medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Running title: Ethnic-specific metabolic profiles of GDM

Unique Metabolic Profiles Associate with Gestational Diabetes and Ethnicity in Low and High-Risk Women Living in the UK

Authors: Harriett Fuller [a], Mark Iles [b,c], J. Bernadette Moore[a] and Michael A. Zulyniak[a]*

[a] Nutritional Epidemiology Group, School of Food Science and Nutrition, University of Leeds, Woodhouse Lane, Woodhouse, Leeds, LS2 9JT, UK.
[b] Leeds Institute of Medical Research, Worsley Building, University of Leeds, Clarendon Way, Woodhouse, Leeds, LS2 9JT, UK.
[c] Leeds Institute for Data Analytics, Worsley Building, University of Leeds, Woodhouse, University of Leeds, Clarendon Way, Woodhouse, Leeds, LS2 9JT, UK

*Corresponding Author: Michael Zulyniak. Email. m.a.zulyniak@leeds.ac.uk. Phone +44 113 343 0685

Data described in the manuscript will not be made available but can be requested through Born in Bradford (https://borninbradford.nhs.uk). Analytic code will be made available upon request pending approval by the research team.

Funding statement

MZ is currently funded by the Wellcome Trust (217446/Z/19/Z). The Born in Bradford data reported here were supported by a Wellcome Trust infrastructure grant [101597] and the National Institute for Health Research ARC Yorkshire and Humber [NIHR200166]. Funding for the metabolomics analyses has been provided by the US National Institutes of Health [R01 DK10324], the European Research Council (ERC) under the European Union's Seventh Framework Programme [FP7/2007-2013] / ERC grant agreement no 669545 and the MRC via the MRC Integrative Epidemiology Unit Programme [MC_UU_00011/6]

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

HF: No conflicts of interest

MI: No conflicts of interest

JBM: No conflicts of interest

MAZ: No conflicts of interest

MI and MAZ designed research. HF and MAZ conducted research. MI and MAZ provided essential materials. HF performed statistical analyses. HF MI JBM and MAZ wrote the paper. MAZ has the primary responsibility for the final content. All authors read and reviewed the final manuscript.

Abbreviations

ApoA1: Apolipoprotein A1

BIB: Born in Bradford

DHA: Docosahexanoic Acid

FAw3: Total Omega-3

FAw6: Total Omega-6

GDM: Gestational Diabetes Mellitus

GRM: Glycolysis Related Metabolites

HDL-C: High Density Lipoprotein Cholesterol

HD2L-C: High Density Lipoprotein-2 Cholesterol

IQR: Inter Quartile Range

LDL_D: Diameter of Low Density Lipoprotein

LPS: Lipoprotein Particle Size

MSEP: Mean Square Error of Prediction

MFA: Monounsaturated Fatty Acids

MW: Mann-Whitney

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

NHS: National Health Service

PCA: Principal Component Analyses

OGTT: Oral Glucose Tolerance Test

oPLSDA: Orthogonal Partial Least Squares Discriminatory Analyses

SAC-H: High Weight South Asian Cases

SANC-H: High Weight South Asian Non-Cases

SAC-N: Healthy Weight South Asian Cases

SANC-N: Healthy Weight South Asian Non-Cases

PLSDA: Partial Least Squares Discriminatory Analyses

RMSEE: Root Mean Square Error of Prediction

ROC: Receiver Operator Curve

SA: South Asian

SE: Standard Error

SFA: Saturated Fatty Acids

sPLSDA: Sparse Partial Least Squares Discriminatory Analyses

TCA: Tricarboxylic Acid Cycle

VIP: Variable Importance in Projection

VLDL_D: Diameter of Very Low Density Lipoprotein

WEs: White Europeans

WEC-N: Healthy Weight White European Cases

WENC-N: Healthy Weight White European Non-Cases

WEC-H: High Weight White European Cases

WENC-H: High Weight White European Non-Cases

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

Background:

Gestational Diabetes Mellitus (GDM) is the most common global pregnancy complication; however, prevalence varies substantially between ethnicities with South Asians (SA) experiencing up to 3-times the risk of the disease compared to white Europeans (WEs). Factors driving this discrepancy are unclear, although the metabolome is of great interest as GDM is known to be characterised by metabolic dysregulation.

Objective:

This primary aim was to characterise and compare the metabolic profiles of GDM in SA and WE women (at < 28 weeks' gestation) from the Born in Bradford (BIB) prospective birth cohort in the UK.

Methods:

146 fasting serum metabolites, from 2668 pregnant WE and 2671 pregnant South Asian (SA) women (average BMI 26.2 kg/m^2, average age 27.3 years) were analysed using partial least squares discriminatory analyses to characterise GDM status. Linear associations between metabolite values and post-oral glucose tolerance test measures of dysglycemia (fasting glucose and 2-hour post glucose) were also examined.

Results:

Seven metabolites associated with GDM status in both ethnicities (variable importance in projection (VIP) ≥1), while 6 additional metabolites associated with GDM only in WE women. Unique metabolic profiles were observed in healthy weight

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

women who later developed GDM, with distinct metabolite patterns identified by ethnicity and BMI status. Of the metabolite values analysed in relation to dysglycemia, lactate, histidine, apolipoprotein A1, HDL cholesterol, HDL2 cholesterol associated with decreased glucose concentration, while DHA and the diameter of very low-density lipoprotein particles (nm) associated with increased glucose concertation in WE women; while in SAs albumin alone associated with decreased glucose concentration.

Conclusions

This study shows that the metabolic risk profile for GDM differs between WE and SA women enrolled in BiB the UK. This suggests that aetiology of the disease differs between ethnic groups and that ethnic-appropriate prevention strategies may be beneficial.

Keywords: GDM, metabolomics, ethnicity, South Asians, pregnancy, maternal health, PLSDA, sPLSDA, personalised nutrition

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Introduction

During pregnancy, there is a natural increase in catabolism to ensure sufficient energy for the foetus (1, 2). This increase is governed by maternal hormones, beginning as a mild change in insulin sensitivity and progressing through hyperinsulinemia to controlled insulin resistance by the third trimester (2-5). For most pregnancies, these changes are safe and controlled, with insulin sensitivity returning to a healthy state following pregnancy. However, for approximately one in seven pregnancies, insulin resistance exceeds normal “healthy“ levels and enters a diabetic state, putting the mother and her growing offspring in danger of short- and long-term health risks (6, 7). This pregnancy-induced state of diabetes, gestational diabetes mellitus (GDM), is a major global health concern with varying prevalence between populations.

In Middle Eastern, North Africa, and South Asian countries, GDM prevalence can exceed 20% of pregnancies, whereas in European countries prevalence of GDM is more commonly around 5% (5). Numerous lifestyle, biological, and genetic factors are thought to contribute to this disparity of risk (5, 8). Despite the numerous factors, diet is the mainstay of most prevention and treatment strategies because of its demonstrated efficacy for managing glucose concentrations (9-11). Nonetheless, we and others have demonstrated that the effects of dietary prevention strategies on maternal and offspring health are not generalisable across populations or ethnic groups, with dietary patterns demonstrating varied effects between ethnic groups in relation to both GDM prevention and birth weight (12-15). These data suggest that the metabolism and pathology of GDM may differ across populations, where some ethnic groups have unique metabolic profiles that make them more susceptible to

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

GDM (4, 5, 16-18). Specifically, elevated concentrations of alanine, numerous fatty acids (e.g., myristic acid, palmitic acid, palmitoleic acid) and lower amounts of glutamate, proline, and phospholipids in blood have been identified as predictors of GDM risk in early pregnancy (i.e., before 16 weeks) (4), with recent evidence demonstrating significant differences in the abundance of these metabolites between ethnic groups (19). Notably, evidence from Born in Bradford (BiB), a prospective multi-ethnic pregnancy and birth cohort , has demonstrated the need for potentially modified GDM assessment criteria for South Asian (SA) women because of increased risks of delivery complication and newborn macrosomia at significantly lower glucose thresholds, compared to white European (WE) women (20). Indeed, currently the UK's National Health System (NHS), routinely screens all women of South Asian ancestry for GDM while only high-risk WE women are screened (21).

As a consequence of this, the Diabetic Pregnancy Study Group called for increased research into the role of the metabolome on GDM in 2018 (22). To date, however, the metabolic drivers of GDM remain unclear with numerous discrepancies within the field, likely due to small, heterogenous cohorts of heterogenous cohorts of varying populations, cultures, and ancestral groups (23). Indeed, only one study has conducted an analysis of individual metabolites and GDM in an ethnic-specific fashion (1). This work investigated univariate associations between numerous metabolites in WE (n = 4072) and SA (n = 4702) women and demonstrated that concentrations of lipoproteins and cholesterols are typically higher in WE women and are stronger predictors of GDM (ie. have a higher VIP score), compared to SA women. However, metabolite profiles are heterogenous mixtures of metabolites, many of which are strongly correlated and may depend on other metabolites to exhibit an effect. In light of this, multivariate approaches that assess all variables

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

simultaneously along with their inter-variable correlations (24) can be used to identify (i) patterns of uncorrelated metabolites that associate with GDM risk, and (ii) cardinal metabolites that independently associate with GDM risk. Therefore, this study aims to build upon existing work by applying multivariate statistical techniques within an ethnically diverse population to (i) determine underlining metabolite patterns that correlate with GDM, (ii) identify ethnic-specific metabolic drivers of GDM risk.

Materials and Methods

i. Population Characteristics

The BiB cohort was established to examine determinants of health from pregnancy and childhood into adulthood in an ethnically diverse region in the north of England (25). Between 2007 and 2010, BiB recruited 12,453 women (26-28 weeks' gestation, mean maternal age 27.8), collecting baseline data on 13,776 pregnancies and 13,858 births, with 45% of the cohort of SA origin (25, 26). BiB aimed to recruit all mothers giving birth at the Bradford Royal Infirmary, the largest hospital within Bradford. Bradford is a northern English city with high levels of deprivation and a large SA population, the majority of which have Pakistani ancestry. All women were invited to partake in an oral glucose tolerance test (OGTT) for GDM diagnosis at approximately 26-28 weeks during their standard antennal care. Almost all UK citizens utilize the NHS for antenatal care.

Of these, 11,480 women provided blood samples for metabolite analyses during the same visit as their OGTT. Written consent was gained from all participants and ethical approval was granted by the Bradford Research Ethnics Committee (ref07/H1302/112)(25).

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

ii. Blood Metabolite Analysis

Full details of venous blood sample collection, preparation, metabolite quantification and validation have previously been described in detail (1). In brief, fasted blood samples were taken at the Bradford Royal Infirmary by trained phlebotomists, processed within 2.5 hours and stored at -80℃ in the absence of freeze-thaw cycles. (27) Samples were processed using a high-throughput automated NMR platform and have previously been validated (Nightingale Health©; Helsinki, Finland). Metabolite values expressed as a percentage or ratio were excluded to minimize redundancy, resulting in a panel of 146 metabolite values expressed in absolute quantitative measures. This panel comprised measures of 97 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 (**Supplemental S1**).

iii. Participant Selection

Of the 11,480 blood samples analyzed for metabolites, 54 samples were excluded because they failed one of five Nightingale© quality control measures (low glucose, high lactate, high pyruvate, low protein concentration and plasma samples). Of the 11,426 available samples, ~3% of mothers were missing ≥1 metabolite values. The structure of missing metabolite data was assessed via the visualization and imputation of missing values (VIM) package within R (28) and multiple correspondence analysis (MCA). There was no evidence that the missing metabolite data occurred in a non-random pattern. It was therefore deemed appropriate to impute missing values. Optimized multiple imputation with iterative principal component analysis (PCA; 100 simulations, K-fold cross validation) based upon the

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

minimization of mean square error of prediction (MSEP) was performed using the missMDA package (29). A sensitivity analysis was performed to test the effect of mothers with higher rates of missingness (≥3% missing metabolite values) on imputation. No detectable difference in imputation quality was noticed. As such, the metabolite data of all available 11,426 maternal samples were included for imputation.

Imputed metabolite data were combined with descriptive BiB reported characteristics, including participant's ethnicity, age moved to UK (if born abroad) GDM status, gestational age at sample collection (obtained from obstetric records), history of diabetes, age, BMI, smoking status, parity and whether they were carrying a singleton/multiple pregnancy. Length of residence was calculated by subtracting the age the mother moved to the UK from maternal age. When an individual was born within the UK, length of residency was taken to be mothers age.

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 (1, 25). Using this criteria, a women was diagnosed with GDM if either their fasting glucose concentration exceed ≥ 6.1 mmol/L or if 2-hour post-load glucose concentrations was ≥ 7.8 mmol/L following a 75g oral glucose tolerance test (OGTT). The OGTT was completed in the morning following an overnight fast and involved the consumption of a standard solution over a 5-minute period containing the equivalent of 75g of anhydrous glucose (30). Following a GDM diagnosis all SA and WE mothers receive the same standardised care following a GDM diagnosis. Initially GDM management involves referral to a dietitian and the management of glucose

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

concentration through diet and increased exercise. If unsuccessful, managed by metformin or insulin injections will be prescribed. Women with GDM will also be offered additional antenatal appointments to monitor the health of both mother and baby throughout the pregnancy. Irrespective of GDM status, basic nutritional counselling is offered to all mothers as part of standard antenatal classes offered throughout pregnancy by the UK National Health Services (NHS) (31).

Ethnicity was self-reported. If ethnicity was not collected, details were obtained from primary care records along with information on parity and the number of registered births. Maternal age was recorded at pregnancy booking (ie., the first routine antenatal visit) 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 (underweight: ≤ 18.5 kg/m^2 in WE and SAs, normal/healthy weight: 18.6-25 kg/m^2 in WEs or 18.6-22.9 kg/m^2 in SAs, 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) (32). When analysed as a binary variable, women were grouped as having a ‘healthy’ or ‘high’ BMI if they were above/below the BMI cut-off for overweight status utilizing these ethnic-specific cut offs. Smoking status was self-reported at baseline and during pregnancy. Recruitment and the baseline assessment of covariates was the same in both ethnic groups. Summary statistics for each variable were presented as a mean and standard error (SE). Difference in baseline characteristics were calculated between women with and without GDM for continuous variables via a Mann-Whitney (MW) test, while differences for categorical variables were tested using Pearson’s Chi-squared test.

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Participants whose samples were collected after GDM diagnosis (28th week or later) were excluded from the analysis as well as mothers with a history of diabetes. Individuals who reported being of a South Asian origin other than Pakistani (SA) were also excluded, due to the small sample size (therefore limited power) of other South Asian ancestry groups. In total, 5,339 participants, 2,671 SAs (all of Pakistani descent) and 2,688 White European (WE) women, were retained for analysis. **(Figure 1)**

Ethnicity was self-reported and the homogeneity of the WE group has been confirmed in previous genetic analyses within BIB (33). 93.2% of the included WEs were born in the British Isles (i.e., the UK, Republic of Ireland, Channel Islands or Isle of Man), with the majority in England (91.4%). Of these women, 95.5% reported that both of their parents were also born in the British Isles. Within the group of WE women not born in the British Isles, 3.7% were born in Eastern Europe (Czech Republic, Poland, Slovakia) with the remaining proportion reporting 'other' or 'unknown'. Within the SA population, 43.7% were born within the UK. Of the SA women born in the UK, 93% reported that their mother was born in Pakistan (87.4%) or India (5.6%) and 95% reported that their father was born in Pakistani (88.6%) or India (6.7%). A small proportion did not know their mother's (1.4%) or father's (1.3%) place of birth. Of the women born outside of the UK, the average age of immigration to the UK was 18.8 yrs (IQR 18 – 23).

iv. Metabolite Discriminatory Analysis

Partial least squares discriminatory analysis (PLSDA) is a supervised dimensionality reduction technique that uses all included variables to discriminate group data based

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

upon predefined outcome groups. 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 projection (VIPs), where VIPs ≥1 denote a variable with good discriminatory quality and predictive ability (34, 35).

PLSDA allowed an overall assessment of the predictive capacity of metabolites for GDM, in models with and without known GDM risk factors (i.e., BMI, maternal age, parity, multiple pregnancy, and smoking status), with ethnicity added to visually assess its effect on the model. Following this, both sets of PLSDA models were performed within each ethnic group. To assess bi-directionality, models predicting ethnicity were also executed within the overall population and GDM cases/ non-cases separately using the same criteria as above.

The optimum number of components to include within the model was selected based upon the component's ability to significantly predict group membership within the training ($pR^2Y \leq 0.05$) and validation ($pQ^2Y \leq 0.05$) datasets (7-fold cross validation, 'nipals' algorithm). When multiple components were significantly predictive, the predictive component that best discriminated between groups (i.e. maximization of outcome variance explained, R^2Y) with the minimal error (root mean squared error of estimation (RMSEE)) was selected. Data were pareto scaled and mean-centered prior to analysis. External validity was assessed via 7-fold cross validation. PLSDA models were performed via the 'ropls' package within R (36). When the size of the outcome groups differed by ≥ 1% the larger group was randomly sampled (n=20) to minimize error. VIPs were mean averaged and SEs calculated across all significant iterations ($pR^2Y \leq 0.05$, $pQ^2Y \leq 0.05$) for each metabolite following the removal of outlier VIPs, defined as 1.5 x interquartile range of VIP values. Differences in the distribution of VIP values between both ethnicities

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

and case-status were assessed for significant iterations via a MW test; this was possible because all comparisons were tested against the same panel of metabolite measures. To assess the impact of smoking on PLSDA results, PLSDA models predicting smoking in the overall study population were also performed.

iii. Post-Hoc multivariate analyses

BMI is a suspected mediator along the casual pathway that links metabolism and GDM, was a significant driver of GDM within SA women and WE women. To explore this, the ethnic-specific impact of BMI on the metabolome and subsequent GDM diagnoses was investigated using sparse PLSDA (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 and balancing variable selection with group classification (37). 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), which is the same approach used by the NHS (38). The analyses focussed on low-risk WE (n= 872) and low-risk SA women (n= 864) — i.e., only women (i) in their first pregnancy, (ii) that did not smoke during pregnancy, and (iii) were < 35 years of age were included. This was done to prevent these covariates from overpowering the models, and allowing the contributing roles of BMI on GDM to be more clearly appreciated within and between each ethnic group.

Metabolites selected by sPLSDA in each comparison were fed into PLSDA models (20 iterations) alongside highly correlated metabolites (Pearsons correlation coefficient ≥0.9) in order determine metabolie values contributing to the separation of

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

the outcome groups whilst balancing dimensionality reduction and group discrimination. PLSDA models were adjusted for maternal age (continuous), BMI (continous), smoking status, partiy and multiple pregnancy such as before. Differences in the distributions of metabolites within each group were also compared by a MW test.

iii. Linear regression analyses for identified metabolite associations

Linear regression models investigating the relationship between post- oral OGTT measures (fasting glucose and 2-hour post-OGTT) were performed on all metabolite values identified as important (VIP≥1) in characterizing GDM status. Normality of glucose measures and metabolite values were assessed using histograms and Q-Q plots. Most metabolites (136/146) required normalization. Normality was most often achieved by log transformation (59 metabolite values); however, in some cases square-root and normal score transformation (NST) were implemented via the 'rcompanion' package(39). All glucose measures were log normalized. Known GDM risk factors of maternal age (years), gestational age (days), parity, and smoking status during pregnancy (yes/no), were initially including in the models. When significant associations were observed between metabolite values and glucose in this exploratory analysis ($P<0.05$), BMI was added to the models (initially as a continuous and then as a binary variable utilizing ethnic-specific BMI cut-offs for overweight status) to assess the role of early pregnancy BMI as a mediator of metabolite-dysglycemia associations. Within SAs, a final additional adjustment of length of residency within the UK was made to account for any effects of acculturation.

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Results

i. Population Characteristics

The mean age of participant was 26.7 years and had a mean BMI of 26 kg/m^2. WE women were significantly older and had higher BMIs compared to SA women (**Table 1**). Parity was significantly higher in SA women compared to WE women (P<0.001) and parity was only significantly higher in GDM cases compared to non-cases within SA women (**Table 1**). Smoking during pregnancy was significantly more common in WE women compared to SA women (25% vs 3%; P<0.001). No difference in 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 time of sample collection was 187 gestational days.

ii. Primary Analysis

Metabolite Characterisation of GDM

In the 1st model, an overall analysis of the full cohort (i.e., both ethnic groups), PLSDA explained 21.7% of the variation between the GDM and non-GDM groups and confirmed maternal age and BMI as primary risk factors for GDM risk followed by parity, smoking status, and having a non-singleton pregnancy as the primary drivers of GDM (**Table 2**). In the full model, 7 metabolite values reported VIPs ≥ 1, including 4 fatty acid metabolite measures (total fatty acids, 18:2 linoleic acid, total MUFA and total SFA), and one glycolysis related metabolite (lactate) (**Figure 2**). Modelled independently, the PLSDA with only covariates explained 12.4% of the variation in GDM status and significantly predicted GDM status, whereas the model with only metabolites explained 13.5% of the variance in GDM but was non-significant. The 2nd model, which included ethnicity as a covariate, accounted for

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

26.6% of the variation between the GDM and non-GDM groups. The same 6 metabolites we reported as predictors of GDM with an additional cholesterol metabolite measure (total esterified cholesterol). Notably, model 2 confirmed ethnicity (SA vs WE) as a major risk factor for GDM, after age and BMI. Modelled independently, 'ethnicity' and other covariates explained 15.2% of the variance in GDM status; therefore, the addition of metabolites into the model increased the amount of variance explained by over 11%.

Ethnically-stratified analysis of metabolites characterising GDM

In an ethnically stratified analysis (20 iterations), models only including metabolites accounted for a median average of 6.5% of the variation in GDM status in SA women and 5.8% of the variation in WE women in optimised models (ie. minimisation RMSEE and maximisation of R^2Y) although no model comprising metabolites alone was signficant. Conversely, models only including established clinical risk factors (age, BMI, parity, smoking status and multiple pregnancy) were significantly predictive (p- value $R^2 < 0.05$, $Q^2<0.05$) of GDM status and explained 13.3% of the variation in SAs 12.8% of the variation in WEs. The addition of metabolites to these covariate models also resulted in the significant prediction of GDM. These models resulted in 26% of the variance in GDM status in WE women and 20% of the variance in SA women being accounted for, an increase of 13.6% and 6.8% when compared to covariate models in WE and SA respectively. Following adjustments for maternal age, parity, BMI, and smoking status, GDM could be predicted within both ethnicites. Maternal age, parity and BMI were predictors of GDM in both ethnicities (VIP≥1), with BMI the most important predictor of GDM in SA women , while in WE women maternal age was most important predictor

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

(**Supplementary table 1**). Smoking was a predictor of GDM only in WE women. After adjustment for confounders, 7 metabolite variables characterised GDM status (VIP≥1) in both ethnicities (total fatty acids, total MUFA, total SFA, linoleic acid, glycoprotein acetyls, lactate, and diameter of VLDL) (**Figure 3**, **Supplementary table 2**). Of these metabolites, the VIPs of three (lactate, glycoprotein acetyls, and linoleic acid) characterised GDM status comparatively well between ethnicites (VIP≥1; MW P>0.05), whereas four metabolite measures (total fatty acids, total MUFA, total SFA, and VLDL_D) characterised GDM in both ethnic groups, but were significantly stronger markers of GDM in WE women (VIP≥1; MW P<0.05 between ethnicites). 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. On average, the optimised models explained 26% of the variance of GDM in WE women, and 20% of the variance in SA, women (**Supplementary table 3**).

Metabolites characterised by ethnicity

To explore underlying metabolic profiles within each ethnic group, we identified metabolites that most strongly distinguished WE women and SA women. In a PLSDA including known GDM risk factors as covariates (maternal age, smoking status, parity, BMI, and GDM status), 12 metabolic measures VIP≥1 in statistically significant models (models p- values $R^2 > 0.05$ and $Q^2>0.05$) and therefore were believed to have characterised ethnicity in GDM and non-GDM women: total fatty acids, serum cholesterol, SFA, MUFA, FAw6, esterified cholesterol, LA, LDL cholesterol, remnant cholesterol, phosphatidylcholine and total cholesterol. (**Supplementary table 4**).

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Additionally, ethnicity was characterised by 6 metabolites values exclusivley in women diagnosed with GDM [i.e alanine, total fatty acids, linoleic acid (LA), glycoprotein acetyls, lactate and diameter of VLDL] whilst 5 metabolites values were exclusive in those not diagnosed with GDM [i.e., apolipoprotein A1, remnant choelsterol, docosahexanoic acid (DHA), and phosphatidylcholine]. An additional 9 metabolite values [ie., total serum cholesterol, LDL cholesterol, total esterfied cholesterol, n-3 fatty acids, PUFA, MUFA, SFA, phosphatidylcholine and total cholines] were predictive of ethnicity in both GDM cases and non-cases (**Supplementary figure 1**).

iv. Post-Hoc Analyses

Characterisation of GDM in Low-Risk Women

BMI was classifed as an important variable (VIP ≥ 1) in the overall analysis and in both ethnic subgroup analyses. However, a greater mean VIP (± SE) was observed in SA women compared to WE women (VIP_{SA} = 7.06 ± 0.22 vs. VIP_{WE} = 4.33 ± 0.22; P<0.001) (**Supplementary table 1**) indicating that BMI may be a more important predictor of GDM status within WEs. Indeed, healthy weight SA women who developed GDM ($SA_{Healthy\text{-}GDM}$) presented the most distinct metabolic profile (Receiver Operator Curve; ROC = 0.783), but were most similar to healthy WE women who developed GDM ($WE_{Healthy\text{-}GDM}$; ROC = 0.691) (**Supplementary figure 2**). The reason for this shared and distinct pattern of metabolites in ‘healthy’ weight women who developed GDM is unclear and many hypotheses are possible. One hypothesis may be that the pattern is an artifact of their fetal programming. Adult offspring from GDM pregnancies are at increased risk of dysglycemia, diabetes, and

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

GDM that has been attributed to metabolic dysregulation, and early dysglycemia that progresses in later life (40-43).

Future work in established cohorts that investigate trans-generational pregnancy risks (such as Born in Bradford, Generation R, and Nutrigen) are integral to unravel the source of this unique metabolic profile which distinguishes healthy weight GDM cases of SA ancestry from non-cases, overweight SA cases, and WE cases (25, 44, 45). Due to the higher proportion of underweight mothers of SA ancestry, a sensitivity analysis was performed where underweight mothers were removed ($n_{removed}$ = 93, BMI≤18.5 kg/m^2) to determine if their profiles were unique. No difference in the outcome was observed following the removal of these individuals.

Metabolites selected by sPLSDA in each comparison were fed into PLSDA models (20 iterations) alongside highly correlated metabolites (Pearsons correlation coefficient ≥0.9) to identify key metabolic drivers of this separation (**Supplementary figure 3**). Alanine, glutamine, and glycerol were important to distinguish healthy weight SA women who developed GDM (SAC-N) from all others, whereas fatty acids were important to distinguish SAC-N from other GDM cases. Interestingly, in healthy women, aromatic and branched chain amino acids distinguished GDM and non-GDM women between (but not within) ethnic groups. Glycerol distributions were significantly different in all comparisons (MW <0.05).

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

Orthogonal partial least squares discriminant analysis (oPLSDA) is a supervised multivariate technique that separates variation within each predictor variables based upon its linear (correlated) and orthogonal (uncorrelated) association with the outcome variable (46, 47). This can provide better separation along fewer components when a large proportion of variance within the dataset does not directly

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

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 separating the main group of interest (healthy weight SA cases, SAC-N) with the two most relevant biological comparisons (healthy weight SA non-cases, SANC-N and healthy weight WE cases, WEC-N).

No significant separation of the SAC-N vs SANC-N, SAC-N vs SAC-H and SAC-N vs WEC-N groups were identified via SUS plots with oPLSDA. Following the inclusion of BMI and age within the models the SAC-N group was found to separate from all other groups (**Supplementary figure 4**). BMI was the only variable found to be responsible for this separation with a high magnitude and reliability. Pyruvate, L-HDL and XL-HDL had a small impact on the separation of the SAC-N group but with a low reliability, as shown within SUS-plots.

Association between important metabolites and gestational dysglycemia

Overall, 8 of 146 metabolite measures were associated with fasting glucose or 2-hour post glucose (**Table 3**), all of which were identified as GDM predictors via PLSDA or sPLSDA. The analysis in WE women demonstrated the greatest number of associations between metabolite and glucose measures. Six metabolites positively associated fasting glucose concentration (albumin, lactate, histidine, apolipoprotein A1, HDLC and HDL2C), while one negatively associated with fasting glucose (mean density of LDL) (**Supplementary table 5, Supplementary table 6)**. Only DHA associated with 2-hr post OGTT in WEs, where a 1 mmol/L increase in DHA associated with a 0.20 mmol/L increase in 2-hour post glucose. In the analysis of SA women, only albumin was associated with dysglycemia, where higher albumin associated with lower concentration of fasting glucose and 2-hr post OGTT. In an additional analysis, length of residency within the UK was added to the fully-adjusted

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

model to evaluate the role of UK acculturation as a modifier of the association between albumin and postprandial glucose measures. In both models significant associations were identified with albumin (P-value$_{fasting}$ =0.031, β= -0.79, SE=0.37, P-value$_{2\text{-hour}}$ =0.028 β=-1.75 ,SE=0.80). Length of residency was not found to be a significant variable in the model, but the magnitude of associations decreased slightly following its inclusion (**Supplementary table 5, Supplementary table 6**). In the ethnic-combined analysis, associations between albumin, lactate, and mean diameter of LDL with fasting glucose retained significance. Adjusting for BMI as a continuous or binary variable had no impact on the associations.

Discussion

Using a prospective birth cohort with an equal proportion of WE and SA women, we identified 7 metabolite measures that characterized GDM in both WE and SA women — 4 of which were more predictive in WE women. These results agree with the Omega cohort (78.5% non-Hispanic white; nested case-control; 46 cases, 47 controls) that highlighted a distinct metabolic profile at 16-weeks' gestation (comprised of fatty acids, sugars, alcohols, amino acids and organic acids), associated with future GDM diagnosis (48). Although the metabolite patterns identified by the Omega study were not predictive, our predictive multivariate analysis (and a previous univariate analyses) (1) found similar associations between GDM and many of these metabolites (i.e., amino acids, glycolysis related metabolites, and fatty acids), and offers further evidence of ethnic-specific associations.

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Given the overall elevated risk of GDM observed in SA women compared to WE women, even at a healthy BMI (i.e., OR≈3) (49), and the role of ethnicity in predicting GDM, in the present study, we sought to characterize distinct metabolic profiles of SA and WE women. Of the 146 metabolite values tested, 7 were important for stratifying GDM and non-GDM women in the overall population (lactate, mean density of VLDL particles, total fatty acids, total MUFAs, 18:2 linoleic acid, total SFA and esterified cholesterol). Following stratification by ethnicity, alanine, glutamine, total serum cholesterol, n-6 fatty acids, PUFAs, and citrate distinguished GDM and non-GDM in WE women whereas no metabolite values were predictive solely within SA women.

Although no metabolite value identified solely within WE women was associated with post-OGTT measures of glucose in *post-hoc* analyses, our evidence agrees with previous work from (i) a small case-control study (26 T2Ds vs 7 controls) that reported alanine, glutamine, and citrate to characterize GDM and controls, with citrate being a key marker of diabetics with underlying complications (e.g., CVD) (50), and (ii) a cohort study of 431 pregnant Chinese women (12-16 weeks' gestation), where alanine and glutamine were associated with GDM (51). Biologically, alanine, glutamine, and citrate are 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 (52, 53). Taken together, it may be that alanine and glutamine are more robust markers of dysglycemia, whereas citrate is a marker of metabolic or physiologic stress in diabetic individuals — such as pregnancy. The role of total cholesterol is uncertain as it is not convincingly associated with dysglycemia (a meta-analysis of 73 observational studies found no association)(54), suggesting

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

that associations between total cholesterol and GDM are complex and/or subject to confounding.

In the ethnic subgroup analyses, fatty acids were identified as the most important family (ie. VIP ≥ 1) to characterize GDM status. In WE women and SA women respectively, 75% and 50% of the fatty acids included within the metabolite panel were considered 'important' to characterize GDM within WE women. Furthermore, in SA women fatty acids constituted more than half of all metabolites with a VIP ≥ 1.

This reflects earlier work by Taylor *et al*. [1], which 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 (55, 56).

Furthermore, fatty acids (total MUFAs, total n-3 PUFAs, total n-6 PUFAs, total PUFAs, and DHA) were identified as key metabolic factors to distinguish healthy-weight SA and WE women who developed GDM. Interestingly, we highlighted associations between n-6 PUFA and total PUFAs with GDM that were specific to WE women. Given the equal sample sizes between groups, and that fatty acids were important to characterize ethnicity, it is suggestive of ethnic differences in PUFA metabolism (57-59) and a role in ethnic-associated GDM risk (57, 60, 61). Indeed, n6 PUFA-derived eicosanoids show discriminatory qualities between type-2 diabetics and controls with good accuracy ($R^2X = 0.824$, $R^2Y = 0.995$, $Q^2 = 0.779$) and were identified as proposed mediators of dysglycemia within a Chinese population (62). Longitudinal analyses to evaluate the association between changes in PUFA and

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

eicosanoids concentrations on dysglycemia during pregnancy are required to better understand this association.

The association between VLDL diameter and dysglycemia is supported by a recent hypothesis linking insulin resistance, triglyceride synthesis, and increased VLDL diameter (63, 64). Although we cannot disregard that VLDLs are sensitive to level of fasting (65) (*as our participants were subjected to prior to blood collection*), evidence also suggests that ethnic-specific genetic variants associate with ethnic-specific differences in VLDL diameter (66). Although there has been less work on the possible association between glycoprotein acetyls (a marker of systemic inflammation) and GDM, and future work is required in this area

Lactate was one of the strongest predictors of GDM within both groups, in agreement with evidence from a case-control study in China (n=12 GDM; n=10 controls) (67) and pathway analyses that propose lactate as a regulator of insulin resistance and a marker metabolic syndrome severity (68, 69). *Post-hoc* analysis demonstrated no association between glycoprotein acetyls and glucose concerntrations, while lactate and mean diameter of VLDL were associated with fasting glucose in WE women but not SA women. The multi-ethnic HAPO cohort demonstrated a similar ethnic-specific association between lactate and fasting glucose within individuals of Northern European ancestry but not minority ethnic groups (48, 70, 71).

Of the numerous fatty acid measures that were associated with GDM, only DHA was associated with a post-OGTT measures of glucose and only in WE women. Overall, DHA is considered a protective metabolite against insulin resistance (e.g., HOMA-IR); however, recent evidence suggests high heterogeneity (56, 72, 73).

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

As we did, 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<0.05) (56), while conversely, the DOMINO trial (n=1990 pregnant women) reported no difference in 1-hr post-OGTT glucose concentrations between DHA supplemented mothers and controls (74). The reason for such discrepancies is unclear but may be that n-3 PUFAs (such as DHA) require interactions with other metabolites (e.g., Vitamin D) (75) to impart an effect, concentrations of which vary considerably between populations, seasons, and geographic region (76-78).

The study aimed to increase and test generalizability of results within a diverse population; however, our results may not be generalizable across other ethnic groups or geographic regions. Nonetheless, this study has four main limitations. Firstly, samples were 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 generalizable across the full-term of pregnancy. Secondly, as with all observational studies, the effect of confounding cannot be disregarded and causality cannot be inferred. Despite this, this is the first study to use a panel of multivariate statistical techniques to characterize GDM within a large prospective cohort with an equal representation of WE women and women from a non-WE population, meaning that statistical power to measure the same effect size is comparable between groups. Thirdly, the biological validity of the identified metabolites was tested and many correlated with postprandial glucose measures; and although confounding cannot be eliminated, all models included known GDM confounders and modelling characterizing the overall metabolic differences between ethnicities were also performed to test whether differences in

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

metabolite profiles were found between ethnicities in relation to GDM status. Finally, diet is a contributor to metabolite concentrations, but comprehensive dietary data was not available for our analysis. Future work with comprehensive dietary records are needed evaluate the presence of a moderating effect of diet on metabolism and GDM risk.

Conclusion

In conclusion, this study has identified unique and shared metabolic profiles that characterize GDM in WE and SA women. Future work exploring the moderating role of lifestyle on the metabolome and the underlying biological mechanisms driving the identified associations will provide a better understanding of the etiological factors responsible for the heightened level of GDM risk experienced by SA women and shed light on improved prevention strategies

Acknowledgements

Born in Bradford (BiB) is only possible because of the enthusiasm and commitment of the children and parents in BiB. We are grateful to all the participants, health professionals, schools and researchers who have made Born in Bradford happen

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

References

1. Taylor K, Ferreira DLS, West J, Yang T, Caputo M, Lawlor DA. Differences in Pregnancy Metabolic Profiles and Their Determinants between White European and South Asian Women: Findings from the Born in Bradford Cohort. Metabolites. 2019;9(9).
2. Zeng Z, Liu F, Li S. Metabolic Adaptations in Pregnancy: A Review. Ann Nutr Metab. 2017;70(1):59-65.
3. Egan AM, Dow ML, Vella A. A Review of the Pathophysiology and Management of Diabetes in Pregnancy. Mayo Clin Proc. 2020;95(12):2734-46.
4. Chen Q, Francis E, Hu G, Chen L. Metabolomic profiling of women with gestational diabetes mellitus and their offspring: Review of metabolomics studies. J Diabetes Complications. 2018;32(5):512-23.
5. McIntyre HD, Catalano P, Zhang C, Desoye G, Mathiesen ER, Damm P. Gestational diabetes mellitus. Nat Rev Dis Primers. 2019;5(1):47.
6. Behboudi-Gandevani S, Amiri M, Bidhendi Yarandi R, Ramezani Tehrani F. The impact of diagnostic criteria for gestational diabetes on its prevalence: a systematic review and meta-analysis. Diabetol Metab Syndr. 2019;11:11.
7. Zhu Y, Zhang C. Prevalence of Gestational Diabetes and Risk of Progression to Type 2 Diabetes: a Global Perspective. Curr Diab Rep. 2016;16(1):7.
8. Saravanan P, Magee LA, Banerjee A, Coleman MA, Von Dadelszen P, Denison F, Farmer A, Finer S, Fox-Rushby J, Holt R, et al. Gestational diabetes: opportunities for improving maternal and child health. The Lancet Diabetes & Endocrinology. 2020;8(9):793-800.
9. Mijatovic-Vukas J, Capling L, Cheng S, Stamatakis E, Louie J, Cheung NW, Markovic T, Ross G, Senior A, Brand-Miller JC, et al. Associations of Diet and Physical Activity with Risk for Gestational Diabetes Mellitus: A Systematic Review and Meta-Analysis. Nutrients. 2018;10(6):698.
10. American Diabetes A. 14. Management of Diabetes in Pregnancy: <em>Standards of Medical Care in Diabetes—2020</em&gt. Diabetes Care. 2020;43(Supplement 1):S183.
11. NICE. Diabetes in pregnancy: management from preconception to the postnatal period 2021 [updated 16th December 2020. Available from: https://www.nice.org.uk/guidance/ng3/chapter/Recommendations#preconception-planning-and-care.
12. de Seymour J, Chia A, Colega M, Jones B, McKenzie E, Shirong C, Godfrey K, Kwek K, Saw S-M, Conlon C, et al. Maternal Dietary Patterns and Gestational Diabetes Mellitus in a Multi-Ethnic Asian Cohort: The GUSTO Study. Nutrients. 2016;8(9):574.
13. Zulyniak MA, Souza RJD, Shaikh M, Desai D, Lefebvre DL, Gupta M, Wilson J, Wahi G, Subbarao P, Becker AB, et al. 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. 2017;7.
14. Fuller H, Moore JB, Iles MM, Zulyniak MA. Ethnic-specific associations between dietary consumption and gestational diabetes mellitus incidence: a meta-analysis. PLOS Global Public Health. 2022.

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

15. Zulyniak MA, de Souza RJ, Shaikh M, Ramasundarahettige C, Tam K, William N, Desai D, Lefebvre DL, Gupta M, Subbarao P. Ethnic differences in maternal diet in pregnancy and infant eczema. PLoS One. 2020;15(5):e0232170.
16. Wang Q, Wurtz P, Auro K, Makinen VP, Kangas AJ, Soininen P, Tiainen M, Tynkkynen T, Jokelainen J, Santalahti K, et al. Metabolic profiling of pregnancy: cross-sectional and longitudinal evidence. BMC Med. 2016;14(1):205.
17. Law KP, Zhang H. The pathogenesis and pathophysiology of gestational diabetes mellitus: Deductions from a three-part longitudinal metabolomics study in China. Clin Chim Acta. 2017;468:60-70.
18. McCabe CF, Perng W. Metabolomics of Diabetes in Pregnancy. Curr Diab Rep. 2017;17(8):57.
19. Mitro SD, Wu J, Rahman ML, Cao Y, Zhu Y, Chen Z, Chen L, Li M, Hinkle SN, Bremer AA, et al. Longitudinal Plasma Metabolomics Profile in Pregnancy—A Study in an Ethnically Diverse U.S. Pregnancy Cohort. Nutrients. 2021;13(9):3080.
20. Farrar D, Fairley L, Santorelli G, Tuffnell D, Sheldon TA, Wright J, van Overveld L, Lawlor DA. Association between hyperglycaemia and adverse perinatal outcomes in south Asian and white British women: analysis of data from the Born in Bradford cohort. The Lancet Diabetes & Endocrinology. 2015;3(10):795-804.
21. NHS. Overview Gestational Diabetes www.NHS.co.uk: NHS; 2019 [updated 06/08/2019. Available from: https://www.nhs.uk/conditions/gestational-diabetes/.
22. Schaefer-Graf U, Napoli A, Nolan CJ, Diabetic Pregnancy Study G. Diabetes in pregnancy: a new decade of challenges ahead. Diabetologia. 2018;61(5):1012-21.
23. Mao X, Chen X, Chen C, Zhang H, Law KP. Metabolomics in gestational diabetes. Clin Chim Acta. 2017;475:116-27.
24. Saccenti E, Hoefsloot HCJ, Smilde AK, Westerhuis JA, Hendriks MMWB. Reflections on univariate and multivariate analysis of metabolomics data. Metabolomics. 2013;10(3):361-74.
25. Wright J, Small N, Raynor P, Tuffnell D, Bhopal R, Cameron N, Fairley L, Lawlor DA, Parslow R, Petherick ES, et al. Cohort Profile: the Born in Bradford multi-ethnic family cohort study. Int J Epidemiol. 2013;42(4):978-91.
26. Bird PK, McEachan RRC, Mon-Williams M, Small N, West J, Whincup P, Wright J, Andrews E, Barber SE, Hill LJB, et al. Growing up in Bradford: protocol for the age 7-11 follow up of the Born in Bradford birth cohort. BMC Public Health. 2019;19(1):939-.
27. Taylor K, McBride N, J Goulding N, Burrows K, Mason D, Pembrey L, Yang T, Azad R, Wright J, A Lawlor D. Metabolomics datasets in the Born in Bradford cohort [version 2; peer review: 1 approved, 1 approved with reservations]. Wellcome Open Research. 2021;5(264).
28. Alexander Kowarik MT. Imputation with the R Package VIM. Journal of Statistical Software. 2016;74(7):1-16.
29. Josse J, Husson F. missMDA: A Package for Handling Missing Values in Multivariate Data Analysis. Journal of Statistical Software. 2016;70(1).
30. Lawlor DA, West J, Fairley L, Nelson SM, Bhopal RS, Tuffnell D, Freeman DJ, Wright J, Whitelaw DC, Sattar N. Pregnancy glycaemia and cord-blood levels of insulin and leptin in Pakistani and white British mother–offspring pairs: findings from a prospective pregnancy cohort. Diabetologia. 2014;57(12):2492-500.
31. NHS. Antental Classes www.NHS.co.uk: NHS; 2021 [updated 17/03/2021. Available from: https://www.nhs.uk/pregnancy/labour-and-birth/preparing-for-the-birth/antenatal-classes/.

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

32. WHO. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. The Lancet. 2004;363(9403):157-63.
33. Arciero E, Dogra SA, Malawsky DS, Mezzavilla M, Tsismentzoglou T, Huang QQ, Hunt KA, Mason D, Sharif SM, van Heel DA, et al. Fine-scale population structure and demographic history of British Pakistanis. Nature Communications. 2021;12(1):7189.
34. Perreault M, Zulyniak MA, Badoud F, Stephenson S, Badawi A, Buchholz A, Mutch DM. A distinct fatty acid profile underlies the reduced inflammatory state of metabolically healthy obese individuals. PLoS One. 2014;9(2):e88539-e.
35. Badoud F, Lam KP, DiBattista A, Perreault M, Zulyniak MA, Cattrysse B, Stephenson S, Britz-McKibbin P, Mutch DM. Serum and Adipose Tissue Amino Acid Homeostasis in the Metabolically Healthy Obese. Journal of Proteome Research. 2014;13(7):3455-66.
36. Thevenot EA, Roux A, Xu Y, Ezan E, Junot C. 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. 2015;14(8):3322-35.
37. Lê Cao K-A, Boitard S, Besse P. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. BMC Bioinformatics. 2011;12(1):253.
38. NHS. BMI Healthy Weight Calculator: NHS; 2018 [updated 5/11/2018. Available from: https://www.nhs.uk/live-well/healthy-weight/bmi-calculator/.
39. Mangiafico S. rcompanion: Functions to Support Extension Education Program Evaluation. R package version 2.4.1. ed2021.
40. Egeland G, Meltzer SJ. Following in mother's footsteps? Mother–daughter risks for insulin resistance and cardiovascular disease 15 years after gestational diabetes. Diabetic Medicine. 2010;27(3):257-65.
41. Clausen TD, Mathiesen ER, Hansen T, Pedersen O, Jensen DM, Lauenborg J, Schmidt L, Damm P. Overweight and the metabolic syndrome in adult offspring of women with diet-treated gestational diabetes mellitus or type 1 diabetes. The Journal of Clinical Endocrinology & Metabolism. 2009;94(7):2464-70.
42. McLean M, Chipps D, Cheung NW. Mother to child transmission of diabetes mellitus: does gestational diabetes program Type 2 diabetes in the next generation? Diabetic Medicine. 2006;23(11):1213-5.
43. Anand SS, Gupta M, Teo KK, Schulze KM, Desai D, Abdalla N, Zulyniak M, de Souza R, Wahi G, Shaikh M. Causes and consequences of gestational diabetes in South Asians living in Canada: results from a prospective cohort study. CMAJ open. 2017;5(3):E604.
44. Jaddoe VW, Mackenbach JP, Moll HA, Steegers EA, Tiemeier H, Verhulst FC, Witteman JC, Hofman A. The Generation R Study: Design and cohort profile. Eur J Epidemiol. 2006;21(6):475-84.
45. de Souza RJ, Zulyniak MA, Desai D, Shaikh MR, Campbell NC, Lefebvre DL, Gupta M, Wilson J, Wahi G, Atkinson SA, et al. Harmonization of Food-Frequency Questionnaires and Dietary Pattern Analysis in 4 Ethnically Diverse Birth Cohorts. The Journal of nutrition. 2016;146(11):2343-50.
46. Worley B, Powers R. Multivariate Analysis in Metabolomics. Curr Metabolomics. 2013;1(1):92-107.
47. Blasco H, Blaszczynski J, Billaut JC, Nadal-Desbarats L, Pradat PF, Devos D, Moreau C, Andres CR, Emond P, Corcia P, et al. Comparative analysis of targeted metabolomics:

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

dominance-based rough set approach versus orthogonal partial least square-discriminant analysis. J Biomed Inform. 2015;53:291-9.
48. Enquobahrie DA, Denis M, Tadesse MG, Gelaye B, Ressom HW, Williams MA. Maternal Early Pregnancy Serum Metabolites and Risk of Gestational Diabetes Mellitus. J Clin Endocrinol Metab. 2015;100(11):4348-56.
49. Makgoba M, Savvidou M, Steer P. An analysis of the interrelationship between maternal age, body mass index and racial origin in the development of gestational diabetes mellitus. BJOG: An International Journal of Obstetrics & Gynaecology. 2012;119(3):276-82.
50. Del Coco L, Vergara D, De Matteis S, Mensa E, Sabbatinelli J, Prattichizzo F, Bonfigli AR, Storci G, Bravaccini S, Pirini F, et al. NMR-Based Metabolomic Approach Tracks Potential Serum Biomarkers of Disease Progression in Patients with Type 2 Diabetes Mellitus. J Clin Med. 2019;8(5).
51. Jiang R, Wu S, Fang C, Wang C, Yang Y, Liu C, Hu J, Huang Y. Amino acids levels in early pregnancy predict subsequent gestational diabetes. J Diabetes. 2020;12(7):503-11.
52. Newsholme P, Brennan L, Rubi B, Maechler P. New insights into amino acid metabolism, beta-cell function and diabetes. Clin Sci (Lond). 2005;108(3):185-94.
53. Haber EP, Procopio J, Carvalho CR, Carpinelli AR, Newsholme P, Curi R. New insights into fatty acid modulation of pancreatic beta-cell function. Int Rev Cytol. 2006;248:1-41.
54. Ryckman KK, Spracklen CN, Smith CJ, Robinson JG, Saftlas AF. Maternal lipid levels during pregnancy and gestational diabetes: a systematic review and meta-analysis. BJOG. 2015;122(5):643-51.
55. 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, Martinez-Portilla RJ. 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;2019:7098470-.
56. Chen X, Stein TP, Steer RA, Scholl TO. 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. 2019;7(1):e000632.
57. Gray RG, Kousta E, McCarthy MI, Godsland IF, Venkatesan S, Anyaoku V, Johnston DG. 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. 2013;12:25-.
58. Benedetti S, Al-Tannak NF, Alzharani M, Moir HJ, Stensel DJ, Thackray AE, Naughton DP, Dorak MT, Spendiff O, Hill N. 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. 2019;9(4):71.
59. Ralston JC, Zulyniak MA, Nielsen DE, Clarke S, Badawi A, El-Sohemy A, Ma DW, Mutch DM. Ethnic-and sex-specific associations between plasma fatty acids and markers of insulin resistance in healthy young adults. Nutrition & metabolism. 2013;10(1):42.
60. Yuen L, Wong VW, Simmons D, Simmons D. Ethnic Disparities in Gestational Diabetes. Current Diabetes Report. 2018;18(68).
61. Cheung NW. The influence of ethnicity on the development of type 2 diabetes mellitus in women with gestational diabetes: a prospective study and review of the literature. ISRN endocrinology. 2012;2012.

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

62. Xia F, He C, Ren M, Xu F-G, Wan J-B. 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. 2020;1120:24-35.
63. Zhao H, Li H, Chung ACK, Xiang L, Li X, Zheng Y, Luan H, Zhu L, Liu W, Peng Y, et al. Large-Scale Longitudinal Metabolomics Study Reveals Different Trimester-Specific Alterations of Metabolites in Relation to Gestational Diabetes Mellitus. J Proteome Res. 2019;18(1):292-300.
64. Krauss RM. Lipids and lipoproteins in patients with type 2 diabetes. Diabetes Care. 2004;27(6):1496-504.
65. Nordestgaard BG. A Test in Context: Lipid Profile, Fasting Versus Nonfasting. Journal of the American College of Cardiology. 2017;70(13):1637-46.
66. Frazier-Wood AC, Manichaikul A, Aslibekyan S, Borecki IB, Goff DC, Hopkins PN, Lai C-Q, Ordovas JM, Post WS, Rich SS, et al. Genetic variants associated with VLDL, LDL and HDL particle size differ with race/ethnicity. Human Genetics. 2013;132(4):405-13.
67. Liu T, Li J, Xu F, Wang M, Ding S, Xu H, Dong F. Comprehensive analysis of serum metabolites in gestational diabetes mellitus by UPLC/Q-TOF-MS. Anal Bioanal Chem. 2016;408(4):1125-35.
68. Wu Y, Dong Y, Atefi M, Liu Y, Elshimali Y, Vadgama JV. Lactate, a Neglected Factor for Diabetes and Cancer Interaction. Mediators Inflamm. 2016;2016:6456018.
69. Jones TE, Pories WJ, Houmard JA, Tanner CJ, Zheng D, Zou K, Coen PM, Goodpaster BH, Kraus WE, Dohm GL. Plasma lactate as a marker of metabolic health: Implications of elevated lactate for impairment of aerobic metabolism in the metabolic syndrome. Surgery. 2019;166(5):861-6.
70. Jacob S, Nodzenski M, Reisetter AC, Bain JR, Muehlbauer MJ, Stevens RD, Ilkayeva OR, Lowe LP, Metzger BE, Newgard CB, et al. Targeted Metabolomics Demonstrates Distinct and Overlapping Maternal Metabolites Associated With BMI, Glucose, and Insulin Sensitivity During Pregnancy Across Four Ancestry Groups. Diabetes care. 2017;40(7):911-9.
71. Chen YD VB, Reaven GM. Plasma lactate concentration in obesity and type 2 diabetes. Diabete & Metabolisme. 1993;19(4).
72. Zhu Y, Li M, Rahman ML, Hinkle SN, Wu J, Weir NL, Lin Y, Yang H, Tsai MY, Ferrara A, et al. 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. 2019;16(9):e1002910.
73. Brown TJ, Brainard J, Song F, Wang X, Abdelhamid A, Hooper L, Group P. 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. 2019;366:l4697.
74. Zhou SJ, Yelland L, McPhee AJ, Quinlivan J, Gibson RA, Makrides M. Fish-oil supplementation in pregnancy does not reduce the risk of gestational diabetes or preeclampsia. Am J Clin Nutr. 2012;95(6):1378-84.
75. Jamilian M, Samimi M, Ebrahimi FA, Hashemi T, Taghizadeh M, Razavi M, Sanami M, Asemi Z. 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. 2017;11(2):459-68.

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

76. Darling AL, Hart KH, Macdonald HM, Horton K, Kang'ombe AR, Berry JL, Lanham-New SA. Vitamin D deficiency in UK South Asian Women of childbearing age: a comparative longitudinal investigation with UK Caucasian women. Osteoporos Int. 2013;24(2):477-88.
77. Sedhain A, Bhattarai GR, Yadav SR, Pandey BR, Pant TP. Geographic and Seasonal Variation of Vitamin D: A Retrospective Study in Two Centers of Nepal. J Nepal Health Res Counc. 2020;18(1):103-7.
78. Lagunova Z, Porojnicu AC, Lindberg F, Hexeberg S, Moan J. The dependency of vitamin D status on body mass index, gender, age and season. Anticancer Res. 2009;29(9):3713-20.

Table 1: Population Characteristics at < 28 weeks' gestation (mean gestational age 26.7 weeks) from the Born in Bradford

	Overall	White European				South Asian				White European vs South Asian P-values		
	Total (n=5339)	**Total (n=2668)**	**GDM (n=128)**	**Non-GDM (n=2540)**	**P-value**	**Total (n=2,671)**	**GDM (n=286)**	**Non-GDM (n=2385)**	**P-value**	**Overall (n=5339)**	**GDM (n=414)**	**Non-GDM (n=4925)**
Age (years)	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
Mothers weight at baseline assessment (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
Mothers' 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[1]												
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 or 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					0.47				< 0.001	< 0.001	< 0.001	< 0.001
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)		695 (26)	49 (17.1)	646 (28.3)				
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

[1] Summary table of population characteristics, expressed as a mean (SE) for continuous variables and counts (%) for categorical variables. Differences between women with and without GDM for continuous variables were tested using a Mann-Whitney test, while differences for categorical variables were tested using Pearson's Chi-squared test.

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

(BiB) cohort.[2]

[2] Body Mass Index (BMI) cut-offs: underweight, ≤ 18.5 kg/m^2 in WE and SAs; normal weight, 18.6-25 kg/m^2 in WEs or 18.6-22.9 kg/m^2 in SAs; 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. SA, South Asian; WE, white European.

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 2: Key metabolite measures (VIP ≥1) that discriminate women diagnosed as GDM from non-GDM women in Partial Least Squares Discriminatory Analysis (PLSDA).[1]

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 Fatty Acids	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)

[1] Variables of Importance in Projection (VIPs) scores [mean and (standard error)] across 20 PLSDA iterations discriminating between GDM (n=414) and non-GDM women (n=414). Model 1: Includes all metabolite measures plus BMI (continuous), maternal age (years), smoking status, parity and multiple pregnancy status. Model 2: Model 1 + ethnicity. GDM, Gestational Diabetes Mellitus; MUFA, total monounsaturated fatty acids; SFA, total saturated fatty acids; VLDL_D, mean diameter of very-low density lipoproteins.

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 3: Metabolite measures associated with dysglycemia in South Asian and white European pregnant women before 28 weeks' gestation (mean gestational age 26.7 weeks)[1]

Dysglycemia (mmol/L)	**Combined analyses of South Asian and White Europeans** (n=5538)	**South Asian** (n=2671)	**White European** (n=2267)
Fasting glucose	Albumin (-)[2] Lactate (-) LDL_D (+)		Lactate (-) Histidine (-) ApoA1 (-) HDL-C (-) HDL2-C (-) LDL_D (+)
2-hour post glucose		Albumin (-)	DHA (+)

[1] Multivariable linear regression analysis was undertaken in the overall population of pregnant women of 2671 South Asians and 2667 White Europeans. Metabolites associated (P<0.05) with measures of fasting glucose or 2-hr post oral glucose tolerance test (OGTT) in the overall population or in ethnic-specific analyses are presented. All models included maternal age (years), gestational age (days), parity, BMI (continuous), and smoking status during pregnancy. No differences in association were identified when BMI was included within the model as a categorical variable.

[2] Direction of associations are presented in brackets — i.e. positive (+) or negative (-). ApoA1, Apolipoprotein A1; DHA, Docosahexaenoic acid; HDL-C, High-density lipoprotein cholesterol; HDL2-C, High-density lipoprotein-2 cholesterol; LDL_D, mean diameter of low-density lipoprotein.

medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 1: Flow chart of study participants from the Born in Bradford (BiB) cohort included within this study.

Figure 2: Circular bar plot identifying key metabolites (VIP ≥ 1) that distinguished 414 women with gestational diabetes mellitus (GDM) from 414 women without GDM.

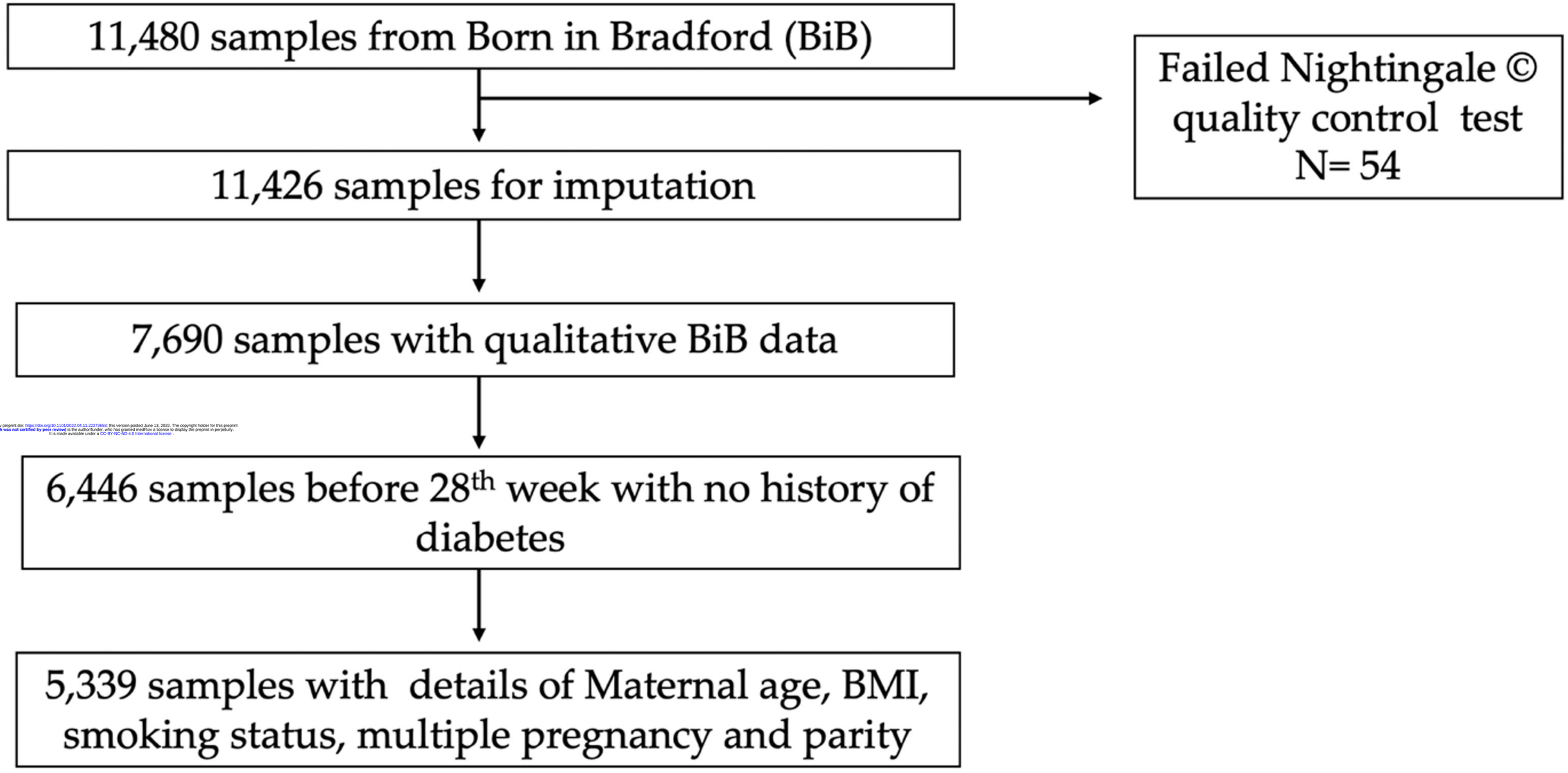
Mean average VIP scores across 20 PLSDA model iterations (n_{cases} = 414). Bars represent standard errors (SEs). The PLSDA included maternal age (years), BMI (continuous), smoking status, parity, and multiple pregnancy status, and ethnicity. Red line denotes VIP cut-off of 1. Units mmol/L unless stated. 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 3: Circular bar plot of ethnically stratified analyses identifying key metabolites (VIP ≥ 1) that distinguished GDM women from non GDM women in South Asians (n_{cases}=286) 2and white Europeans (n_{cases}=128).

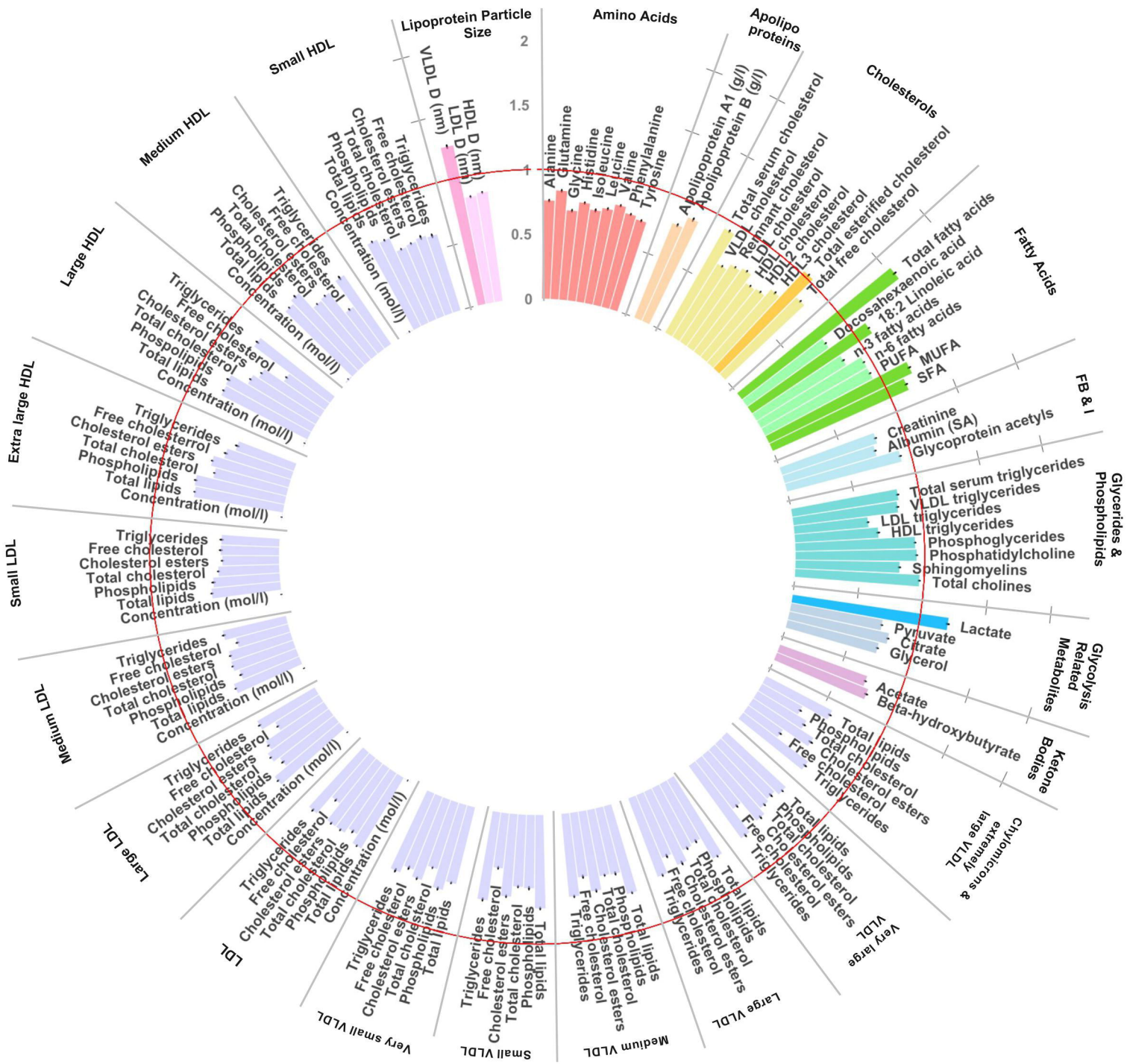
Mean average VIP scores across 20 PLSDA model iterations ($n_{casesSA}$ = 286, $n_{casesWE}$=128). Bars represent standard errors (SEs). The PLSDA was run separately for SA (blue) and WE (red) women and included maternal age (years), BMI (continuous), smoking status, parity, and multiple pregnancy status. Red circular line denotes VIP cut-off of 1. No lipoproteins demonstrated a VIP >1 and were not included in the figure to preserve space. Units mmol/L unless stated. 18.2 LA, 18.2 Linoleic acid; GRM, Glycolysis Related Metabolites; LPS, Lipoprotein Particle Size;

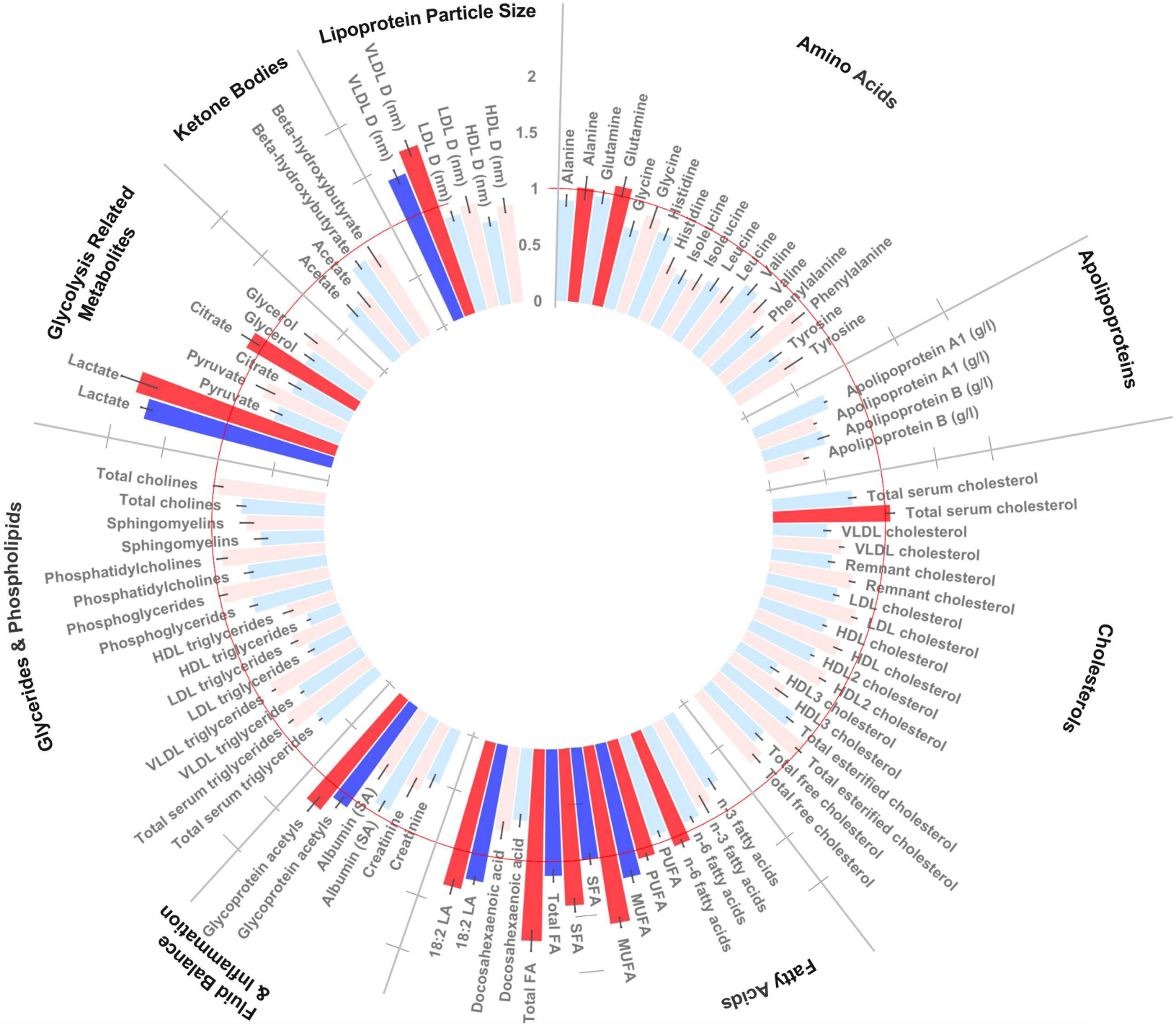
medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

MUFA, total monounsaturated fatty acids; SFA, total saturated fatty acids; Tot FA, Total fatty acids; VLDL_D, mean diameter of very-low density lipoproteins.



medRxiv preprint doi: https://doi.org/10.1101/2022.04.11.22273658; this version posted June 13, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .


Lipoprotein Particle Size
Amino Acids
Apolipo proteins
Cholesterols
Fatty Acids
FB & I
Glycerides & Phospholipids
Glycolysis Related Metabolites
Ketone Bodies
Chylomicrons & extremely large VLDL
Very large VLDL
Large VLDL
Medium VLDL
Small VLDL
Very small VLDL
LDL
Large LDL
Medium LDL
Small LDL
Extra large HDL
Large HDL
Medium HDL
Small HDL
0
0.5
1
1.5
2
VLDL D (nm)
LDL D (nm)
HDL D (nm)
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 cholesterol
HDL cholesterol
HDL2 cholesterol
HDL3 cholesterol
Total esterified cholesterol
Total free cholesterol
Total fatty acids
Docosahexaenoic acid
18:2 Linoleic acid
n-3 fatty acids
n-6 fatty acids
PUFA
MUFA
SFA
Creatinine
Albumin (SA)
Glycoprotein acetyls
Total serum triglycerides
VLDL triglycerides
LDL triglycerides
HDL triglycerides
Phosphoglycerides
Phosphatidylcholine
Sphingomyelins
Total cholines
Lactate
Pyruvate
Citrate
Glycerol
Acetate
Beta-hydroxybutyrate
Total lipids
Phospholipids
Total cholesterol
Cholesterol esters
Free cholesterol
Triglycerides
Concentration (mol/l)


Lipoprotein Particle Size
VLDL D (nm)
VLDL D (nm)
LDL D (nm)
LDL D (nm)
HDL D (nm)
HDL D (nm)
2
1.5
1
0.5
0
Amino Acids
Alanine
Alanine
Glutamine
Glutamine
Glycine
Glycine
Histidine
Histidine
Isoleucine
Isoleucine
Leucine
Leucine
Valine
Valine
Phenylalanine
Phenylalanine
Tyrosine
Tyrosine
Apolipoproteins
Apolipoprotein A1 (g/l)
Apolipoprotein A1 (g/l)
Apolipoprotein B (g/l)
Apolipoprotein B (g/l)
Cholesterols
Total serum cholesterol
Total serum cholesterol
VLDL cholesterol
VLDL cholesterol
Remnant cholesterol
Remnant cholesterol
LDL cholesterol
LDL cholesterol
HDL cholesterol
HDL cholesterol
HDL2 cholesterol
HDL2 cholesterol
HDL3 cholesterol
HDL3 cholesterol
Total esterified cholesterol
Total esterified cholesterol
Total free cholesterol
Total free cholesterol
Fatty Acids
n-3 fatty acids
n-3 fatty acids
n-6 fatty acids
n-6 fatty acids
PUFA
PUFA
MUFA
MUFA
SFA
SFA
Total FA
Total FA
Docosahexaenoic acid
Docosahexaenoic acid
18:2 LA
18:2 LA
Fluid Balance & Inflammation
Creatinine
Creatinine
Albumin (SA)
Albumin (SA)
Glycoprotein acetyls
Glycoprotein acetyls
Glycerides & Phospholipids
Total serum triglycerides
Total serum triglycerides
VLDL triglycerides
VLDL triglycerides
LDL triglycerides
LDL triglycerides
HDL triglycerides
HDL triglycerides
Phosphoglycerides
Phosphoglycerides
Phosphatidylcholines
Phosphatidylcholines
Sphingomyelins
Sphingomyelins
Total cholines
Total cholines
Glycolysis Related Metabolites
Lactate
Lactate
Pyruvate
Pyruvate
Citrate
Citrate
Glycerol
Glycerol
Ketone Bodies
Acetate
Acetate
Beta-hydroxybutyrate
Beta-hydroxybutyrate