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.
[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]

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

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 ethnical approval was granted by the Bradford Research Ethnics 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℃ 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 (pR^2Y =0.05) and validation (pQ^2Y= 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, R^2Y) 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 (pR^2Y =0.05, pQ^2Y=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.

188 **Results**

189 **i. Population Characteristics**

190 The mean age of participant was 26.7 years and had a mean BMI of 26 kg/m^2. WE
191 women were significantly older and had higher BMIs compared to PK women (**Table**
192 **1**). Parity was significantly higher in PK women compared to WE women (P<0.001)
193 and parity was only significantly higher in GDM cases compared to non-cases within
194 PK women (**Table 1**). Smoking during pregnancy was significantly more common in
195 WE women compared to PK women (25% vs 3%; P<0.001). No difference in
196 proportion of singleton pregnancies (>97%) was observed between WE women and
197 PK women. Alcohol intake was not assessed because it was reported by only 1% of
198 PK women. The mean time of sample collection was 187 gestational days.

199 **ii. Primary Analysis**

200 **Metabolite Characterisation of GDM**

201 In the 1st model, an overall analysis of the full cohort (i.e., both ethnic groups),
202 PLSDA explained 21.7% of the variation between the GDM and non-GDM groups
203 and confirmed maternal age and BMI as primary risk factors for GDM risk followed
204 by parity, smoking status, and having a non-singleton pregnancy as the primary
205 drivers of GDM (**Table 2**). In the full model, 7 metabolite values reported VIPs ≥ 1,
206 including 4 fatty acid metabolite measures (total fatty acids, 18:2 linoleic acid, total
207 MUFA and total SFA), one glycolysis related metabolite (lactate) (**Figure 2**).
208 Modelled independently, the covariates alone explained 12.4% of the variation in
209 GDM status while a model with only metabolites explained 13.5% of the variance in
210 GDM and was non-significant. The 2nd model, which included ethnicity as a
211 covariate, accounted for 26.6% of the variation between the GDM and non-GDM

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

218

219 **Ethnically-stratified analysis of metabolites characterising GDM**

220 In an ethnically stratified analysis (20 iterations), models only including metabolites
221 accounted for a median average of 6.5% of the variation in GDM status in PK
222 women and 5.8% of the variation in WE women in optimised models (ie.
223 minimisation RMSEE and maximisation of R^2Y) although no model comprising
224 metabolites alone was signficant (p- value $R^2 > 0.05$, $Q^2>0.05$). Conversely, models
225 only including established clinical risk factors (age, BMI, parity, smoking status and
226 multiple pregnancy) were significantly predictive of GDM status and explained 13.3%
227 of the variation in PKs 12.8% of the variation in WEs. The addition of metabolites to
228 these covariate models also resulted in the significant prediction of GDM. These
229 models resulted in 26% of the variance in GDM status in WE women and 20% of the
230 variance in PK women being accounted for, an increase of 13.6% and 6.8% when
231 compared to covariate models in WE and PK respectively. Following adjustments for
232 maternal age, parity, BMI, and smoking status, GDM could be predicted within both
233 ethnicites. Maternal age, parity and BMI were predictors of GDM in both ethnicities
234 (VIP≥1), with BMI the most impotant predictor of GDM in PK women , while in WE
235 women maternal age was most important predictor (**Supplementary S2**). Smoking
236 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 acetyls, lactate, and diameter of VLDL) (**Figure 3**, **Supplementary S3**). 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), 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, phosphatidylchlorine 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 cholesterol, 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/m^2 for WE women and BMI≥ 23kg/m^2 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

285 overpowering the models, while allowing for the formation of a low-weight PK group
286 whose only major observable GDM risk factor was ethnicity.
287 Healthy weight PK women who developed GDM ($PK_{Healthy\text{-}GDM}$) presented the
288 most distinct metabolic profile (Receiver Operator Curve; ROC = 0.783), but were
289 most similar to healthy WE women who developed GDM ($WE_{Healthy\text{-}GDM}$; ROC =
290 0.691) (**Supplementary S7**). The reason for this shared and distinct pattern of
291 metabolites in 'healthy' weight women who developed GDM is unclear and many
292 hypotheses are possible. One hypothesis may be that the pattern is an artifact of
293 their fetal programming. Adult offspring from GDM pregnancies are at increased risk
294 of dysglycemia, diabetes, and GDM that has been attributed to epigenome-wide
295 modifications (32, 33), metabolic dysregulation, and early dysglycemia that
296 progresses in later life (34-37). It is tempting to speculate that this subgroup of
297 women were exposed to dysglycemia in utero and, despite being of healthy weight,
298 carry a metabolic signature that increased their risk of GDM. Future work in
299 established cohorts (such as Born in Bradford, Generation R, and Nutrigen) looking
300 at trans-generational pregnancy risks are integral to unravel the source of this unique
301 metabolic profile (23, 38, 39). The remaining groups of PK women and WE women,
302 including all non-GDM and overweight women, were indistinguishable from each
303 other. When the dataset was reduced to only the 4 GDM groups, healthy PK women
304 ($PK_{Healthy\text{-}GDM}$) remained distinguishable. A sensitivity analysis after removal of
305 underweight mothers ($n_{removed}$ = 93, BMI≤18.5 kg/m^2) did not effect the outcome.
306 Metabolites selected by sPLSDA in each comparison were fed into PLSDA
307 models (20 iterations) alongside highly correlated metabolites (Pearsons correlation
308 coefficient ≥0.9) to identify key metabolic drivers of this separation (**Supplementary**
309 **S8**). Alanine, glutamine, and glycerol were important to distinguish healthy weight PK

310 women who developed GDM ($PK_{Healthy- GDM}$) from all others, while fatty acids were
311 important to distinguish $PK_{Healthy- GDM}$ other GDM cases. Interestingly, in healthy
312 women, aromatic and branched chain amino acids distinguished GDM and non-GDM
313 women between (but not within) ethnic groups. Glycerol distributions were
314 significantly different in all comparisons (MW <0.05).

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

316 Orthogonal partial least squares discriminant analysis (oPLSDA) is a supervised
317 multivariate technique that separates variation within each predictor variables based
318 upon its linear (correlated) and orthogonal (uncorrelated) association with the
319 outcome variable (40, 41). This can provide better separation along fewer
320 components when a large proportion of variance within the dataset does not directly
321 correlate with the outcome variable. Furthermore, through the creation of Shared and
322 Unique Structure (SUS) plots it is possible to determine shared and unique factors
323 separating the main group of interest (healthy weight PK cases, PKC-N) with the
324 two most relevant biological comparisons (healthy weight PK non-cases, PKNC-N
325 and healthy weight WE cases, WEC-N).

326 No significant separation of the PKC-N vs PKNC-N, PKC-N vs PKC-H and PKC-
327 N vs WEC-N groups were identified via SUS plots with oPLSDA. Following the
328 inclusion of BMI and age within the models the PKC-N group was found to separate
329 from all other groups (**Supplementary S9**). BMI was the only variable found to be
330 responsible for this separation with a high magnitude and reliability. Pyruvate, L-HDL
331 and XL-HDL had a small impact on the separation of the PKC-N group but with a low
332 reliability, as shown within SUS-plots.

333 **Association between important metabolites and gestational dysglycemia**

334 Linear regression models investigating the relationship between post- oral glucose
335 tolerance test (OGTT) measures (fasting glucose and 2-hour post-OGTT) were
336 performed on all metabolite values identified as important (VIP≥1) in characterizing
337 GDM status. Normality of glucose measures and metabolite values were assessed
338 using histograms and Q-Q plots. Most metabolites (136/146) required normalisation.
339 Normality was most often achieved by log transformation (59 metabolite values);
340 however, in some cases square-root and normal score transformation (NST) were
341 implemented via the 'rcompanion' package(42). All glucose measures were log
342 normalised. Known GDM risk factors, maternal age, gestational age, parity, and
343 smoking status during pregnancy (yes/no), were initially including within the models.
344 When significant associations were observed between metabolite values and
345 glucose in this exploratory analysis (P<0.05), BMI was added to the models (initially
346 as a continuous and then as a binary variable ustilising ethnic-specific BMI cut-offs
347 for overweight status) to assess the role of early pregnancy BMI as a mediator of
348 metabolite-dysglycemia associations.

349 Overall, 8 of 146 metabolite measures associated with fasting glucose or 2-hour post
350 glucose (**Table 3**), all of which were identified as GDM predictors via PLSDA or
351 sPLSDA. The ethnic-specific analysis in WE women demonstrated the greatest
352 number of associations between metabolite measures (i.e., lactate, histidine, ApoA1,
353 HDL cholesterol, HDL 2 cholesterol, and mean diameter of LDL) with fasting
354 glucose, while only a single metabolite measure (DHA) associated with 2-hr post
355 OGTT. In the ethnic-specific PK analysis, only albumin associated with dysglcyemia,
356 both fasting glucose and 2-hr post OGTT. In the ethnic-combined analysis,
357 associations between albumin, lactate, and mean diameter of LDL with fasting

358 glucose retained significance. Adjusting for BMI as a continuous or binary variable
359 had no impact on the associations. (**Supplementary S10 & S11)**.

360

361 **Discussion**

362 Using a prospective birth cohort with an equal proportion of WE and PK women, we
363 identified 7 metabolite measures that characterized GDM in both WE and PK women
364 — 4 of which were more predictive in WE women. These results agree with the
365 Omega cohort (78.5% non-Hispanic white; nested case-control; 46 cases, 47
366 controls) that highlighted a distinct metabolic profile at 16-weeks' gestation
367 (comprised of fatty acids, sugars, alcohols, amino acids and organic acids),
368 associated with future GDM diagnosis (43). Although the metabolite patterns
369 identified by the Omega study were not predictive, our predictive multivariate
370 analysis (and a previous univariate analyses [1]) found similar associations between
371 GDM and many of these metabolites (i.e., amino acids, glycolysis related
372 metabolites, and fatty acids), and offers further evidence of ethnic-specific
373 associations.

374 Given the overall elevated risk of GDM observed in PK women compared to
375 WE women, even at a healthy BMI (i.e., OR≈3) (44), and the role of ethnicity in
376 predicting GDM, in the present study, we sought to characterize distinct metabolic
377 profiles of PK and WE women. Of the 146 metabolite values tested, 7 were important
378 for stratifying GDM and non-GDM women in the overall population (lactate, mean
379 density of VLDL particles, total fatty acids, total MUFAs, 18:2 linoleic acid, total SFA
380 and esterified cholesterol). Following stratification by ethnicity, alanine, glutamine,
381 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

407 that fatty acids alter insulin resistance and insulin secretion during pregnancy (50,
408 51).

409 Furthermore, fatty acids (total MUFAs, total n-3 PUFAs, total n-6 PUFAs, total
410 PUFAs, and DHA) were identified as key metabolic factors to distinguish healthy-
411 weight PK and WE women who developed GDM. Interestingly, we highlighted
412 associations between n-6 PUFA and total PUFAs with GDM that were specific to WE
413 women. Given the equal sample sizes between groups, and that fatty acids were
414 important to characterize ethnicity, it is suggestive of ethnic differences in PUFA
415 metabolism (52-54) and a role in ethnic-associated GDM risk (52, 55, 56). Indeed, n6
416 PUFA-derived eicosanoids show discriminatory qualities between type-2 diabetics
417 and controls with good accuracy ($R^2X = 0.824$, $R^2Y = 0.995$, $Q^2 = 0.779$) and were
418 identified as proposed mediators of dysglycemia within a Chinese population (57).
419 Longitudinal analyses to evaluate the association between changes in PUFA and
420 eicosanoids levels on dysglycemia during pregnancy are required to better
421 understand this association.

422 The association between VLDL diameter and dysglycemia is supported by
423 recent hypothesis linking insulin resistance, triglyceride synthesis, and increased
424 VLDL diameter (58, 59). Although we cannot disregard that VLDLs are sensitive to
425 level of fasting (60) (*as our participants were subjected to prior to blood collection*),
426 evidence also suggests that ethnic-specific genetic variants associate with ethnic-
427 specific differences in VLDL diameter (61). Although there has been less work on the
428 possible association between glycoprotein acetyls (a marker of systemic
429 inflammation) and GDM, and future work is required in this area, our work agrees
430 with recent evidence that has demonstrated differences in glycoprotein acetyls

431 between GDM (n=239) and non-GDM (n=818) pregnant women (GDM Z-score =
432 +0.48; 95% CI: 0.22, 0.74 vs non-GDM).

433 Lactate was one of the strongest predictors of GDM within both groups, in
434 agreement with evidence from a case-control study in China (n=12 GDM; n=10
435 controls) (62) and pathway analyses that propose lactate as a regulator of insulin
436 resistance and a marker metabolic syndrome severity (63, 64). *Post-hoc* analysis
437 demonstrated no association between glycoprotein acetyls and glucose levels, while
438 lactate and mean diameter of VLDL associated with fasting glucose in WE women
439 but not PK women. The multi-ethnic HAPO cohort demonstrated a similar ethnic-
440 specific association between lactate and fasting glucose within individuals of
441 Northern European ancestry but not minority ethnic groups (43, 65, 66).

442 Of the numerous fatty acid measures that associated with GDM, only DHA
443 associated with a post-OGTT measures of glucose and only in WE women. Overall,
444 DHA is considered a protective metabolite against insulin resistance (e.g., HOMA-
445 IR); however, recent evidence suggests high heterogeneity (51, 67, 68). As we did,
446 researchers investigating the Camden pregnancy cohort (n=1,368) reported a
447 significant positive linear association between DHA and HOMA-IR (0.303 ± 0.152 per
448 unit DHA %; $P<0.05$) (51), while conversely, the DOMINO trial (n=1990 pregnant
449 women) reported no difference in 1-hr post-OGTT glucose levels between DHA
450 supplemented mothers and controls (69). The reason for such discrepancies is
451 unclear but may be that n-3 PUFAs (such as DHA) require interactions with other
452 metabolites (e.g., Vitamin D) (70) to impart an effect, levels of which vary
453 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 chracterise 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

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, 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, 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, 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, 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. Wang Q, Wurtz P, Auro K, Makinen VP, Kangas AJ, Soininen P, et al. Metabolic profiling of pregnancy: cross-sectional and longitudinal evidence. BMC Med. 2016;14(1):205.
15. 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.
16. McCabe CF, Perng W. Metabolomics of Diabetes in Pregnancy. Curr Diab Rep. 2017;17(8):57.

17. Chen Q, Francis E, Hu G, Chen L. Metabolomic profiling of women with gestational diabetes mellitus and their offspring: Review of metabolomics studies. Journal of Diabetes and its Complications. 2018;32(5):512-23.
18. Mitro SD, Wu J, Rahman ML, Cao Y, Zhu Y, Chen Z, et al. Longitudinal Plasma Metabolomics Profile in Pregnancy—A Study in an Ethnically Diverse U.S. Pregnancy Cohort. Nutrients. 2021;13(9):3080.
19. Farrar D, Fairley L, Santorelli G, Tuffnell D, Sheldon TA, Wright J, et al. 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.
20. 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.
21. Mao X, Chen X, Chen C, Zhang H, Law KP. Metabolomics in gestational diabetes. Clin Chim Acta. 2017;475:116-27.
22. 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.
23. Wright J, Small N, Raynor P, Tuffnell D, Bhopal R, Cameron N, et al. Cohort Profile: the Born in Bradford multi-ethnic family cohort study. Int J Epidemiol. 2013;42(4):978-91.
24. Bird PK, McEachan RRC, Mon-Williams M, Small N, West J, Whincup P, et al. Growing up in Bradford: protocol for the age 7-11 follow up of the Born in Bradford birth cohort. BMC Public Health. 2019;19(1):939-.
25. Taylor K, McBride N, J Goulding N, Burrows K, Mason D, Pembrey L, et al. Metabolomics datasets in the Born in Bradford cohort [version 2; peer review: 1 approved, 1 approved with reservations]. Wellcome Open Research. 2021;5(264).
26. Alexander Kowarik MT. Imputation with the R Package VIM. Journal of Statistical Software. 2016;74(7):1-16.
27. Josse J, Husson F. missMDA: A Package for Handling Missing Values in Multivariate Data Analysis. Journal of Statistical Software. 2016;70(1).
28. consultation We. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. The Lancet. 2004;363(9403):157-63.
29. Perreault M, Zulyniak MA, Badoud F, Stephenson S, Badawi A, Buchholz A, et al. A distinct fatty acid profile underlies the reduced inflammatory state of metabolically healthy obese individuals. PloS one. 2014;9(2):e88539-e.
30. Badoud F, Lam KP, DiBattista A, Perreault M, Zulyniak MA, Cattrysse B, et al. Serum and Adipose Tissue Amino Acid Homeostasis in the Metabolically Healthy Obese. Journal of Proteome Research. 2014;13(7):3455-66.
31. 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.
32. Moen G-H, Sommer C, Prasad RB, Sletner L, Groop L, Qvigstad E, et al. Mechanisms in endocrinology: Epigenetic modifications and gestational diabetes: A systematic review of published literature. European journal of endocrinology. 2017;176(5):R247-R67.
33. Yang I, Zhang W, Davidson E, Fingerlin T, Kechris K, Dabelea D. Epigenetic marks of in utero exposure to gestational diabetes and childhood adiposity outcomes: the EPOCH study. Diabetic Medicine. 2018;35(5):612-20.

34. 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.
35. Clausen TD, Mathiesen ER, Hansen T, Pedersen O, Jensen DM, Lauenborg J, et al. 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.
36. 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.
37. Anand SS, Gupta M, Teo KK, Schulze KM, Desai D, Abdalla N, et al. Causes and consequences of gestational diabetes in South Asians living in Canada: results from a prospective cohort study. CMAJ open. 2017;5(3):E604.
38. Jaddoe VW, Mackenbach JP, Moll HA, Steegers EA, Tiemeier H, Verhulst FC, et al. The Generation R Study: Design and cohort profile. Eur J Epidemiol. 2006;21(6):475-84.
39. de Souza RJ, Zulyniak MA, Desai D, Shaikh MR, Campbell NC, Lefebvre DL, 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.
40. Worley B, Powers R. Multivariate Analysis in Metabolomics. Curr Metabolomics. 2013;1(1):92-107.
41. Blasco H, Blaszczynski J, Billaut JC, Nadal-Desbarats L, Pradat PF, Devos D, et al. Comparative analysis of targeted metabolomics: dominance-based rough set approach versus orthogonal partial least square-discriminant analysis. J Biomed Inform. 2015;53:291-9.
42. Mangiafico S. rcompanion: Functions to Support Extension Education Program Evaluation. R package version 2.4.1. ed2021.
43. 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.
44. 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.
45. Del Coco L, Vergara D, De Matteis S, Mensa E, Sabbatinelli J, Prattichizzo 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).
46. Jiang R, Wu S, Fang C, Wang C, Yang Y, Liu C, et al. Amino acids levels in early pregnancy predict subsequent gestational diabetes. J Diabetes. 2020;12(7):503-11.
47. 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.
48. 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.
49. 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.
50. Villafan-Bernal JR, Acevedo-Alba M, Reyes-Pavon R, Diaz-Parra GA, Lip-Sosa DL, Vazquez-Delfin HI, et al. 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-.
51. 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.
52. Gray RG, Kousta E, McCarthy MI, Godsland IF, Venkatesan S, Anyaoku V, et al. 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 in health and disease. 2013;12:25-.
53. Benedetti S, Al-Tannak NF, Alzharani M, Moir HJ, Stensel DJ, Thackray AE, et al. 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.
54. Ralston JC, Zulyniak MA, Nielsen DE, Clarke S, Badawi A, El-Sohemy A, et al. Ethnic- and sex-specific associations between plasma fatty acids and markers of insulin resistance in healthy young adults. Nutrition & metabolism. 2013;10(1):42.
55. Yuen L, Wong VW, Simmons D, Simmons D. Ethnic Disparities in Gestational Diabetes. Current Diabetes Report. 2018;18(68).
56. 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.
57. 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.
58. Zhao H, Li H, Chung ACK, Xiang L, Li X, Zheng 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.
59. Krauss RM. Lipids and lipoproteins in patients with type 2 diabetes. Diabetes Care. 2004;27(6):1496-504.
60. Nordestgaard BG. A Test in Context: Lipid Profile, Fasting Versus Nonfasting. Journal of the American College of Cardiology. 2017;70(13):1637-46.
61. Frazier-Wood AC, Manichaikul A, Aslibekyan S, Borecki IB, Goff DC, Hopkins PN, et al. Genetic variants associated with VLDL, LDL and HDL particle size differ with race/ethnicity. Human Genetics. 2013;132(4):405-13.
62. Liu T, Li J, Xu F, Wang M, Ding S, Xu H, et al. Comprehensive analysis of serum metabolites in gestational diabetes mellitus by UPLC/Q-TOF-MS. Anal Bioanal Chem. 2016;408(4):1125-35.
63. 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.
64. Jones TE, Pories WJ, Houmard JA, Tanner CJ, Zheng D, Zou K, et al. 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.
65. Jacob S, Nodzenski M, Reisetter AC, Bain JR, Muehlbauer MJ, Stevens RD, 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.
66. Chen YD VB, Reaven GM. Plasma lactate concentration in obesity and type 2 diabetes. Diabete & Metabolisme. 1993;19(4).
67. Zhu Y, Li M, Rahman ML, Hinkle SN, Wu J, Weir NL, 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.
68. Brown TJ, Brainard J, Song F, Wang X, Abdelhamid A, Hooper L, et al. 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.
69. 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.
70. Jamilian M, Samimi M, Ebrahimi FA, Hashemi T, Taghizadeh M, Razavi M, et al. 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.
71. Darling AL, Hart KH, Macdonald HM, Horton K, Kang'ombe AR, Berry JL, et al. 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.
72. 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.
73. 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.

	Overall	White European (WE)				Pakistani (PK)				WE vs PK 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	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 Booking Weight (Kg)	68.8 (0.2)	72.0 (0.3)	76.8 (1.6)	71.8 (0.3)	0.002	65.5 (0.3)	72.3 (0.9)	64.7 (0.3)	< 0.001	< 0.001	0.03	< 0.001
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												
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

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.

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)

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.

Dysglycemia (mmol/L)	**Pakistani and White European** (n=5538)	**Pakistani** (n=2671)	**White European** (n=2267)
Fasting glucose	Albumin (-) Lactate (-) LDL_D (+)	Albumin (-)	Lactate (-) Histidine (-) ApoA1 (-) HDL-C (-) HDL2-C (-) LDL_D (+)
2-hour post glucose		Albumin (-)	DHA (+)

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

Figure 1: Flow of study participants.

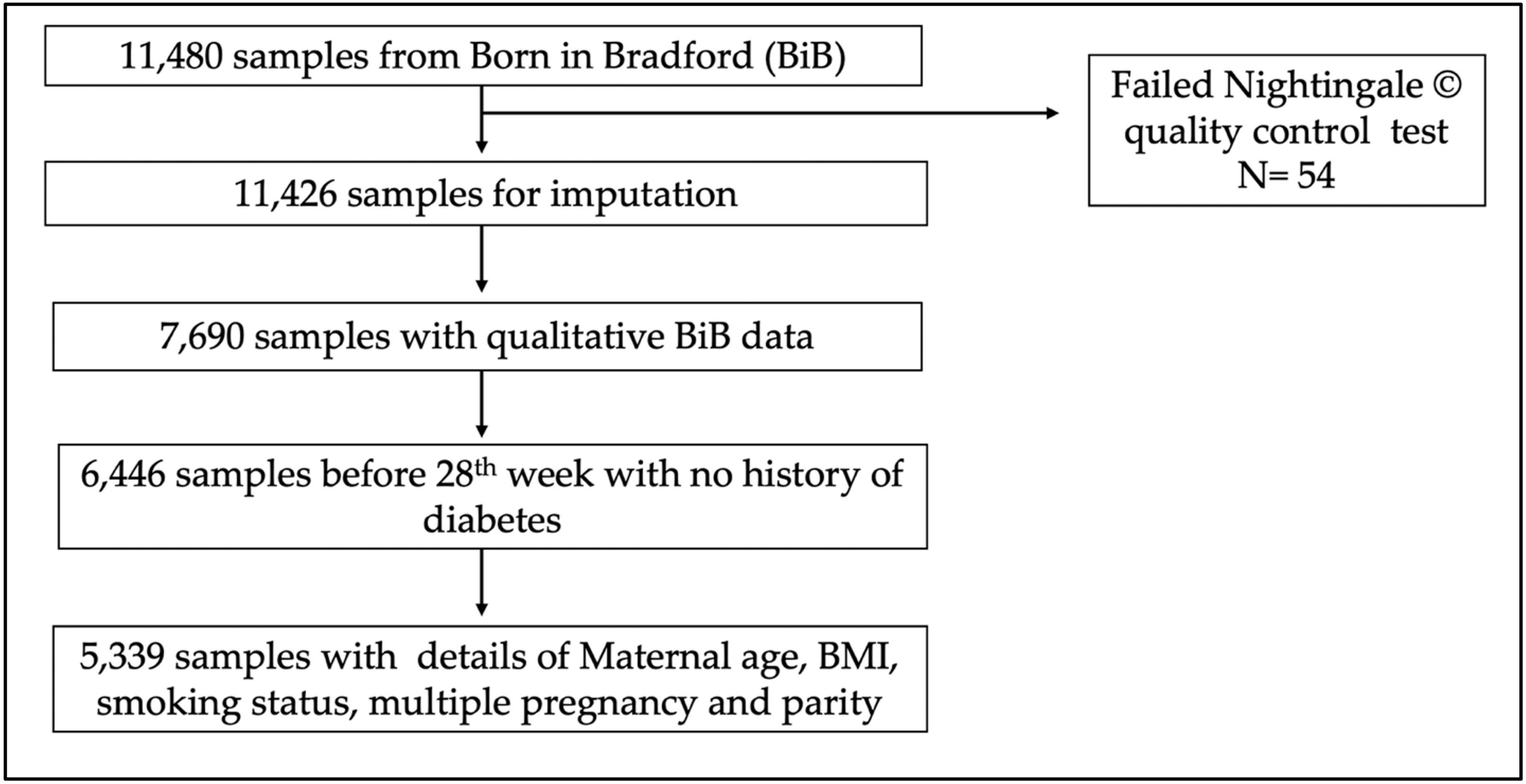
PCA: Principal Component Analysis. PLSDA: Partial Least Squares Discriminatory Analysis. sPLSDA: Sparse PLSDA. oPLSDA: Orthogonal PLSDA.

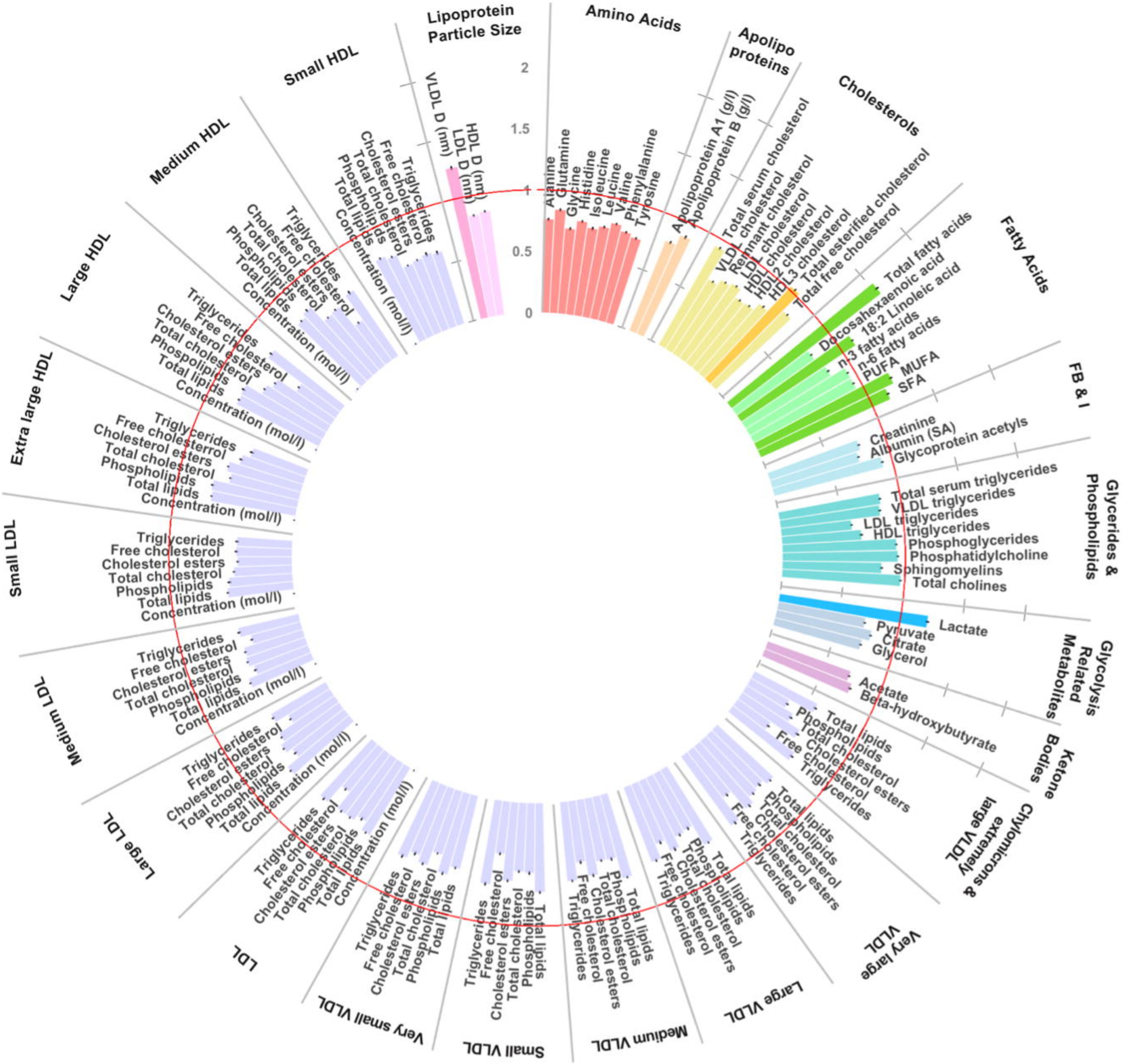
Figure 2: Circular bar plot of metabolites (VIP ≥ 1) distinguishing the metabolomes of women with gestational diabetes (GDM) cases from non-GDM women.

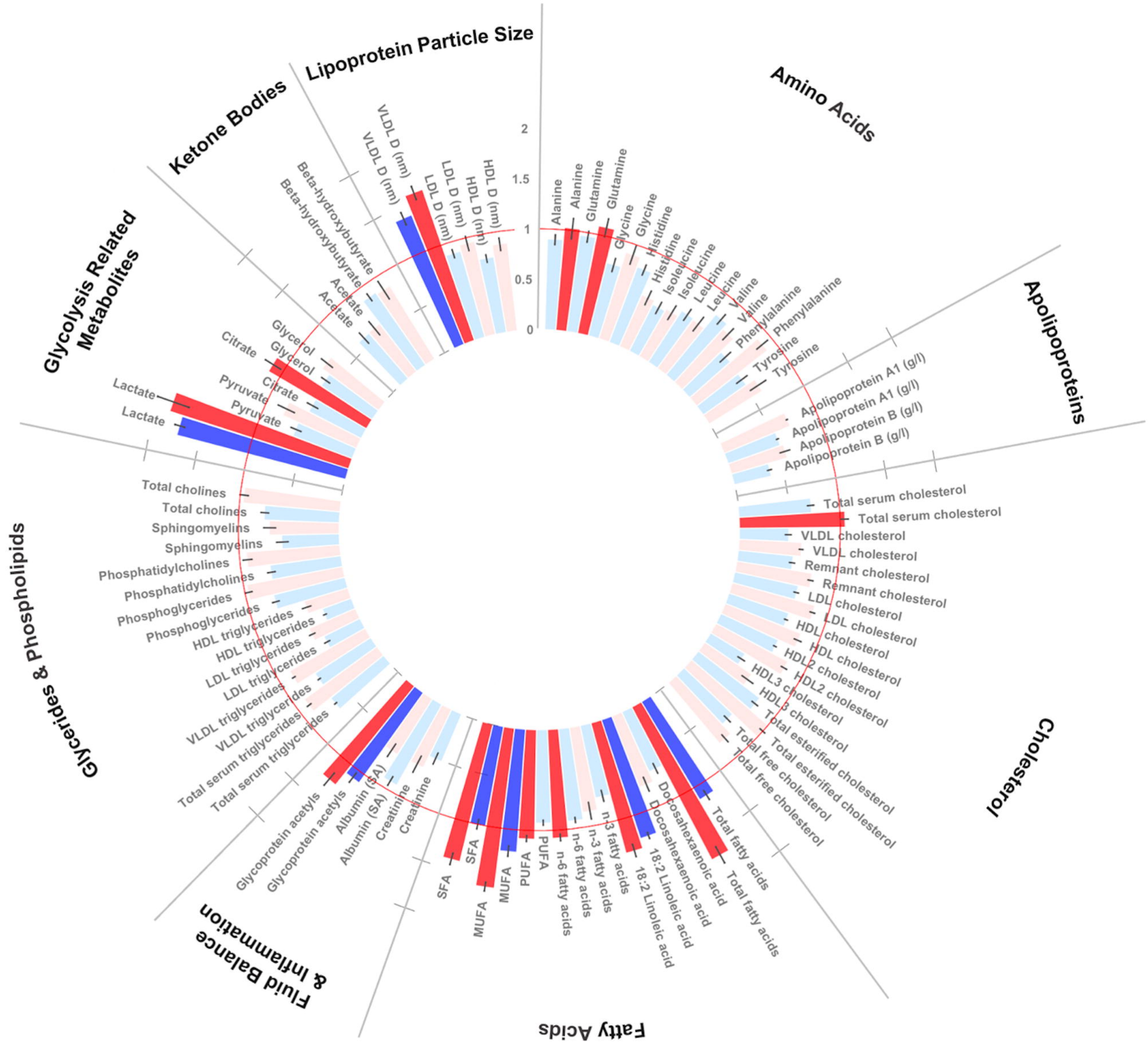
The PLSDA included maternal age, BMI, smoking status, parity, and multiple pregnancy status, and ethnicity. Red line denotes VIP cut-off of 1. SFA = total saturated fatty acids; MUFA = total monounsaturated fatty acids; VLDL_D = diameter of very-low density lipoproteins. GRM: Glycolysis Related Metabolites. 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)
Failed Nightingale © quality control test N= 54
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


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
2
1.5
1
0.5
0
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
Triglycerides
Free cholesterol
Cholesterol esters
Total cholesterol
Phospholipids
Total lipids
Concentration (mol/l)


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