Running title: Ethnic-specific metabolic profiles of GDM

Unique Metabolic Profiles in South Asian Women Characterizes Gestational Diabetes in Low and High-Risk Women

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.
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NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
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
LDL_D: Diameter of Low Density Lipoprotein
LPS: Lipoprotein Particle Size
MSEP: Mean Square Error of Prediction
MFA: Monounsaturated Fatty Acids
MW: Mann-Whitney
PCA: Principal Component Analyses
PKs: Pakistanis

CC-BY-NC-ND 4.0 International license
It is made available under a CC-BY-NC-ND 4.0 International license.

OGTT: Oral Glucose Tolerance Test

oPLSDA: Orthogonal Partial Least Squares Discriminatory Analyses

PKC-H: High Weight Pakistani Cases

PKNC-H: High Weight Pakistani Non-Cases

PKC-N: Healthy Weight Pakistani Cases

PKNC-N: Healthy Weight Pakistani 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
Abstract

Background:

Gestational Diabetes Mellitus (GDM) is the most common pregnancy complication to occur worldwide, however prevalence varies substantially between ethnicities with South Asians experiencing up to 3-times the risk of the disease compared to white Europeans (WEs). Factors driving this discrepancy in prevalence and the pathogenesis of GDM are unclear, although the metabolome is of great importance due to the metabolic dysregulation characterised GDM.

Objective:

To characterise and distinguish the metabolic profiles of GDM in two distinct ethnic populations at < 28 weeks’ gestation.

Design:

146 fasting serum metabolite values, quantified by nuclear magnetic resonance, from 2668 pregnant WE and 2671 pregnant Pakistani (PK) women from the Born in Bradford (BIB) cohort, were analysed using partial least squares discriminatory analyses (PLSDA). The presence of a linear relationships 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 were associated with GDM status in both ethnicities, while 6 additional metabolites associated with GDM only in WE women. Unique metabolic profiles were observed in healthy weight women who later developed GDM, with distinct metabolite patterns identified by ethnicity and BMI status. Of the 146
metabolite values analysed in relation to dysglycemia, quantities of lactate, histidine, apolipoprotein A1, HDL cholesterol, HDL2 cholesterol, and DHA, as well as the diameter of very low density lipoprotein particles (nm) were associated with dysglycemia in WE women; while in PK women albumin alone associated with dysglycemia.

**Conclusion**

This is the first study to show that the metabolic risk profile for GDM differs between ethnicities and highlights a need for ethnically appropriate GDM prevention strategies.

**Keywords:** GDM, metabolomics, ethnicity, Pakistani, pregnancy, maternal health, PLSDA, sPLSDA, personalised nutrition
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 levels of prevalence between nations and ethnic groups.

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 contribute to this disparity of risk (5, 8), but diet is the mainstay of most prevention and treatment strategies because of its demonstrated efficacy for managing glucose levels (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 (12, 13). These data suggest that the metabolism and pathology of GDM differ between populations, where some ethnic groups have unique shifts in metabolism that could make them more susceptible to GDM (4, 5, 14-16). 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 (17), with recent evidence demonstrating significant differences in the abundance of these metabolites between ethnic groups (18).

Notably, evidence from Born in Bradford (BiB), a prospective multi-ethnic pregnancy and birth cohort, has demonstrated the need for 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 (19).

As a consequence of this, the Diabetic Pregnancy Study Group called for increased research into the role of the metabolome on GDM in 2018 (20). To date, however, the metabolic drivers of GDM remain unclear with numerous discrepancies within the field, likely due to small, ethnically heterogenous cohorts (21). 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 more robust markers of GDM, 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 simultaneously along with their inter-variable correlations (22) 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 (23). Between 2007 and 2010, BiB recruited 12,453 women (26-28 weeks’ gestation, mean maternal age 27.8) at the Bradford Royal Infirmary, collecting baseline data on 13,776 pregnancies and 13,858 births, with 45% of the cohort of SA origin (23, 24). Of these, 11,480 women provided blood samples for metabolite analyses. Written consent was gained from all participants and ethical approval was granted by the Bradford Research Ethics Committee (ref07/H1302/112)(23).

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°C in the absence of freeze-thaw cycles. (25) Samples were processed using a high-throughput automated NMR platform and 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 remaining samples, ~3% were missing ≥1 metabolite values for unknown reasons. To test if these values were missing at random, the structure of missing data was assessed via the visualization and imputation of missing values (VIM) package within R (26). In addition, multiple correspondence analysis (MCA) was also implemented to assess the randomness of missing data. There was no evidence that the missing data occurred in a non-random pattern, suggesting that it was appropriate to impute missing values. 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 therefore performed using the missMDA package (27). The impact of mothers with ≥3% metabolite values missing on imputation quality was assessed through the exclusion of these mothers, resulting in no detectable difference in imputation quality. Therefore, all 11,426 samples were included for imputation.

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. 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 South Asian origin other than Pakistani (PK) 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 PK and 2,688 White European (WE) women, were retained for analysis. (Figure 1)

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, 23). 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 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 PK women; obese: > 30kg/m$^2$ for WE or >27.5kg/m$^2$ for PK women) (28). Smoking status was self-reported at baseline and during pregnancy.

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 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 (29, 30).

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 (pR²Y =0.05) and validation (pQ²Y= 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. maximisation of outcome variance explained, R²Y) with the minimal error (root mean squared error of estimation (RMSEE)) was selected. Data were pareto scaled and mean-centred prior to analysis. External validity was assessed via 7-fold cross validation. PLSDA models were performed via the ropls package within R (31).

When the size of the outcome groups differed by ≥ 1% the larger group was randomly sampled (n=20) to minimise error. VIPs were mean averaged and standard errors (SEs) calculated across all significant iterations (pR²Y =0.05, pQ²Y=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 and case-status were assessed for significant iterations via a Mann-Whitney (MW) test; this was possible because all comparisons were tested against the same panel of metabolite measures. In order to assess the impact of smoking on PLSDA results, PLSDA models predicting smoking in the overall study population were also performed.
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 PK women (Table 1). Parity was significantly higher in PK women compared to WE women (P<0.001) and parity was only significantly higher in GDM cases compared to non-cases within PK women (Table 1). Smoking during pregnancy was significantly more common in WE women compared to PK women (25% vs 3%; P<0.001). No difference in proportion of singleton pregnancies (>97%) was observed between WE women and PK women. Alcohol intake was not assessed because it was reported by only 1% of PK 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), one glycolysis related metabolite (lactate) (Figure 2).

Modelled independently, the covariates alone explained 12.4% of the variation in GDM status while a model with only metabolites explained 13.5% of the variance in GDM and was non-significant. The 2nd model, which included ethnicity as a covariate, accounted for 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 (PK 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 PK women and 5.8% of the variation in WE women in optimised models (ie. minimisation RMSEE and maximisation of R²Y) although no model comprising metabolites alone was significant (p-value R² > 0.05, Q²>0.05). Conversely, models only including established clinical risk factors (age, BMI, parity, smoking status and multiple pregnancy) were significantly predictive of GDM status and explained 13.3% of the variation in PKs 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 PK women being accounted for, an increase of 13.6% and 6.8% when compared to covariate models in WE and PK respectively. Following adjustments for maternal age, parity, BMI, and smoking status, GDM could be predicted within both ethnicities. Maternal age, parity and BMI were predictors of GDM in both ethnicities (VIP≥1), with BMI the most important predictor of GDM in PK women, while in WE women maternal age was most important predictor (Supplementary S2). Smoking was a predictor of GDM only in WE women. After adjustment for confounders, 7
metabolite variables characterised GDM status (VIP≥1) within both ethnicities (total fatty acids, total MUFA, total SFA, linoleic acid, glycoprotein acetylts, lactate, and diameter of VLDL) (Figure 3, Supplementary S3). Of these metabolites, the VIPs of three (lactate, glycoprotein acetylts, and linoleic acid) characterised GDM status comparatively well between ethnicites (VIP≥1; MW P>0.05), while four metabolite measures (total fatty acids, total MUFA, total SFA, and VLDL_D) characterised GDM in both ethnic groups, but were signficantly more robust 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 PK women. On average, the optimised models explained 26% of the variance of GDM in WE women, and 20% of the variance in PK, women (Supplementary S4).

Metabolites characterised by ethnicity
To explore underlying metabolic profiles within each ethnic group, we identified metabolites that most strongly distinguished WE women and PK 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 S5).

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 choelsterol, n-3 fatty acids, PUFA, MUFA, SFA, phosphatidylcholine and total cholines] were predictive of ethnicity in both GDM cases and non-cases (Supplementary S6).

iii. Post-Hoc Analyses

Characterisation of GDM in Low-Risk Women

BMI, as a suspected mediator along the casual pathway that links metabolism and GDM, was a significant driver of GDM within PK women and WE women. However, a greater mean VIP (± SE) was observed in PK women compared to WE women (VIP_PK = 7.06 ± 0.22 vs. VIP_WE = 4.33 ± 0.22; P<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 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 [23]. Women were classified as ‘healthy or ‘overweight’ based upon ethnic-specific cut-offs (BMI ≥ 25kg/m² for WE women and BMI≥ 23kg/m² for PK women). The analyses focussed on low-risk WE (n= 872) and PK women (n= 864) meaning mothers who (i) had previous children, (ii) smoked during pregnancy, and/or (iii) were ≥35 years of age were excluded. This was done to prevent these covariates from
overpowering the models, while allowing for the formation of a low-weight PK group whose only major observable GDM risk factor was ethnicity.

Healthy weight PK women who developed GDM (PK_{Healthy-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-GDM}; ROC = 0.691) (Supplementary S7). 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 GDM that has been attributed to epigenome-wide modifications (32, 33), metabolic dysregulation, and early dysglycemia that progresses in later life (34-37). It is tempting to speculate that this subgroup of women were exposed to dysglycemia in utero and, despite being of healthy weight, carry a metabolic signature that increased their risk of GDM. Future work in established cohorts (such as Born in Bradford, Generation R, and Nutrigen) looking at trans-generational pregnancy risks are integral to unravel the source of this unique metabolic profile (23, 38, 39). The remaining groups of PK 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 PK women (PK_{Healthy-GDM}) remained distinguishable. A sensitivity analysis after removal of underweight mothers (n_{removed} = 93, BMI≤18.5 kg/m²) did not effect the outcome.

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 S8). Alanine, glutamine, and glycerol were important to distinguish healthy weight PK
women who developed GDM (PK_{Healthy-GDM}) from all others, while fatty acids were important to distinguish PK_{Healthy-GDM} 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 (40, 41). This can provide better separation along fewer components when a large proportion of variance within 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 separating the main group of interest (healthy weight PK cases, PKC-N) with the two most relevant biological comparisons (healthy weight PK non-cases, PKNC-N and healthy weight WE cases, WEC-N).

No significant separation of the PKC-N vs PKNC-N, PKC-N vs PKC-H and PKC-N vs WEC-N groups were identified via SUS plots with oPLSDA. Following the inclusion of BMI and age within the models the PKC-N group was found to separate from all other groups (**Supplementary S9**). 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 PKC-N group but with a low reliability, as shown within SUS-plots.
Association between important metabolites and gestational dysglycemia

Linear regression models investigating the relationship between post-oral glucose tolerance test (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 normalisation. 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(42). All glucose measures were log normalised. Known GDM risk factors, maternal age, gestational age, parity, and smoking status during pregnancy (yes/no), were initially including within 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 utilising ethnic-specific BMI cut-offs for overweight status) to assess the role of early pregnancy BMI as a mediator of metabolite-dysglycemia associations.

Overall, 8 of 146 metabolite measures associated with fasting glucose or 2-hour post glucose (Table 3), all of which were identified as GDM predictors via PLSDA or sPLSDA. The ethnic-specific analysis in WE women demonstrated the greatest number of associations between metabolite measures (i.e., lactate, histidine, ApoA1, HDL cholesterol, HDL 2 cholesterol, and mean diameter of LDL) with fasting glucose, while only a single metabolite measure (DHA) associated with 2-hr post OGTT. In the ethnic-specific PK analysis, only albumin associated with dysglycemia, both fasting glucose and 2-hr post OGTT. 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. (Supplementary S10 & S11).

Discussion

Using a prospective birth cohort with an equal proportion of WE and PK women, we identified 7 metabolite measures that characterized GDM in both WE and PK 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 (43). 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.

Given the overall elevated risk of GDM observed in PK women compared to WE women, even at a healthy BMI (i.e., OR≈3) (44), and the role of ethnicity in predicting GDM, in the present study, we sought to characterize distinct metabolic profiles of PK 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 exclusively within WE women while no metabolite values were predictive solely within PK women.

Although no metabolite value identified solely within WE women 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) (45), and (ii) a cohort study of 431 pregnant Chinese women (12-16 weeks’ gestation), where alanine and glutamine associated with GDM (46). Biologically, alanine, glutamine, and citrate are connected and could moderate dysglycemia through their interact with the tricarboxylic acid cycle (TCA) to promote the formation of TCA intermediates, fatty acid synthesis, and modulate glucagon and insulin secretion (47, 48). 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 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)(49), suggesting 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 (according to VIPs) to characterize GDM status. In WE women and PK women respectively, 75% and 50% of included fatty acid measures were ‘important' to characterize GDM within WE women, while in PK women fatty acids constituted more than half of all important metabolites. 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 (50, 51).

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 PK 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 (52-54) and a role in ethnic-associated GDM risk (52, 55, 56). Indeed, n6 PUFA-derived eicosanoids show discriminatory qualities between type-2 diabetics and controls with good accuracy ($R^2_X = 0.824$, $R^2_Y = 0.995$, $Q^2 = 0.779$) and were identified as proposed mediators of dysglycemia within a Chinese population (57).

Longitudinal analyses to evaluate the association between changes in PUFA and eicosanoids levels on dysglycemia during pregnancy are required to better understand this association.

The association between VLDL diameter and dysglycemia is supported by recent hypothesis linking insulin resistance, triglyceride synthesis, and increased VLDL diameter (58, 59). Although we cannot disregard that VLDLs are sensitive to level of fasting (60) (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 (61). 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, our work agrees with recent evidence that has demonstrated differences in glycoprotein acetyls.
between GDM (n=239) and non-GDM (n=818) pregnant women (GDM Z-score =
+0.48; 95% CI: 0.22, 0.74 vs non-GDM).

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) (62) and pathway analyses that propose lactate as a regulator of insulin
resistance and a marker metabolic syndrome severity (63, 64). Post-hoc analysis
demonstrated no association between glycoprotein acetylts and glucose levels, while
lactate and mean diameter of VLDL associated with fasting glucose in WE women
but not PK 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 (43, 65, 66).

Of the numerous fatty acid measures that associated with GDM, only DHA
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 (51, 67, 68). 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) (51), 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 (69). 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) (70) to impart an effect, levels of which vary
considerably between populations, seasons, and geographic region (71-73).
The study aimed to increase and test generalisability of results within a diverse population; however, our results may not be generalisable 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 generalisable 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 characterise GDM within a large prospective cohort with an equal representation of WE women and women from an minority ethnic population. Thirdly, the biological validity of the identified metabolites was tested and many correlated with continual glucose measures; and although confounding cannot be eliminated, all models included known GDM confounders and modelling characterising the overall metabolic differences between ethnicities were also performed to test the presence of reverse associations. Finally, diet is a contributor to metabolite levels but comprehensive dietary data was not available for our analysis. Future work with comprehensive dietary records are needed evaluate the moderating effect of diet on metabolism and GDM risk, and its interaction with ethnicity.

**Conclusion**

In conclusion, this study has identified unique and shared metabolic profiles that characterise GDM in WE and PK 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 aetiological
factors responsible for the heightened level of GDM risk experienced by PK 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
References


Diabetes and Its Intrinsic and Extrinsic Determinants: Systematic Review and Meta-Analysis.


### Table 1: Population Characteristics

Summary table of population characteristics, expressed as a mean (SD) 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.
Table 2: Key metabolite measures that discriminate women diagnosed as GDM from non-GDM.

Variables of Importance in Projection (VIPs) scores [mean and (standard error)] across 20 partial least squares discriminatory analyses (PLSDA) iterations discriminating between GDM and non-GDM women. Model 1: **Additional covariates:** maternal age, BMI, smoking status, parity, and multiple pregnancy status. Model 2: Model 1 covariates + ethnicity. SFA = total saturated fatty acids; MUFA = total monounsaturated fatty acids; VLDL_D = diameter of very-low density lipoproteins.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>6.4 (0.03)</td>
<td>5.9 (0.03)</td>
</tr>
<tr>
<td>BMI</td>
<td>5.4 (0.04)</td>
<td>5.1 (0.02)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>-</td>
<td>2.9 (0.02)</td>
</tr>
<tr>
<td>Parity</td>
<td>2.4 (0.01)</td>
<td>2.3 (0.01)</td>
</tr>
<tr>
<td>Smoking Status</td>
<td>1.9 (0.02)</td>
<td>1.7 (0.01)</td>
</tr>
<tr>
<td>Multiple Pregnancy</td>
<td>1.5 (0.01)</td>
<td>1.3 (0.009)</td>
</tr>
<tr>
<td>Lactate</td>
<td>1.5 (0.01)</td>
<td>1.2 (0.008)</td>
</tr>
<tr>
<td>VLDL_D</td>
<td>1.3 (0.01)</td>
<td>1.3 (0.01)</td>
</tr>
<tr>
<td>Total Fatty Acids</td>
<td>1.2 (0.01)</td>
<td>1.5 (0.01)</td>
</tr>
<tr>
<td>Total MUFA</td>
<td>1.2 (0.001)</td>
<td>1.2 (0.008)</td>
</tr>
<tr>
<td>18:2 Linoleic Acid</td>
<td>1.1 (0.01)</td>
<td>1.1 (0.004)</td>
</tr>
<tr>
<td>Total SFA</td>
<td>1.1 (0.01)</td>
<td>1.2 (0.007)</td>
</tr>
<tr>
<td>Esterified Cholesterol</td>
<td>-</td>
<td>1.0 (0.008)</td>
</tr>
</tbody>
</table>
Table 3: Metabolite measures associated with dysglycemia.

Analysis was undertaken in the overall population of pregnant women (white Europeans and Pakistani) and within each ethnic group. Metabolites associated (P<0.05) with measures of fasting glucose or 2-hr post oral glucose tolerance test (OGTT) in multivariable linear regression in the overall population or in ethnic-specific analyses are presented. All models included maternal age, gestational age, parity, BMI, and smoking status during pregnancy. For ease of interpretation, only direction of associations are presented in brackets — i.e. positive (+) or negative (-). Effect sizes and measures of variances are available in Supplementary S8 and S9. ApoA1, apolipoprotein A1; high-density lipoprotein cholesterol, HDL-C; high-density lipoprotein-2 cholesterol, HDL2-C; DHA, docosahexaenoic acid; and low-density lipoprotein diameter, LDL_D

<table>
<thead>
<tr>
<th>Dysglycemia (mmol/L)</th>
<th>Pakistani and White European (n=5538)</th>
<th>Pakistani (n=2671)</th>
<th>White European (n=2267)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fasting glucose</td>
<td>Albumin (-)</td>
<td>Albumin (-)</td>
<td>Lactate (-)</td>
</tr>
<tr>
<td></td>
<td>Lactate (-)</td>
<td></td>
<td>Histidine (-)</td>
</tr>
<tr>
<td></td>
<td>LDL_D (+)</td>
<td></td>
<td>ApoA1 (-)</td>
</tr>
<tr>
<td>2-hour post glucose</td>
<td>Albumin (-)</td>
<td>DHA (+)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ApoA1, apolipoprotein A1; high-density lipoprotein cholesterol, HDL-C; high-density lipoprotein-2 cholesterol, HDL2-C; DHA, docosahexaenoic acid; and low-density lipoprotein diameter, LDL_D
Figure 1: Flow of study participants.

PCA: Principal Component Analysis. PLSDA: Partial Least Squares Discriminatory Analysis. sPLSDA: Sparse PLSDA. oPLSDA: Orthogonal PLSDA.

Figure 2: Circular bar plot of metabolites (VIP ≥ 1) distinguishing the metabolomes of women with gestational diabetes (GDM) cases from non-GDM women.

The PLSDA included maternal age, BMI, smoking status, parity, and multiple pregnancy status, and ethnicity. Red line denotes VIP cut-off of 1. SFA = total saturated fatty acids; MUFA = total monounsaturated fatty acids; VLDL_D = diameter of very-low density lipoproteins. GRM: Glycolysis Related Metabolites. LPS: Lipoprotein Particle Size. Units mmol/L unless stated.

Figure 3: Circular bar plot of all metabolite variables VIP scores as predictors of gestational diabetes (GDM) in ethnically stratified analysis.

The PLSDA was run separately for PK (blue) and WE (red) women and included maternal age, BMI, 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. SFA = total saturated fatty acids; MUFA = total monounsaturated fatty acids; VLDL_D = diameter of very-low density lipoproteins. GRM: Glycolysis Related Metabolites. LPS: Lipoprotein Particle Size. Units mmol/L unless stated.
11,480 samples from Born in Bradford (BiB)

11,426 samples for imputation

7,690 samples with qualitative BiB data

6,446 samples before 28th week with no history of diabetes

5,339 samples with details of Maternal age, BMI, smoking status, multiple pregnancy and parity