modelling study. *BMJ Open.* 7, e018893.
Yamamoto, JM, Kellett, JE, Balsells, M, Garcia-Patterson, A, Hadar, E, Sola, I, Gich, I, van der Beek, EM, Castaneda-Gutierrez, E, Heinonen, S,

Hod, M, Laitinen, K, Olsen, SF, Poston, L, Rueda, R, Rust, P, van Lieshout, L, Schelkle, B, Murphy, HR and Corcoy, R (2018) Gestational diabetes mellitus and diet: A systematic review and meta-analysis of randomized controlled trials examining the impact of modified dietary interventions on maternal glucose control and neonatal birth weight. *Diabetes Care.* 41, 1346-1361.

Yang, J, Lee, SH, Goddard, ME and Visscher, PM (2011) GCTA: A tool for genome-wide complex trait analysis. *Am J Hum Genet.* 88, 76-82.

Yang, J, Zaitlen, NA, Goddard, ME, Visscher, PM and Price, AL (2014) Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet.* 46, 100-106.

Yi, DUH, Hong, J, Karmin, O, Bo, C and Ji, XL (2017) Association of dietary pattern during pregnancy and gestational diabetes mellitus : A prospective cohort. *Biomed. Environ. Sci.* 30, 887-897.

Yong, HY, Shariff, ZM, Yusof, B-nM and Rejali, Z (2020) The association between dietary patterns before and in early pregnancy and the risk of gestational diabetes mellitus ( GDM ): Data from the Malaysian secost cohort. *PLoS ONE.* 1-16.

Yu, Y, Arah, OA, Liew, Z, Cnattingius, S, Olsen, J, Sørensen, HT, Qin, G and Li, J (2019) Maternal diabetes during pregnancy and early onset of cardiovascular disease in offspring: Population based cohort study with 40 years of follow-up. *BMJ.* 367, l6398.

Yuen, L, Wong, VW and Simmons, D (2018) Ethnic disparities in gestational diabetes. *Curr Diab Rep.* 18l, 68.

Zhang, C, Schulze, MB, Solomon, CG and Hu, FB (2006) A prospective study of dietary patterns, meat intake and the risk of gestational diabetes mellitus. *Diabetologia.* 49, 2604-2613.

Zhang, C, Tobias, DK and Chavarro, JE (2014) Adherence to healthy lifestyle and risk of gestational diabetes mellitus : Prospective cohort study. *the BMJ.* 5450, 1-11.

Zhao, H, Li, H, Chung, ACK, Xiang, L, Li, X, Zheng, Y, Luan, H, Zhu, L, Liu, W, Peng, Y, Zhao, Y, Xu, S, Li, Y and Cai, Z (2019) Large-scale longitudinal metabolomics study reveals different trimester-specific alterations of metabolites in relation to gestational diabetes mellitus. *J Proteome Res.* 18, 292-300.

Zhou, SJ, Yelland, L, McPhee, AJ, Quinlivan, J, Gibson, RA and Makrides, M (2012) Fish-oil supplementation in pregnancy does not reduce the risk of gestational diabetes or preeclampsia. *Am J Clin Nutr.* 95, 1378-1384.

Zhou, X, Chen, R, Zhong, C, Wu, J, Li, X, Li, Q, Cui, W, Yi, N, Xiao, M, Yin, H, Xiong, G, Han, W, Hao, L, Yang, X and Yang, N (2018) Maternal dietary pattern characterised by high protein and low carbohydrate intake in pregnancy is associated with a higher risk of gestational diabetes mellitus in Chinese women: A prospective cohort study. *Br. J. Nutr.* 210, 1045-1055.

Zhu, Y, Li, M, Rahman, ML, Hinkle, SN, Wu, J, Weir, NL, Lin, Y, Yang, H, Tsai, MY, Ferrara, A and Zhang, C (2019) Plasma phospholipid n-3 and n-6 polyunsaturated fatty acids in relation to cardiometabolic markers and gestational diabetes: A longitudinal study within the prospective nichd fetal growth studies. *PLoS Med.* 16, e1002910.

Zinsser, LA, Stoll, K, Wieber, F, Pehlke-Milde, J and Gross, MM (2020) Changing behaviour in pregnant women: A scoping review. *Midwifery.* 85l, 102680.

Zulyniak, MA, de Souza, RJ, Shaikh, M, Desai, D, Lefebvre, DL, Gupta, M, Wilson, J, Wahi, G, Subbarao, P, Becker, AB, Mandhane, P, Turvey, SE, Beyene, J, Atkinson, S, Morrison, KM, McDonald, S, Teo, KK, Sears, MR, Anand, SS and Alliance Nutrigen (2017) Does the impact of a plant-based diet during pregnancy on birth weight differ by ethnicity ? A dietary pattern analysis from a prospective Canadian birth cohort alliance. *BMJ Open.* 7,11.

Zulyniak, MA, Fuller, H and Iles, MM (2020) Investigation of the causal association between long-chain n-6 polyunsaturated fatty acid synthesis and the risk of type 2 diabetes: A mendelian randomization analysis. *Lifestyle Genom.* 13, 146-153.

Zulyniak, MA and Mutch, DM (2011) Harnessing metabolomics for nutrition research. *Curr. Pharm Biotechnol.* 12, 1005-1015.