An unbiased lipidomics approach identifies early second trimester lipids predictive of Maternal Glycemic Traits and Gestational Diabetes Mellitus

Liangjian Lu[1], Albert Koulman[2], Clive J. Petry[1], Benjamin Jenkins[2], Lee Matthews[2], Ieuan A. Hughes[1], Carlo L. Acerini[1], Ken K. Ong[1, 3], David B. Dunger[1, 4]

[1]Department of Paediatrics, University of Cambridge, Cambridge, UK

[2]Medical Research Council Human Nutrition Research, Cambridge, UK

[3]Medical Research Council Epidemiology Unit, University of Cambridge, Cambridge, UK

[4]Wellcome Trust-MRC Institute of Metabolic Science, Cambridge, UK

Running title: Early second trimester lipid biomarkers of GDM

Word count: 3997

Tables and Figures: 4

Corresponding author:

Professor David B. Dunger

University Department of Paediatrics

Level 8, Box 116, Addenbrooke's Hospital

Hills Road, Cambridge CB2 0QQ, UK

Tel.: +44 (0) 1223 336886

Fax: +44 (0) 1223 336996

Email: dbd25@cam.ac.uk

Abstract

Objective: To investigate the relationship between early second trimester serum lipidomic variation and maternal glycemic traits at 28 weeks, and to identify predictive lipid biomarkers for Gestational Diabetes (GDM).

Research Design and Methods: Prospective study of 817 pregnant women (Discovery cohort, n=200; Validation cohort, n=617) who provided an early second trimester serum sample, and underwent oral glucose tolerance testing (OGTT) at 28 weeks. In the discovery cohort, lipids were measured using direct infusion mass spectrometry, and correlated with OGTT results. Variable Importance in Projection (VIP) scores were used to identify candidate lipid biomarkers. Candidate biomarkers were measured in the validation cohort using Liquid Chromatography- Mass Spectrometry, and tested for associations with OGTT results and GDM status.

Results: Early second trimester lipidomic variation was associated with 1-hour post-load glucose levels, but not with fasting plasma glucose. Of the 13 lipid species identified by VIP scores, 10 had nominally significant associations with post-load glucose levels. In the validation cohort, 5 of these 10 lipids had significant associations with post-load glucose levels independent of maternal age and BMI, i.e. TG(51:1), TG(48:1), PC(32:1), PCae(40:3) and PCae(40:4). All except the last were also associated with maternal GDM status. Together, these 4 lipid biomarkers had moderate ability to predict GDM (Area under curve (AUC)= 0.71±0.04, $p=4.85x10^{-7}$), and improved the prediction of GDM by age and BMI alone from AUC 0.69 to AUC 0.74.

Conclusions: Specific early second trimester lipid biomarkers can predict maternal GDM status independent of maternal age and BMI, potentially enhancing risk factor-based screening.

Introduction

Gestational diabetes (GDM) affects 9-26% of pregnancies (1). It is clinically important because it increases the risk of obstetric complications (e.g. pre-eclampsia and shoulder dystocia), as well as neonatal complications (e.g. hypoglycemia and hyperbilirubinemia). In the multi-center Hyperglycemia and Adverse Pregnancy Outcomes (HAPO) study, a continuous linear relationship was shown between maternal glucose levels at 24-32 weeks of gestation and the odds of several adverse pregnancy outcomes, even for glucose levels well within the normal range (2). This has provided impetus for the International Association of Diabetes and Pregnancy Study Groups (IADPSG) to recommend universal screening for GDM via the 75g Oral Glucose Tolerance Test (OGTT) between 24-28 weeks gestation (3).

This recommendation by the IADPSG has not been uniformly adopted due to various concerns, including the implications on service provision, and the large number of normoglycemic women who will have undergo what can be an unpleasant and poorly tolerated test (4). Some countries, such as the United Kingdom (4) and Italy (5), perform risk factor-based screening, with only high-risk individuals receiving the diagnostic 75g oral glucose tolerance tests. Specified risk factors include obesity, previous macrosomic baby, previous gestational diabetes, family history of diabetes and minority ethnic family origin with high prevalence of diabetes. In the United States, the American Diabetes Association and American College of Obstetricians and Gynecologists continue to endorse a two-step approach with an initial universal non-fasted 50g Glucose Load Test at 24-28 weeks gestation (6).

Another recognized limitation of screening for GDM at 24-28 weeks gestation is the delay in detecting cases of GDM that developed in the first or early second

trimesters. By the time of screening, significant increased fetal adiposity may have developed (4,7,8). Although performed in a high-risk ethnic minority population, Agarwal and colleagues showed that over 40% of GDM cases could be diagnosed by a 75g OGTT performed before 18 weeks gestation (9). These limitations of universal 24-28 week OGTT support the potential value of predictive early gestational biomarkers for GDM. Biomarkers that have been studied include fasting plasma glucose (9) and more recently maternal metabolites (10,11), plasma proteins (12) and miRNA (13) using high throughput technologies. Although these efforts have not yet informed clinical strategies, the results of metabolomic studies in particular have shed valuable insights on the pathophysiology of GDM (11,14).

Variations in lipid profiles have yet to be comprehensively studied, even though changes in maternal lipid metabolism are well-described from the beginning of pregnancy (15). In early pregnancy, plasma lipids, including triglycerides, phospholipids and cholesterol decrease, before steadily increasing from week 8 onwards (15). The rise in triglycerides is accompanied by an increase in VLDL, LDL and HDL levels (16). Changes in lipids during GDM have also been observed, with GDM women having higher serum triglyceride levels, but lower LDL levels, compared to normoglycemic pregnant women (17).

Therefore, we hypothesized that lipidomic variation in early second trimester maternal serum samples could be associated with later glucose tolerance measured at 28 weeks. Confirmation of this hypothesis would enable identification of candidate lipid biomarkers that are predictive of GDM, so as to improve GDM screening and provide mechanistic insights into the pathophysiology of GDM.

Research Design and Methods

Recruitment and sample collection

The Cambridge Baby Growth Study (CBGS) is a prospective longitudinal study that has been described previously (18–20). Briefly, 2212 women in early pregnancy were recruited between 2001 and 2009 from ultrasound clinics at the Rosie Maternal Hospital, Cambridge, UK. Such dating scans are routinely offered to all pregnant women receiving antenatal care, and are performed at 8-14 weeks gestation.

Shortly after recruitment, at 15.2 ± 0.07 weeks gestation, a non-fasting venous blood sample was collected if women consented. After clotting and within 2 hours of sample collection, these samples were centrifuged at 3,000G for 10 min and the serum separated and stored at -80 ^{o}C. They were maintained at -80 °C until analysis, with the exception of a single freeze-thaw cycle to prepare the necessary aliquots for lipid analysis. A total of 1260 serum samples were collected.

All participants were also invited for a standard 75g OGTT, which was performed at 28 weeks gestation after an overnight fast. A total of 1069 women underwent the OGTT. Plasma glucose levels were analyzed by the standard glucose oxidase method.

Cohort selection

For this study, we excluded OGTT participants who (i) were missing either fasting or 1-hour (1h) post-load venous plasma glucose level measurements (n=10), (ii) subsequently gave birth to twins (n=17), or (iii) did not provide a early second trimester serum sample (n=219). A very small number of participants (n=6) were also excluded for various other reasons, e.g. inadequate remaining serum samples, no

paired DNA sample (for use in other studies). This yielded a total of 817 women, who were assigned to a Discovery Cohort of 200 women, and a Validation Cohort of 617 women. Women in the discovery cohort were selected because they had data on other genetic or phenotypic traits, which other ongoing studies in our group were interested in correlating lipidomic variation with.

There were two differences in clinical characteristics between the discovery and validation cohorts. First, 1h post-load glucose levels were 0.27mM higher in the validation cohort (Table 1). This result was of borderline significance on univariate testing (p=0.05), and non-significant when multiple testing was accounted for using the Benjamini-Hochberg method (p=0.175). Importantly, there was no significant difference in the proportion of cases with GDM in the two groups. Second, samples were taken in the validation cohort at a slightly later gestation, approximately 0.5 weeks later.

Ethical approval

The study protocol was approved by the local research ethics committee, Addenbrooke's Hospital, Cambridge, UK. Written informed consent was obtained from all participants.

Lipid biomarker analyses

In the discovery cohort the lipids were profiled by direct infusion mass spectrometry as described previously (21,22). For biomarker validation, we used a Liquid Chromatography-Mass Spectrometry (LC-MS) method as described before (23).

Statistical analysis

Partial least squares (PLS) regression and PLS- Discriminant analysis (PLS-DA) were used to identify associations between lipidomics variables and OGTT results in the discovery cohort. Fitted models were considered significant if the Q^2, i.e. R^2 of the model as estimated by cross-validation, was positive. The importance of individual lipid species was quantified via the Variable Importance in Projection (VIP) score, and used to identify candidate lipid biomarkers. The VIP score is a widely used method of variable selection. It takes into account the amount of Y-variance explained by the projection, and the loadings of each variable on this projection, while adjusting for the absolute magnitude of each X-variable. As such, 2 variables with identical contribution to the explanatory power of the model will have identical VIP scores, regardless of which component they have a large influence on, or their absolute magnitudes.

Standard linear and logistic regression techniques were used to assess the association between candidate lipid biomarkers and maternal OGTT results or GDM status. GDM was defined based on fasting and 1h post-load glucose levels using IADPSG thresholds, i.e. $\geqslant$5.1 and 10.0mM respectively. 2h post-load data was unavailable for most women and was omitted from our case definition for uniformity. This is acceptable as only 7% of UK women with GDM are diagnosed based on the 2h measurement alone (1).

Logistic regression was used to combine the predictive ability of candidate lipid biomarkers, which was then assessed using Receiver Operating Characteristics (ROC) plots. Where backward stepwise selection was used, a significance threshold of 0.10 for removal was employed. Linear discriminant analysis was used to ensure the robustness of these results.

The threshold for statistical significance was 0.05. For the discovery cohort uncorrected p-values were considered, whereas in the validation cohort p-values were corrected for multiple testing using the Benjamini-Hochberg method. Values in the text are given as mean ± SE unless otherwise specified. Regression coefficients were standardized by the predictor, i.e. change in response variable for each standard deviation increase in the predictor.

PLS and PLS-DA regression was performed using SIMCA version 14 (MKS Umetrics AB, Umeå, Sweden). All other analyses were performed using SPSS version 21 (IBM, Armonk, NY, USA).

Results

Association between early second trimester lipidomic variation and 28-week glucose tolerance

Analysis of the discovery cohort was confined to 196 samples as 4 samples were found to be unsuitable for analysis due to hemolysis. In these samples, 189 lipid species were detected.

A PLS model was constructed to examine the extent to which early second trimester lipidomic variation explained post-load venous glucose levels during a OGTT at 28 weeks. The resulting model yielded 1 fitted component, which used 15% of lipidomic variation (R^2X) to explain 11% (R^2Y) of variation in post-load glucose levels. This was robust to internal cross-validation, yielding a Q^2 of 4%.

Because some lipid species may show non-linear relationships with post-load glucose levels, we divided participants into tertiles of OGTT levels, and constructed a PLS-DA model to explain membership in the top tertile. The resulting model also yielded a single component, which used 15% of lipidomic variation to explain 9% of the variation in top tertile membership, with a Q^2 of 1.31%.

Similar PLS and PLS-DA models were also constructed to explore the relationship between lipidomic variation and fasting plasma glucose levels. However, the models were overfitted, with the PLS and PLS-DA models yielding a Q^2 of -3% (R^2X=15%, R^2Y=9%) and -10% (R^2X=9%, R^2Y=8%) respectively. These models were not used in subsequent analyses.

Identification of lipid biomarkers of post-load plasma glucose levels

From each of the two models considering post-load glucose levels, approximately 10 lipids with the highest VIP scores were selected, with the exact cut-point selected using a graphical method (Fig. 1a, b). The PLS model yielded 9 lipid species and the PLS-DA model 10 species, with 5 lipid species being identified in both models. One of the species annotated as DG-H2O(32:0) was likely to be an in-source fragment of a different lipid species. As in-source fragments are artifacts of mass spectrometry, this species was disregarded, leaving a total of 13 lipid species identified.

The 13 lipid species were regressed against post-load glucose levels or membership in the top tertile thereof. To ensure that they were not simply surrogates for known risk factors of GDM, maternal age and pre-pregnancy BMI were adjusted for. This resulted in 10 of the 13 lipid species having a nominally significant association with post-load glucose levels and/or the top tertile thereof (Supplementary Table S1). The 2 lipid species showing the strongest association with post-load glucose levels were the triglyceride TG(51:1) (0.40mM per SD increase, p=8.88E-4) and the choline ether phospholipid PCae(40:3) (-0.41mM per SD increase, p=9.73E-4) (Supplementary Table S2).

Further examination of the correlations between these 10 lipid species revealed clustering into 2 large groups (Fig. 1c). The first group contained the choline ether phospholipids, whereas the second group contained triglycerides and a phosphatidylcholines.

Validation of lipid biomarkers of post-load plasma glucose levels

These 10 candidate lipid biomarkers were measured in the validation set of 617 subjects using a LC-MS method. This provided additional chromatographic information, eliminating any possible artifacts introduced by the shotgun approach in

the discovery set. With this method we were unable to detect PCae(44:4) in our samples, suggesting that this putative signal in the discovery set was an artifact, and it was omitted from further analysis. We also omitted 20 samples in which none of the remaining 9 lipid species were detectable, as this was likely indicative of poor sample quality. This left data on the 9 candidate lipid biomarkers in 597 subjects for analysis.

Of the 9 remaining lipid species, 5 showed significant associations with post-load glucose levels even after adjustment for maternal age and BMI and correction for multiple testing (Table 2). TG(51:1), TG(48:1) and PC(32:1) were positively associated with maternal post-load glucose levels, whereas PCae(40:3) and PCae(40:4) were inversely associated with maternal post-load glucose levels. Logistic regression against membership in the top tertile of post-load glucose levels did not validate any additional candidate lipid species, nor did adjusting for gestational age at the time of serum sample collection (data not shown).

As fasting and post-load glucose levels have common pathophysiological determinants and were moderately correlated in our cohort ($r=0.343$, $p<0.001$), we tested the 9 putative lipid biomarkers for an association with fasting glucose levels (data not shown). TG(51:1) and PCae(40:4) showed significant associations with fasting glucose levels, even after adjusting for multiple testing. However, after adjusting for maternal age and BMI, only TG(51:1) remained significant (0.06mM per SD increase, p-value=0.003, Benjamini-Hochberg p-value=0.02).

Lipid predictors of Gestational Diabetes

Of the 597 subjects in the validation cohort, 53 met the criteria for GDM. The 5 validated lipid species were tested for association with GDM. TG(51:1), TG(48:1),

PC(32:1), and PCae(40:4) were significantly associated with GDM, even after adjustment for maternal age and maternal BMI, and correction for multiple testing (Fig. 2a). While not reaching significance, PCae(40:3) nonetheless demonstrated a strong trend in the same direction as PCae(40:4) (Benjamini-Hochberg p-value=0.07).

To assess the combined predictive ability of these 4 lipid species, logistic regression was used to calculate the probability of GDM status of each subject, and the probability scores used to construct a ROC curve (Fig. 2b). This yielded an AUC of 0.709 ± 0.040 (p=4.85E-7). Similar results were obtained using linear discriminant analysis.

Of the 597 subjects, 410 (including 37 GDM cases) had available data on maternal age and BMI. This enabled us to assess the additional predictive power conferred by these 4 lipid biomarkers over maternal age and BMI alone. In this sub-group, maternal age and BMI produced an AUC of 0.689 ± 0.046 (p=1.54E-4), and further inclusion of the 4 lipid biomarkers increased the AUC to 0.741 ± 0.045 (p=1.33E-6) (Fig. 2c). Graphically, the improvement in AUC was most marked at stringent thresholds, i.e. enhancing the sensitivity at high levels of specificity. For instance at 91.7% specificity, sensitivity is 21.6% based on traditional risk factors, but raised to 48.6% when lipid predictors are included. In addition, there was some improvement at high levels of sensitivity, for instance at 97.3% sensitivity, where the inclusion of lipid predictors raised specificity from 9.9% based on traditional risk factors alone to 24.9%.

Finally, we sought to identify the most parsimonious model from these 6 potential predictive variables. Using a backward stepwise selection algorithm, the only terms

left in the model were Maternal BMI, TG(48:1) and PCae(40:4). This yielded an AUC of 0.732 ± 0.045 (p=3.28E-6), which is similar to the model including all 6 predictors.

Conclusions

In this unselected cohort of predominantly Caucasian pregnant women, using an unbiased lipidomics approach and a pre-assigned validation cohort, we show for the first time that specific lipid species in the maternal early second trimester lipid profile are associated with maternal glycemic traits assessed by standard 75g oral glucose tolerance testing at 28 weeks. We identified 4 lipid biomarkers, i.e. TG(51:1), TG(48:1), PC(32:1) and PCae(40:4), that predict later GDM independent of maternal age and BMI, and could potentially enhance the performance of existing risk-factor based screening approaches used in many countries.

The performance of clinical risk-factor based screening has been examined in many different populations. A recent study in an Australian population compared the performance of the NICE, ADA and Australasian Diabetes in Pregnancy Society risk-factor based screening guidelines, yielding sensitivities of 92%, 100% and 99%, and specificities of 32.4%, 3.9% and 13.7% respectively (24). This and other studies reveal limited test performance, with the need for low levels of specificity to achieve the high levels of sensitivity.

The 4 lipid biomarkers that we identified have moderate predictive performance, with an estimated AUC of 0.709. This is comparable to other early pregnancy biomarkers, including conventional biomarkers fasting plasma glucose (estimated AUC=0.579) (9), HbA_{1C} (AUC in high risk population=0.67) (25), triglycerides (AUC=0.55-0.61) and triglycerides to HDL ratio (AUC=0.62) (26), as well as novel biomarkers such as second trimester serum miRNA (AUC=0.669) (13). Furthermore the lipid biomarkers that we derived were specifically identified to predict GDM independent of maternal age and BMI, and thus can enhance the predictive performance of existing risk

factor-based approaches. Indeed, inclusion of the 4 lipid biomarkers to maternal age and BMI increased the AUC from 0.689 to 0.741.

The enhancement of sensitivity at high levels of specificity was particularly marked. Although for the purpose of predicting GDM, our data on the clinical performance of lipid biomarkers must be considered very preliminary, lipid biomarkers may potentially have a role in identifying high-risk women who should receive an immediate/early second trimester OGTT.

The lipid biomarkers we identified can be divided into two groups. TG(51:1), TG(48:1) and PC(32:1) are associated with increased post-load glucose levels and GDM risk, and are moderately correlated. The choline ether phospholipids PCae(40:3) and PCae(40:4) are associated with decreased post-load glucose levels and/or GDM risk, and are strongly correlated.

The association of TG(51:1), TG(48:1) and PC(32:1) with maternal glucose levels is consistent with previous investigations into lipidomic changes associated with Type 2 diabetes. TG(48:1) has been implicated with Type 2 diabetes risk in the Framingham cohort (27), and PC(32:1) levels are raised in AusDiab subjects with Type 2 Diabetes (28). These 3 lipid species are notable for the presence of a single double bond, which implies the presence of a monounsaturated fatty acid (MUFA), predominantly palmitoleate and to a lesser extent, oleate, on closer inspection of LC-MS spectra in our analysis. In one study, palmitoleate content within circulating phospholipids was found to be associated with increased insulin resistance (29). As circulating palmitoleate is principally synthesized in the liver in humans, this may reflect hepatic insulin resistance, in line with our finding that TG(51:1) is associated with fasting plasma glucose levels (30). Mechanistically, palmitoleate and oleate are produced

from palmitate and stearate by the action of Steroyl-CoA desaturase 1 (SCD1), and SCD1 activity has recently been linked in a large cohort to Type 2 Diabetes risk and hepatic steatosis (31). Indeed, $SCD1^{-/-}$ mice display increased insulin sensitivity (32).

The association of TG(51:1), primarily comprising TG(18:1/17:0/16:0) (or positional isomers thereof) in our cohort, with maternal glucose levels was also surprising because odd-chain fatty acids, including heptadecanoic acid, have been associated with reduced risk of Type 2 Diabetes (28,33,34). However, our finding is in keeping with results of an untargeted metabolomic screen using fasted serum samples from women at 28 weeks gestation who were enrolled in the HAPO study (35), in which heptadecanoic acid was raised in subjects with fasting plasma glucose levels in the 90^{th} percentile but with similar BMI to controls. This may be due to genuine differences in the pathophysiology of GDM and Type 2 Diabetes, but may also reflect the fact that the studies of Type 2 Diabetes measured the fatty acid content in phospholipids, whereas the latter study of pregnant women measured free fatty acids (35). This underscores the advantage of intact lipid studies as opposed to fatty-acid profiling (30), which is a strength of our study.

The other group of lipid biomarkers identified, i.e. the choline ether phospholipids PCae(40:3) and PCae(40:4), were inversely associated with maternal glucose levels. This is consistent with an earlier report from the AusDiab cohort in which ether phospholipids were inversely related with post-load glucose levels and reduced in patients with diabetes (28). The physiological function of ether phospholipids remains largely unknown (36), but they have been implicated as physiological ligands of PPARγ (37). Intriguingly, SCD1 is a target of PPARγ, potentially providing a mechanistic link between low ether phospholipids, and high levels of palmitoleate- and oleate- containing lipids (32).

The broad overlap between our lipid biomarkers and those identified in studies of Type 2 Diabetes may be due to the fact that lipid profiles were derived from samples obtained early in pregnancy, reflecting the contribution of pre-existing insulin resistance to the development of Gestational Diabetes. Indeed, it will be ideal to obtain a pre-conceptional sample as well as one during pregnancy, to identify lipid biomarkers which reflect the pathophysiological contribution of pregnancy itself. Nevertheless, this explanation of our findings is made less likely by the fact that GWAS studies have revealed a broadly shared genetic architecture between GDM and Type 2 Diabetes, and metabolomic studies from later in pregnancy, including one using samples from 28 weeks gestation, have shown overlapping metabolic signatures between GDM and Type 2 diabetes (11,14,35).

There are several limitations to our study. First, while we validated our candidate lipid biomarkers using a pre-defined subset, these biomarkers have not been externally validated, for example in populations of high-risk ethnicities. Second, our study was not designed to demonstrate the superiority of a lipid biomarker and risk-factor based approach compared to a risk-factor based screening alone. Thus, we lacked data on other conventional risk factors, e.g. family history of diabetes and personal history of GDM. For similar reasons, we also lacked data on other traditional biochemical risk factors, such as HbA_{1C}, triglycerides and HDL cholesterol, and are thus unable to directly compare the performance of the lipid biomarkers to these alternatives within this study. Third, because we selected only 13 lipids from the 189 lipids measured for univariate analysis, we might have been overly conservative in our approach. Finally, because serum samples for lipidomic analysis were obtained in the non-fasting state without controlling for meal time and meal content, this would have added additional lipidomic variability that was not related to

variation in OGTT results at 28 weeks, thus reducing study power. However, this additional variability might be small compared to existing inter- and intra-subject variation (38). Our re-analysis of data from Begum and colleagues (38) suggests that variance is partitioned between inter-subject differences, intra-subject differences not due to meal time, and the effect of meal time in the proportion 62%, 31% and 7%, albeit the population that they studied was less heterogeneous than the CBGS.

In summary, we report for the first time an association between maternal early second trimester lipid species and glycemic traits at 28 weeks, as assessed by a standard OGTT. We further show that 4 lipid biomarkers, TG(51:1), TG(48:1), PC(32:1) and PCae(40:4) are able to predict maternal GDM status independent of maternal age and BMI, and have potential to improve the performance of clinical risk-factor based screening. The lipid biomarkers identified also revealed marked similarities between the pathophysiology of GDM and Type 2 Diabetes (14). In particular, we highlight the established role of MUFAs (especially palmitoleate) and the emerging role of ether phospholipids, as well as the potential pathological role of odd-chain fatty acids, which might indicate a divergence in the pathophysiologies of GDM and Type 2 Diabetes.

Acknowledgements

L.L., A.K., C.J.P. and D.B.D. designed the study, interpreted the data and edited the manuscript. A.K., B.J. and L.M. collected the data. L.L. analyzed the data and drafted the manuscript. I.A.H., C.L.A. and K.K.O. contributed to discussion and reviewed the manuscript. L.L and D.B.D. are the guarantors of this work and, as such, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

The authors acknowledge the excellent laboratory assistance provided by Dianne Wingate, Rachel Seear, Katrin Mooslehner and Radka Platte. The authors would like to express their appreciation to the families who participated in this study, and acknowledge the important contributions made by the research nurses, including Suzanne Smith, Ann-Marie Wardell and Karen Forbes, as well as the assistance provided by colleagues at the Addenbrooke's Wellcome Trust Clinical Research Facility and midwives at the Rosie Maternity Hospital in data collection.

This part of the Cambridge Baby Growth Study was funded by grants from the Wellbeing of Women (RG1644) and Diabetes UK (11/0004241). The lipidomics assays were supported by the Medical Research Council (UD99999906) and Cambridge Lipidomics Biomarker Research Initiative (G0800783). Core funding was also obtained through the Medical Research Council, European Union Framework 5 World Cancer Research Fund, Mothercare Foundation and the Newlife Foundation for Disabled Children. There has also been support from National Institute for Health Research Cambridge Biomedical Research Centre.

There were no potential conflicts of interests reported. This work has not been published elsewhere, either in whole or in part.

Figure legends

Figure 1. Candidate lipid biomarkers identified from the Discovery Cohort. (a-b) Variable Importance in Projection scores of individual lipid species in a PLS model mapping early 2nd trimester lipid profiles to late 2nd trimester 1-hour post-load glucose levels (b) Variable Importance in Projection Scores of individual lipid species in a PLS-DA model mapping early 2nd trimester lipid profiles to membership in the top tertile of late 2nd trimester 1-hour post-load glucose levels (c) Correlation between 10 candidate lipid biomarkers taken forward to validation cohort. PCae: Choline ether phospholipid; TG: Triglyceride; PE: Phosphatidylethanolamine; PC: Phosphatidylcholine. Underlined lipid species refer to species which had nominally significant associations with 1-hour post-load glucose levels (a) or the top tertile thereof (b), and were taken forward to the Validation cohort.

Figure 2. Validated lipid biomarkers and Gestational Diabetes prediction within the Validation Cohort. (a) Individual predictive power of each lipid, independent of maternal age and BMI. * refers to $P<0.05$, with Benjamini-Hochberg correction for multiple testing (b) Combined predictive power of 4 lipid species (c) Enhancement of predictive power of conventional risk factors. AUC: Area under the curve. Other abbreviations as per Figure 1.

References

1. Sacks DA, Hadden DR, Maresh M, Deerochanawong C, Dyer AR, Metzger BE, et al. Frequency of gestational diabetes mellitus at collaborating centers based on IADPSG consensus panel-recommended criteria: the Hyperglycemia and Adverse Pregnancy Outcome (HAPO) Study. Diabetes Care. 2012 Mar;35(3):526–8.

2. Metzger BE, Lowe LP, Dyer AR, Trimble ER, Chaovarindr U, Coustan DR, et al. Hyperglycemia and adverse pregnancy outcomes. N Engl J Med. 2008 May 8;358(19):1991–2002.

3. Metzger BE, Gabbe SG, Persson B, Buchanan TA, Catalano PA, Damm P, et al. International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. Diabetes Care. 2010 Mar;33(3):676–82.

4. National Collaborating Centre for Women's and Children's Health. Diabetes in Pregnancy: Management of Diabetes and Its Complications from Preconception to the Postnatal Period. London; 2015. Report No.: 3.

5. Lacaria E, Lencioni C, Russo L, Romano M, Lemmi P, Battini L, et al. Selective screening for GDM in Italy: application and effectiveness of National Guidelines. J Matern Fetal Neonatal Med. 2015 Jan;28(15):1842–4.

6. American Diabetes Association. Classification and Diagnosis of Diabetes. Diabetes Care. 2015 Dec 23;38(Supplement_1):S8–S16.

7. Schaefer-Graf UM, Kjos SL, Kilavuz O, Plagemann A, Brauer M, Dudenhausen JW, et al. Determinants of fetal growth at different periods of pregnancies complicated by gestational diabetes mellitus or impaired glucose tolerance. Diabetes Care. 2003 Jan;26(1):193–8.

8. Riskin-Mashiah S, Younes G, Damti A, Auslender R. First-trimester fasting hyperglycemia and adverse pregnancy outcomes. Diabetes Care. 2009 Sep;32(9):1639–43.

9. Agarwal MM, Dhatt GS, Punnose J, Zayed R. Gestational diabetes: fasting and postprandial glucose as first prenatal screening tests in a high-risk population. J Reprod Med. 2007 Apr;52(4):299–305.

10. Sachse D, Sletner L, Mørkrid K, Jenum AK, Birkeland KI, Rise F, et al. Metabolic changes in urine during and after pregnancy in a large, multiethnic population-based cohort study of gestational diabetes. PLoS One. 2012 Jan;7(12):e52399.

11. Lowe WL, Karban J. Genetics, genomics and metabolomics: new insights into maternal metabolism during pregnancy. Diabet Med. 2014 Mar;31(3):254–62.

12. Zhao C, Wang F, Wang P, Ding H, Huang X, Shi Z. Early second-trimester plasma protein profiling using multiplexed isobaric tandem mass tag (TMT) labeling predicts gestational diabetes mellitus. Acta Diabetol. Springer Milan; 2015 Aug 11;

13. Zhao C, Dong J, Jiang T, Shi Z, Yu B, Zhu Y, et al. Early second-trimester serum miRNA profiling predicts gestational diabetes mellitus. PLoS One. 2011 Jan;6(8):e23925.

14. Angueira AR, Ludvik AE, Reddy TE, Wicksteed B, Lowe WL, Layden BT. New insights into gestational glucose metabolism: lessons learned from 21st century approaches. Diabetes. 2015 Feb;64(2):327–34.

15. Butte NF. Carbohydrate and lipid metabolism in pregnancy: normal compared with gestational diabetes mellitus. Am J Clin Nutr. 2000 May;71(5 Suppl):1256S–61S.

16. Di Cianni G, Miccoli R, Volpe L, Lencioni C, Del Prato S. Intermediate metabolism in normal pregnancy and in gestational diabetes. Diabetes Metab Res Rev. 2003;19(4):259–70.

17. Koukkou E, Watts GF, Lowy C. Serum lipid, lipoprotein and apolipoprotein changes in gestational diabetes mellitus: a cross-sectional and prospective study. J Clin Pathol. 1996 Aug;49(8):634–7.

18. Petry CJ, Seear R V, Wingate DL, Acerini CL, Ong KK, Hughes IA, et al. Maternally transmitted foetal H19 variants and associations with birth weight. Hum Genet. 2011 Nov;130(5):663–70.

19. Petry CJ, Seear R V, Wingate DL, Manico L, Acerini CL, Ong KK, et al. Associations between paternally transmitted fetal IGF2 variants and maternal circulating glucose concentrations in pregnancy. Diabetes. 2011 Nov;60(11):3090–6.

20. Prentice P, Acerini CL, Eleftheriou A, Hughes I a, Ong KK, Dunger DB. Cohort Profile: the Cambridge Baby Growth Study (CBGS). Int J Epidemiol. 2016 Feb 31;45(1):35–35g.

21. Koulman A, Prentice P, Wong MCY, Matthews L, Bond NJ, Eiden M, et al. The development and validation of a fast and robust dried blood spot based lipid profiling method to study infant metabolism. Metabolomics. 2014 Jan;10(5):1018–25.

22. Eiden M, Koulman A, Hatunic M, West JA, Murfitt S, Osei M, et al. Mechanistic insights revealed by lipid profiling in monogenic insulin resistance syndromes. Genome Med. 2015 Jan;7(1):63.

23. Koulman A, Woffendin G, Narayana VK, Welchman H, Crone C, Volmer DA. High-resolution extracted ion chromatography, a new tool for metabolomics

and lipidomics using a second-generation orbitrap mass spectrometer. Rapid Commun Mass Spectrom. 2009 May;23(10):1411–8.

24. Teh WT, Teede HJ, Paul E, Harrison CL, Wallace EM, Allan C. Risk factors for gestational diabetes mellitus: implications for the application of screening guidelines. Aust N Z J Obstet Gynaecol. 2011 Feb;51(1):26–30.

25. Amylidi S, Mosimann B, Stettler C, Fiedler GM, Surbek D, Raio L. First-trimester glycosylated hemoglobin in women at high risk for gestational diabetes. Acta Obstet Gynecol Scand. 2016 Jan;95(1):93–7.

26. Wang C, Zhu W, Wei Y, Su R, Feng H, Lin L, et al. The Predictive Effects of Early Pregnancy Lipid Profiles and Fasting Glucose on the Risk of Gestational Diabetes Mellitus Stratified by Body Mass Index. J Diabetes Res. 2016 Jan;2016:3013567.

27. Rhee EP, Cheng S, Larson MG, Walford GA, Lewis GD, McCabe E, et al. Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. J Clin Invest. 2011 Apr;121(4):1402–11.

28. Meikle PJ, Wong G, Barlow CK, Weir JM, Greeve MA, Macintosh GL, et al. Plasma lipid profiling shows similar associations with prediabetes and type 2 diabetes. PLoS One. 2013 Jan;8(9):e74341.

29. Mozaffarian D, Cao H, King IB, Lemaitre RN, Song X, Siscovick DS, et al. Circulating palmitoleic acid and risk of metabolic abnormalities and new-onset diabetes. Am J Clin Nutr. 2010 Dec;92(6):1350–8.

30. Roberts LD, Koulman A, Griffin JL. Towards metabolic biomarkers of insulin resistance and type 2 diabetes: progress from the metabolome. Lancet Diabetes Endocrinol. 2014 Jan;2(1):65–75.

31. Jacobs S, Schiller K, Jansen EHJM, Boeing H, Schulze MB, Kröger J. Evaluation of various biomarkers as potential mediators of the association between Δ5 desaturase, Δ6 desaturase, and stearoyl-CoA desaturase activity and incident type 2 diabetes in the European Prospective Investigation into Cancer and Nutrition-Potsdam . Am J Clin Nutr. 2015 Jul;102(1):155–64.

32. Paton CM, Ntambi JM. Biochemical and physiological function of stearoyl-CoA desaturase. Am J Physiol Endocrinol Metab. 2009 Jul;297(1):E28–37.

33. Forouhi NG, Koulman A, Sharp SJ, Imamura F, Kröger J, Schulze MB, et al. Differences in the prospective association between individual plasma phospholipid saturated fatty acids and incident type 2 diabetes: the EPIC-InterAct case-cohort study. Lancet Diabetes Endocrinol. 2014 Oct;2(10):810–8.

34. Jenkins B, West JA, Koulman A. A review of odd-chain fatty acid metabolism and the role of pentadecanoic Acid (c15:0) and heptadecanoic Acid (c17:0) in health and disease. Molecules. 2015 Jan;20(2):2425–44.

35. Scholtens DM, Muehlbauer MJ, Daya NR, Stevens RD, Dyer AR, Lowe LP, et al. Metabolomics reveals broad-scale metabolic perturbations in hyperglycemic mothers during pregnancy. Diabetes Care. 2014 Jan;37(1):158–66.

36. Lodhi IJ, Semenkovich CF. Peroxisomes: a nexus for lipid metabolism and cellular signaling. Cell Metab. Elsevier Inc.; 2014 Mar 4;19(3):380–92.

37. Lodhi IJ, Yin L, Jensen-Urstad APL, Funai K, Coleman T, Baird JH, et al. Inhibiting adipose tissue lipogenesis reprograms thermogenesis and PPARγ activation to decrease diet-induced obesity. Cell Metab. 2012 Aug 8;16(2):189–201.

38. Begum H, Li B, Shui G, Cazenave-Gassiot A, Soong R, Ong RT-H, et al. Discovering and validating between-subject variations in plasma lipids in healthy subjects. Sci Rep. Nature Publishing Group; 2016 Jan;6:19139.

Tables

Clinical Characteristics	All OGTT		Lipidomics Total		Discovery cohort		Validation cohort		Discovery VS Validation
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	p-value
Total number, n	1069		817		200		617		
Age (Years)	33.39	4.19	33.27	4.11	33.38	3.83	33.22	4.21	0.65
Maternal pre-pregnancy BMI (kg/m^2)	24.10	4.46	24.19	4.48	23.90	4.13	24.30	4.60	0.31
Maternal height (m)	1.66	0.07	1.66	0.07	1.66	0.07	1.66	0.07	0.22
Fasting glucose levels (mM)	4.33	0.55	4.33	0.49	4.33	0.53	4.32	0.47	0.93
1h post-load glucose levels (mM)	6.83	1.72	6.77	1.67	6.57	1.59	6.84	1.69	0.05
Gestational Diabetes (%)	9.45		8.20		6.50		8.75		0.38
Gestational Age at serum sample collection	15.2	2.46	15.0	2.02	14.6	1.73	15.1	2.10	0.004

Table 1. Clinical characteristics of participants in the Lipidomics Study. Fasting and post-load glucose levels were measured using a standard oral glucose tolerance test at 28 weeks of pregnancy. Gestational diabetes was diagnosed based on fasting and 1h post-load venous plasma glucose measurements (details in text). Comparisons were performed using Student's t-test and χ^2 for the proportion of participants with gestational diabetes. P-values were not corrected for multiple testing.

Lipid Species	Univariate Analysis			Adjusted for Maternal age and BMI		
	Regression coefficient (mM per SD)	p-value	BH p-value	Regression coefficient (mM per SD)	p-value	BH p-value
TG(51:1)	0.18	**9.15E-03**	**0.02**	0.18	**0.03**	**0.05**
TG(48:1)	0.27	**1.00E-04**	**3.82E-04**	0.22	**7.40E-03**	**0.02**
TG(50:1)	0.07	0.31	0.40	-0.06	0.45	0.50
PC(32:1)	0.33	**1.45E-06**	**1.31E-05**	0.21	**8.94E-03**	**0.02**
PCae(38:4)	-0.06	0.39	0.44	-0.15	0.06	0.09
PCae(44:6)	0.02	0.80	0.80	-0.03	0.69	0.69
PCae(40:3)	-0.26	**1.70E-04**	**3.82E-04**	-0.24	**3.63E-03**	**0.02**
PCae(40:5)	-0.16	**0.02**	**0.03**	-0.14	0.08	0.10
PCae(40:4)	-0.27	**1.31E-04**	**3.82E-04**	-0.29	**2.60E-04**	**2.34E-03**

Table 2. Relationship between candidate lipid biomarkers and 2nd trimester 1-hour post-load glucose levels in the Validation Cohort. BH: Benjamini-Hochberg corrected. Abbreviations as per Figure 1. P-values in bold type indicate a statistically significant result (P<0.05).