Metabolomic biomarkers and novel dietary factors associated with gestational diabetes in China

Xuyang Chen ^{a,b#}, Jamie V. de Seymour ^{c#}, Ting-Li Han ^{a,c#}, Yinyin Xia ^d, Chang Chen ^{a,e}, Ting Zhang ^{a,b}, Hua Zhang ^{a,b*}, Philip N. Baker ^{a,f}

^a Canada - China -New Zealand Joint Laboratory of Maternal and Fetal Medicine, Chongqing Medical University, Chongqing 400016, China

^b Department of Obstetrics and Gynaecology, The First Affiliated Hospital of Chongqing Medical University, Chongqing 400016, China

^c Liggins Institute, The University of Auckland, Auckland 1023, New Zealand.

^d School of Public Health and Management, Chongqing Medical University, Chongqing 400016, China

^e Institute of Life Sciences, Chongqing Medical University, Chongqing 400016, China

^f College of Medicine, Biological Sciences and Psychology, University of Leicester, UK

Joint first authors

* **Correspondence:** Professor Hua Zhang, Email: zh2844@gmail.com, Phone: +86-23-89011102, Address: Department of Obstetrics and Gynaecology, The First Affiliated Hospital of Chongqing Medical University, No. 1 Youyi Road, Yuzhong District, Chongqing 400016, People's Republic of China

Abstract

Introduction: Gestational diabetes mellitus (GDM) is impaired glucose tolerance first recognised during pregnancy; its development is associated with many adverse outcomes. Mechanisms of GDM development are not fully elucidated and few studies have used Chinese participants.

Objective: The aim of this study was to investigate the maternal metabolome associated with GDM in a Chinese population, and explore the relationship with maternal diet.

Methods: Ninety-three participants were recruited at 26-28 weeks' gestation from Chongqing, China. Maternal urine, serum, and hair metabolomes were analysed using gas and liquid chromatography-mass spectrometry. Dietary intake was assessed using a 96-item food frequency questionnaire.

Results: Of the 1064 metabolites identified, 73 were significantly different between cases and controls ($P < 0.05$), but only 2-aminobutyric acid had both a p- and q-value < 0.05 . A

“snack-based-dietary-pattern” was associated with an increased likelihood of GDM (odds ratio: 2.1; 95% confidence interval: 1.1, 3.9). The association remained significant after adjustment for calorie intake but not food volume.

Conclusion: This study provides a comprehensive characterization of the maternal metabolome. The snack-based dietary pattern associated with GDM suggests that timing and frequency of consumption are important factors in the relationship between maternal diet and GDM.

Key words: gestational diabetes; metabolomics; maternal diet; biomarker

Introduction

Gestational diabetes mellitus (GDM) is a pregnancy complication characterized by the onset of impaired glucose tolerance during pregnancy, in the absence of overt diabetes (Reece et al., 2009). The prevalence of GDM has increased over the last 20 years (Ferrara, 2007) and is currently one of the most common pregnancy complications, affecting an estimated one in seven births worldwide (Federation, 2015). China has some of the highest rates of diabetes, with an estimated 110 million adults diagnosed with diabetes, and up to 20% of all pregnancies affected by GDM (Zhu et al., 2017). The development of GDM is associated with serious short-term and long-term health complications for both the mother and her offspring (Kim et al., 2002), but the pathogenesis of GDM is not well understood. Risk factors for GDM development include a high BMI, advanced age, previous history of GDM, previous delivery to a large for gestational age infant, and a family history of diabetes. Despite the application of these clinical risk factors in early pregnancy screening, not all women that develop GDM fit into these high-risk categories. A study conducted in an Asian cohort found that clinical risk factors were inadequate for detecting almost 50% of women that went on to develop GDM, particularly in Chinese women (Chong et al., 2014): These findings emphasise a profound need for further research to understand the pathophysiology underlying GDM development in Chinese women. Maternal diet has been shown to affect pregnancy outcome for mother and baby, and has been associated with numerous pregnancy complications, including GDM (Chen et al., 2016). Studies of the relationship between diet and GDM have predominantly involved a reductionist approach, whereby observational dietary intake is analysed and reported as associations between individual nutrients and GDM. Recent appreciation that nutrients and foods are not often consumed in isolation, has led to a shift in paradigm to analysing dietary patterns (Chen et al., 2016). Dietary pattern analysis is an approach to dietary investigations that involves analysing the sum of foods and beverages consumed over a given period, and therefore takes into consideration nutrients consumed in combination. A majority of the dietary pattern studies related to GDM however, have been conducted in the United States of America, a Western population with significantly different dietary intake, lifestyle, and genetic make-up to women residing in China. Self-reported dietary assessment using food frequency questionnaires, food records, and dietary recalls are also prone to bias resulting from social desirability and inaccuracy in dietary recall. In free-living subjects this is a difficult barrier to address, and developments into an objective but comprehensive measure of dietary intake are warranted. Dietary

biomarker investigations are one way that self-reported dietary intake can be validated.

Metabolomics is a powerful strategy for studying the low molecular weight compounds, known as metabolites, in a biological system (Nicholson et al., 1999). The metabolome is the comprehensive set of detectable metabolites in the biological specimen or tissue under investigation. The metabolome reflects the internal physiological status of an organism as well as changes in response to environmental stimuli, including dietary consumption. Metabolite profiling is a potential tool that can be used for identifying biomarkers of GDM and elucidating metabolic mechanisms associated with its development. Most of the studies reported recently have focused on the maternal biological samples in the second trimester of pregnancy. Blood and urine have been the most frequently used specimens in the studies of GDM. These studies found that metabolites such as 3-hydroxyisovalerate, 2-hydroxyisobutyrate, amino acids or derivatives, and carnitine in maternal urine were associated with GDM. However, there was no common biomarker candidate found in the studies of the maternal urine and only a few common potential biomarker candidates found in blood specimens, such as triglycerides, urea, 2-hydroxybutyrate, and 3-hydroxybutyrate. In addition to bio-fluids, studies have employed hair, amniotic fluid, and placenta specimens to study the metabolome in GDM. Studies of the metabolome related to GDM are summarised in Table S1. Despite initial findings that offered proof of concept that the maternal metabolome can be used to differentiate women that develop GDM from those who do not, the findings lack consistency, and are yet to produce any clinically useful, validated biomarkers of GDM. In addition, no study to date has explored a range of specimens to gather a comprehensive insight into how the maternal metabolome is affected by the development of GDM.

The metabolome is influenced by dietary intake and as such, metabolomic analysis could be used for identifying objective dietary biomarkers of intake (O'Gorman et al., 2013). Objective markers of intake could significantly improve dietary assessment accuracy and could be used in future analyses to link disease to diet, offering intervention avenues for further exploration. Although both the maternal metabolome and maternal diet have been separately investigated for associations with GDM previously, no study has combined both metabolomic and dietary analyses to understand how the maternal diet may be affecting the maternal metabolome in GDM.

The aim of this study was to comprehensively investigate the maternal metabolome and its association with GDM in a Chinese population, and explore how the maternal diet was related to GDM.

To achieve this aim, the following study comprised three objectives:

- 1) Characterise the maternal urine, plasma, and hair metabolome using gas chromatography-mass spectrometry (GC-MS) and liquid chromatography-mass spectrometry (LC-MS) to explore how the maternal metabolome differs between women who develop GDM and those with a healthy pregnancy, in a Chinese population;
- 2) Derive maternal dietary patterns using factor analysis, and explore their association with GDM;
- 3) Explore associations between the maternal dietary patterns derived and the maternal metabolome.

Methods

Study participants

In this case-control study, 50 GDM and 50 normal pregnancies were recruited from The First Affiliated Hospital of Chongqing Medical University, China, between 24 and 28 weeks' gestation. A 75g oral glucose tolerance test was conducted at the time of recruitment to determine case or control status. Criteria for GDM diagnosis followed the International Association of Diabetic Pregnancy Study Group (IADPSG) guidelines (fasting plasma glucose ≥ 5.1 mmol/L, one-hour post 75g OGTT plasma glucose level ≥ 10 mmol/L, or two-hour post 75g OGTT plasma glucose level ≥ 8.5 mmol/L) (International Association of et al., 2010). Three women were excluded due to the development of gestational hypertension later in pregnancy, three women were excluded due to missing OGTT information, and one participant was excluded due to incomplete dietary information. A total of 49 women remained in the study as cases diagnosed with GDM and 44 remained in the study as healthy controls. Ethical approval was granted by the Ethics committee of the Chongqing Medical University (2015009), all research was performed in accordance with relevant guidelines/regulations, and written informed consent was obtained from all participants. The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Questionnaire collection

Participants completed demographic and dietary questionnaires and provided urine, serum, and hair samples at the time of enrolment. A 96-item food frequency questionnaire (FFQ) was used to assess dietary intake of the participants over the previous three months. Maternal socio-demographic characteristics, family medical history, reproductive, and medical histories were obtained from the provided questionnaires as well as medical records. Information about pregnancy outcomes was

acquired from the maternal and infant medical records in the hospital, after delivery.

Sample collection and randomization

Urine, serum, and hair samples were collected approximately at the same time in the morning, during their morning appointment. Mid-stream urine and fasted serum samples were collected into collection tubes, processed according to a standard operating procedure, and frozen at -80°C . Maternal hair strands were cut 0.5 cm away from the scalp around the occipital area of the rear of the head, and stored in aluminium foil at 4°C . Sample IDs were randomly assigned into five batches, balanced according to case-control status, so that there was an equal number of GDM and normal pregnancies per batch.

Sample preparation for GC-MS analysis

Two 110 μl aliquots of thawed urine were transferred into 1.5 ml microcentrifuge tubes (one from each biofluid for GC-MS analysis, the other for LC-MS), then an internal standard (IS) was added to each sample (20 μl of 2,3,3,3-d₄-alanine [10mM]). The sample-IS urine mixes were stored at -80°C prior to MCF derivatisation.

Two 150 μl aliquots of thawed serum were transferred into 1.5 ml microcentrifuge tubes, then an internal standard (IS) was added to each sample (20 μl of 2,3,3,3-d₄-alanine [10mM]). To precipitate protein from the serum samples, 510 μl of cold methanol was added to the 150 μl serum aliquots, followed by freezing at -20°C for 30 min. Then the supernatant was isolated by centrifugation and dried in a speedvac (LABCONCO, CentriVap Concentrator (catalog #7810041) and Cold Trap (catalog #7385037)) at 3000 rpm for 8 hr at room temperature. The extracted samples were stored at -80°C until derivatisation.

The hair was washed, IS added (20 μl of 2,3,3,3-d₄-alanine [10mM]), and metabolites extracted using the methods published by Sulek *et al* (2014) (Sulek et al., 2014). Dried hair extracts were stored at -80°C until derivatisation.

In term of the quality control (QC) preparation, the urine and serum pooled QC samples were prepared for each biofluid by sub-aliquoting 20 μl directly from all frozen collection tubes. For hair samples, pooled hair QC samples were prepared by aliquoting 20 μl after metabolite extraction. Then the pooled QC samples were further sub-aliquot and subsequently prepared the same ways according to the sample type.

Methyl Chloroformate (MCF) derivatisation and GC-MS analysis

All prepared samples were chemically modified using MCF derivatisation prior to GC-MS analysis, as per the methods reported by Smart *et al* (2010), published in Nature protocols (Smart et al., 2010). The MCF derivatives were analyzed on an Agilent GC7890B system linked to a MSD5977A mass selective detector (EI) set at 70 eV. Details of the GC-MS parameters and analytical method are reported in Table S2.

Sample preparation for LC-MS analysis

Each 110 µl urine aliquot was diluted with 110 µl of milli-Q (MQ) water containing 0.2% formic acid. After a 30s vortex and centrifugation for 10 min at 14 000 g, the supernatant was transferred into a glass LC vial (Waters, screw top vial, 12×32mm) and stored at –80 °C prior to LC-MS analysis.

Serum sample preparation was performed using the Ostro Pass-through 96-well sample preparation plate (Waters, Manchester, UK) to remove high molecular weight compounds, including proteins and some complex lipids. The eluates were collected and dried in a speedvac (LABCONCO, CentriVap Concentrator (catalog #7810041) and Cold Trap (catalog #7385037) at 3000 rpm for 8 hr at room temperature. The dried extracts were stored at –80 °C and reconstituted with 50 µl MQ water containing 5% acetonitrile, directly prior to LC-MS analysis.

The hair metabolite extraction protocol for LC-MS analysis was adapted from Broecker *et al* (2012) (Broecker et al., 2012). After cleaning of the hair as per the protocol in Sulek *et al* (2014), 20 mg was weighed out from each participant and placed in Bead Beating Tubes. The hair was ground into powder using a Bioprep-24 homogenizer then incubated in a mixture containing 2:2:1 methanol/acetonitrile/2mM ammonium formate for 18 hr, with gentle shaking. The sample was centrifuged twice at 10,000 rpm for 10 min, and the supernatant was collected and dried in a speedvac (LABCONCO, CentriVap Concentrator (catalog #7810041) and Cold Trap (catalog #7385037)) at 3000 rpm for 8 hr at room temperature. The dried extract was stored at –80 °C and reconstituted with 200 µl MQ water containing 10% acetonitrile, directly prior to LC-MS analysis.

LC-MS analysis

The samples were analysed on a Waters Acquity UPLC I-Class system, coupled to a Waters Xevo G2 QToF mass spectrometer equipped with an electrospray ionization source. Details of the analytical methods and parameters used for each bio-specimen are reported in Table S3.

GC-MS and LC-MS data extraction and normalization

Raw GC-MS chromatogram deconvolution and metabolite identification were performed using AMDIS software, combined with an in-house MCF mass spectra library of 461 standard compounds, according to the process described in Sulek *et al* (2014). The criteria for metabolite identification included a match to the MS spectrum of the library compound >75% and within a one minute retention time window of its respective chromatographic retention time. The GC-MS data was normalised by the internal standard (2,3,3,3-d4-alanine) and batch correction was performed using median centering of QC samples and EigenMS (Karpievitch *et al.*, 2014, Karpievitch *et al.*, 2012, Karpievitch *et al.*, 2009). A dilution correction was achieved using all identified compounds in urine and serum, while biomass in mg was used for hair.

The LC-MS data was processed and normalised using the same methods described by Law *et al* (2016) (Law *et al.*, 2016). The LC-MS data was imported to Progenesis QI, ver. 2.2 (Nonlinear Dynamics) and the LC-MS chromatograms were automatically deconvoluted based on a reference QC sample. The reference QC chromatogram was chosen by assessing all QC samples in the experiment for suitability.

The chromatographic peaks were considered for analysis if they had a minimum of 1.5s, 1.5s, and, 1.8s width for urine, serum, and hair samples respectively. Global data normalisation was executed by normalizing to all compounds, which uses quality control metabolites to substitute internal standards (De Livera *et al.*, 2015), and is based on an assumption that the majority of compounds in a set of samples will not be changing in an experiment, and that their quantitative abundance ratio should be equal to one. Our previous study demonstrated this normalisation method showed better performance than creatinine and total ion count normalization (Law *et al.*, 2017). This is the default setting of Progenesis QI and data equality was corrected by the modified version of EigenMS (Karpievitch *et al.*, 2014, Karpievitch *et al.*, 2012, Karpievitch *et al.*, 2009).

Metabolite identification was performed using an online database search, using the MetaScope algorithm from the Human Metabolite Database (version 3.6), Toxic Exposome Database, Small Molecule Pathway Database, and Lipid Maps Database (release date 28 June 2015). The confidence of compound identification was determined by the confidence score produced by Progenesis QI. The score was calculated by taking into account theoretical fragmentation, mass error of precursor ion (5 ppm), retention time, and isotope distribution. Only metabolite candidates with a confidence score ≥ 54 were considered for statistical analysis.

Statistical analysis

Metabolite profiles

Binary logistic regression was conducted to identify significant differences in metabolite abundance between case and control groups, with adjustment for confounding demographic variables that were significantly different between case and control groups (maternal age and pre-pregnancy BMI). To account for simultaneous testing of many metabolites, false discovery rates (FDR) were calculated for each metabolite using the qvalue R package (Storey and Tibshirani, 2003). Metabolites with an FDR < 0.05 were considered significant (JD. et al., 2004). Heatmaps were constructed using the ggplot2 R package (Wickham, 2009) to represent the metabolite ratios between GDM and control groups. Chromatographic data was log-transformed and Pareto scaled prior to statistical modelling (Xia and Wishart, 2016). Receiver Operating Characteristic (ROC) curves were produced using the pROC R package (Robin et al., 2011).

Dietary pattern analysis

Food items recorded in the FFQ were combined into 76 groups according to similar nutrient composition. Dietary patterns were derived using principal component factor analysis with varimax rotation. Dietary patterns that explained more than 3% of the variation in dietary intake were selected for further analysis. Factor loadings represent the correlations between each food group and the dietary pattern. Factor scores for each dietary pattern were calculated by summing the consumption frequencies of each food group, weighted by their respective factor loadings. Food groups with loadings $> \pm 0.30$ were used to describe the dietary patterns. Associations between maternal dietary patterns and GDM were determined using logistic regression adjusted for confounding demographic variables (maternal age and pre-pregnancy BMI). All dietary statistical analyses were performed in SPSS version 22 (IBM). $P < 0.05$ was considered statistically significant. To estimate energy intake from the FFQ data, a frequently consumed food item from each food group was selected by a researcher that was involved in data collection from participants and was familiar with both the food environment in Chongqing and with the participant's responses when they were completing the FFQ. The food items chosen to represent the food groups were, where possible, selected according to the most frequently recorded item when there was an option to put examples of foods eaten in that food group, or to circle the example most commonly consumed. The calories (per gram) in that representative food item were then obtained from the Chinese food composition tables (Yang et al., 2009). Estimated daily calorie intake per participant = the sum of daily calorie intake from each food

group (= calories (per gram) of the food group's representative item x amount (in grams) the individual consumed of that food group per day). To calculate food volume, the sum of the volume of food consumed from each food group (per day) was calculated.

Results

Clinical characteristics of study participants

Clinical characteristics for all participants are shown in Table 1. The fasting, one-hour, and two-hour blood glucose levels following an OGTT were statistically different between case and control groups. The GDM group had a significantly greater age and BMI than the control group and tended to deliver earlier than women in the control group.

Metabolomic profiling

A total of 232 metabolites were identified in the urine of the participants in our study using GC-MS, and 181 metabolites using LC-MS. Of the metabolites identified in urine, 35 were found to be significantly different ($P < 0.05$) between GDM cases and non-GDM controls (20 from GC-MS analysis, 15 from LC-MS; Supplementary excel sheet 1). After adjustment for multiple comparisons, none of these metabolites remained significantly associated with GDM.

The serum metabolite profile consisted of 96 metabolites identified using GC-MS analysis, and 224 metabolites identified using LC-MS analysis. A total of 13 metabolites were found to be significantly different in the serum of cases when compared to controls (6 metabolites from GC-MS analysis, 7 from LC-MS; Supplementary excel sheet 1). After adjustment for multiple comparisons, one metabolite remained significant ($P = 0.0010$; FDR = 0.0348); 2-aminoabutyric acid (Table S4, Figure S1).

The hair metabolomic analysis resulted in the identification of 160 metabolites identified using GC-MS analysis and 171 metabolites identified using LC-MS analysis. A total of 25 metabolites were found to be significantly different between cases and controls (13 from GC-MS analysis, 12 from LC-MS analysis; Supplementary excel sheet 1). After adjustment for multiple comparisons, none of these metabolites remained significantly associated with GDM.

The heatmaps displayed in Figures 1-3 demonstrate the difference in ratios between cases and controls of the significant metabolites.

Dietary patterns

Nine maternal dietary patterns were derived from the FFQ data (Table S5), which explained 34.5% of the total variation in dietary intake recorded. The large number of dietary patterns in this study

demonstrates the heterogeneity in the dietary consumption of the participants – participants did not all appear to be consuming the same food combinations/diets. The nature of dietary pattern investigations using factor analysis means that the data informs the researchers what the participants are consuming/how foods are being consumed together, rather than using apriori assumptions.

After adjustment for maternal age and pre-pregnancy BMI, the ‘snack-based’ dietary pattern was significantly associated with an increased risk of GDM ($P = 0.02$), with an odds ratio (OR) of 2.1 (95% $CI = 1.1, 3.9$; Table S6). The snack-based dietary pattern explained 4.8% of the total variance in dietary intake from all participants. The snack-based dietary pattern was characterised by frequent intake of fruits (stone fruit and tropical fruit), yoghurt, bread/buns filled with sweet fillings, bread/buns filled with savoury fillings, low-fat milk, noodles with gravy, dark green leafy vegetables, and peas and green beans (Table S5). Participants with scores in the top tertile of the snack-based dietary pattern consumed significantly more calories (cal) per day than participants in the lowest tertile (3035 cal (± 795) compared to 2291 cal (± 877); $P = 0.0002$). After adjustment for energy intake (calories consumed), the association between the snack-based dietary pattern and GDM remained significant with an OR of 2.2 (95% $CI = 1.1, 4.5$; $P = 0.04$). No significant differences were observed between the upper and lower tertiles in regards to variables such as age, BMI, gestational weight gain, or pregnancy outcomes such as infant birth weight or infant length.

Link between the diet and metabolome

A linear regression, adjusted for maternal age and BMI, was conducted to determine which metabolites were associated with the snacks-based dietary pattern. Overall, 59 metabolites were found to be significantly associated with the snacks-based dietary pattern ($P < 0.05$; Supplementary excel sheet 2).

However, only five metabolites met the significance cut-off ($FDR < 0.05$; Table S7) after adjustment for multiple comparison testing: Glucosyl-3-keto-1,25R-hexacosanediol, E-10-Hydroxynortriptyline, PGF2alpha-dihydroxypropanylamine, Acorusnol, and Tetradecanol.

None of the metabolites found to be significantly associated with the snack-based dietary pattern were also significantly associated with GDM. However, 2-aminobutyric acid had a significant linear relationship with the snack-based dietary pattern prior to adjustment, although its high FDR reflected a likely false positive discovery ($P = 0.0279$; $FDR = 0.9418$).

Discussion

This study was the first to extensively characterise the maternal metabolome of Chinese participants;

the extensive hair, serum, and urine analysis identified a total of 1064 metabolites.

Of the metabolites identified, 73 (6.8%) were found to be significantly different between GDM cases and controls, after adjustment for maternal age and BMI. After further adjustment to account for the effect of multiple comparisons, one serum metabolite remained significantly different between GDM cases and healthy controls: 2-aminobutyric acid, which was higher in the serum of women that developed GDM. Interestingly, it was also higher in the urine of women who developed GDM, but its significance was attenuated after adjustment for multiple comparisons (Supplementary excel sheet 1).

2-aminobutyric acid supports glutathione production (Lu, 2009), and glutathione is one of several important peptides with antioxidant capabilities (Wu et al., 2004). Dysregulation of glutathione synthesis is being recognised as contributing to the pathogenesis of many pathological conditions, including diabetes mellitus (Lu, 2009, Wu et al., 2004). 2-aminobutyric acid can also be converted to 2-hydroxybutyric acid (Landaas, 1975), a metabolite which has been found previously as an early biomarker of insulin resistance in a non-diabetic population (Gall et al., 2010).

A limitation of untargeted metabolomics investigations – which can result in a large number of metabolite detections relative to the number of samples – is the potential for false positive discoveries, resulting from multiple comparison testing (Broadhurst and Kell, 2006). To reduce the likelihood of false positive discoveries in this study we performed multiple comparison adjustments using the Storey and Tibshirani (2003) method, and reported findings with a false discovery rate < 0.05 as statistically significant. However, stringent adjustments can introduce false negatives, and risk ignoring potentially valid markers. For the benefit of future studies, we have listed all metabolites with $P < 0.05$ in Supplementary excel sheet 1.

The snack-based dietary pattern identified in this study was found to be significantly associated with an increased likelihood of GDM. Participants who scored in the top tertile on this pattern consumed significantly more calories and a higher volume of food. However, the association with GDM remained significant after adjustment for energy intake (calories consumed). After adjustment for total volume of food consumed, the dietary pattern no longer remained significant, suggesting that the volume of food consumed played an important role in the association between the snack-based dietary pattern and GDM. High food volume can be consumed in two different ways:

- 1) Few meals throughout the day, each containing a high volume of food from a range of sources,
- 2) Frequent snacking.

Although the FFQs do not provide information on daily consumption patterns, based on the types of foods in the snack-based dietary pattern, we hypothesise that the second option was most likely – frequent snacking. Frequent snacking may be a risk factor for GDM if the snacks are high in glycaemic index or glycaemic load. Frequent high Glycaemic Index (GI) snacking may result in frequent insulin spiking which could lead to insulin resistance or eventual exhaustion of the pancreatic beta cells (Willett et al., 2002). Information on meal timing and meal loads should be collected in future studies to understand how food is being consumed.

The only study reported to investigate the association between dietary patterns and GDM in China found that a vegetable-based dietary pattern was associated with a reduced likelihood of GDM, whereas a sweets and seafood-based pattern was associated with an increased likelihood of GDM (He et al., 2015). Neither of the patterns identified in the study conducted by He *et al* (2015) were similar to the snack-based dietary pattern found in the current study. The disparity may be due to the limited number of samples in our study, but may also reflect the difference in dietary consumption across different regions of China, and is something that should be taken into consideration when translating dietary findings into recommendations.

Five metabolites were found to be significantly associated with the snack-based dietary pattern, which was positively associated with GDM. All of the five metabolites were identified in the maternal serum, indicating that serum is a better specimen than urine and hair in this cohort for detecting the effect of habitual dietary intake. Although none of the metabolites associated with the snack-based dietary pattern were also significantly linked to GDM, they are related to potential maternal exposures which offer additional insight into the link between maternal diet and GDM.

Glucosyl-3-keto-1,25R-hexacosanediol is a glycolipid that is often contained in the heterocysts of certain cyanobacteria, specifically, *Anabaena* species (Bauersachs et al., 2009). *Anabaena* are nitrogen-fixing cyanobacteria which are most likely to be consumed by humans through contaminated water sources or other food sources (2017). Some strains of *Anabaena* are also used as natural fertilisers in rice paddy fields in China (Hong, 1992). Therefore, it is plausible that this metabolite is related to dietary exposure in our study. E-10-hydroxynortriptyline is a pharmaceutical metabolite that is present with the use of anti-depressant medication (Nordin and Bertilsson, 1995). Data on maternal mental health and medication was not collected in this study. However, anti-depressant use has been previously found to increase appetite, and is associated with weight gain and higher risk of metabolic

complications (Choong et al., 2012). Women who scored highly on the snack-based dietary pattern consumed greater food volume and total number of calories. Future studies should explore the role of depression and anti-depressant medication on the maternal diet and how this might be related to GDM development.

Conclusions

This study provides the most comprehensive characterisation of the maternal metabolome of pregnant women in China. We found that 2-aminobutyric acid was significantly different between GDM cases and healthy controls. In addition, we identified a dietary pattern that was associated with GDM in this Chinese cohort – a snack-based dietary pattern. Woman who scored highly on this pattern consumed significantly more calories and a higher volume of food. Commonly conducted dietary assessments such as FFQs do not capture information on timing and frequency of food consumed within a typical 24-hour period. Future studies should consider these aspects of dietary behaviour when seeking to understand the relationship between maternal diet and GDM development.

Additional Information

Acknowledgments: We acknowledge the LC-MS scientific support provided by Dr. Kai Law. We also thank Xun Mao and Diqi Zhao for the help of data and sample collection.

Author contributions: YX, PB and HZ conceived the idea and designed the study. TZ and XC collected the questionnaires and samples. TH, CC, and JdS analysed the data. XC, JdS, and TH contributed to the writing of the manuscript.

Conflicts of Interest: The authors declare that they have no conflict of interest.

Compliance with Ethical Standards: Ethical approval was granted by the Ethics committee of the first affiliated hospital of Chongqing Medical University and written informed consent was obtained from all participants.

Funding: This work was supported by the National Natural Science Foundation of China [No.81571453, 81771607, 81701477, 81650110522] and The 111 Project [Yuwaizhuan (2016)32]. The funders were not involved in the study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

References

- [2017. *Health and Ecological Effects* [Online]. Available: www.epa.gov/nutrient-policy-data/health-and-ecological-effects [Accessed 23 July 2017].
- BAUERSACHS, T., COMPAORE, J., HOPMANS, E. C., STAL, L. J., SCHOUTEN, S. & SINNINGHE DAMSTE, J. S. 2009. Distribution of heterocyst glycolipids in cyanobacteria. *Phytochemistry*, 70, 2034-9.
- BROADHURST, D. I. & KELL, D. B. 2006. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics*, 2, 171-196.
- BROECKER, S., HERRE, S. & PRAGST, F. 2012. General unknown screening in hair by liquid chromatography-hybrid quadrupole time-of-flight mass spectrometry (LC-QTOF-MS). *Forensic Sci Int*, 218, 68-81.
- CHEN, X., ZHAO, D., MAO, X., XIA, Y., BAKER, P. N. & ZHANG, H. 2016. Maternal Dietary Patterns and Pregnancy Outcome. *Nutrients*, 8.
- CHONG, Y. S., CAI, S., LIN, H., SOH, S. E., LEE, Y. S., LEOW, M. K., CHAN, Y. H., CHEN, L., HOLBROOK, J. D., TAN, K. H., RAJADURAI, V. S., YEO, G. S., KRAMER, M. S., SAW, S. M., GLUCKMAN, P. D., GODFREY, K. M., KWEK, K. & GROUP, G. S. 2014. Ethnic differences translate to inadequacy of high-risk screening for gestational diabetes mellitus in an Asian population: a cohort study. *BMC Pregnancy Childbirth*, 14, 345.
- CHOONG, E., BONDOLFI, G., ETTER, M., JERMANN, F., AUBRY, J. M., BARTOLOMEI, J., GHOLAM-REZAEI, M. & EAP, C. B. 2012. Psychotropic drug-induced weight gain and other metabolic complications in a Swiss psychiatric population. *J Psychiatr Res*, 46, 540-8.
- DE LIVERA, A. M., SYSI-AHO, M., JACOB, L., GAGNON-BARTSCH, J. A., CASTILLO, S., SIMPSON, J. A. & SPEED, T. P. 2015. Statistical methods for handling unwanted variation in metabolomics data. *Anal Chem*, 87, 3606-15.
- FEDERATION, I. D. 2015. *IDF Diabetes Atlas (7th ed)* [Online]. Available: www.diabetesatlas.org [Accessed 23 July 2017].
- FERRARA, A. 2007. Increasing prevalence of gestational diabetes mellitus: a public health perspective. *Diabetes Care*, 30 Suppl 2, S141-6.
- GALL, W. E., BEEBE, K., LAWTON, K. A., ADAM, K. P., MITCHELL, M. W., NAKHLE, P. J., RYALS, J. A., MILBURN, M. V., NANNIPIERI, M., CAMASTRA, S., NATALI, A., FERRANNINI, E. & GROUP, R. S. 2010. alpha-hydroxybutyrate is an early biomarker of insulin resistance and glucose intolerance in a nondiabetic population. *PLoS One*, 5, e10883.
- HE, J. R., YUAN, M. Y., CHEN, N. N., LU, J. H., HU, C. Y., MAI, W. B., ZHANG, R. F., PAN, Y. H., QIU, L., WU, Y. F., XIAO, W. Q., LIU, Y., XIA, H. M. & QIU, X. 2015. Maternal dietary patterns and gestational diabetes mellitus: a large prospective cohort study in China. *Br J Nutr*, 113, 1292-300.
- HONG, G.-F. 1992. *The Nitrogen Fixation and its Research in China* Berlin, Springer-Verlag..
- INTERNATIONAL ASSOCIATION OF, D., PREGNANCY STUDY GROUPS CONSENSUS, P., METZGER, B. E., GABBE, S. G., PERSSON, B., BUCHANAN, T. A., CATALANO, P. A., DAMM, P., DYER, A. R., LEIVA, A., HOD, M., KITZMILER, J. L., LOWE, L. P., MCINTYRE, H. D., OATS, J. J., OMORI, Y. & SCHMIDT, M. I. 2010. International association of diabetes and pregnancy study groups recommendations on the diagnosis and classification of hyperglycemia in pregnancy. *Diabetes Care*, 33, 676-82.
- JD., S., JE., T. & D., S. 2004. Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: A unified approach. *Journal of the Royal Statistical Society, Series B*, 187-205.

KARPIEVITCH, Y. V., DABNEY, A. R. & SMITH, R. D. 2012. Normalization and missing value imputation for label-free LC-MS analysis. *BMC Bioinformatics*, 13 Suppl 16, S5.

KARPIEVITCH, Y. V., NIKOLIC, S. B., WILSON, R., SHARMAN, J. E. & EDWARDS, L. M. 2014. Metabolomics data normalization with EigenMS. *PLoS One*, 9, e116221.

KARPIEVITCH, Y. V., TAVERNER, T., ADKINS, J. N., CALLISTER, S. J., ANDERSON, G. A., SMITH, R. D. & DABNEY, A. R. 2009. Normalization of peak intensities in bottom-up MS-based proteomics using singular value decomposition. *Bioinformatics*, 25, 2573-80.

KIM, C., NEWTON, K. M. & KNOPP, R. H. 2002. Gestational diabetes and the incidence of type 2 diabetes: a systematic review. *Diabetes Care*, 25, 1862-8.

LANDAAS, S. 1975. The formation of 2-hydroxybutyric acid in experimental animals. *Clin Chim Acta*, 58, 23-32.

LAW, K. P., HAN, T. L., MAO, X. & ZHANG, H. 2017. Tryptophan and purine metabolites are consistently upregulated in the urinary metabolome of patients diagnosed with gestational diabetes mellitus throughout pregnancy: A longitudinal metabolomics study of Chinese pregnant women part 2. *Clin Chim Acta*, 468, 126-139.

LAW, K. P., MAO, X., HAN, T. L. & ZHANG, H. 2016. Unsaturated plasma phospholipids are consistently lower in the patients diagnosed with gestational diabetes mellitus throughout pregnancy: A longitudinal metabolomics study of Chinese pregnant women part 1. *Clin Chim Acta*, 465, 53-71.

LU, S. C. 2009. Regulation of glutathione synthesis. *Mol Aspects Med*, 30, 42-59.

NICHOLSON, J. K., LINDON, J. C. & HOLMES, E. 1999. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica*, 29, 1181-9.

NORDIN, C. & BERTILSSON, L. 1995. Active hydroxymetabolites of antidepressants. Emphasis on E-10-hydroxy-nortriptyline. *Clin Pharmacokinet*, 28, 26-40.

O'GORMAN, A., GIBBONS, H. & BRENNAN, L. 2013. Metabolomics in the identification of biomarkers of dietary intake. *Comput Struct Biotechnol J*, 4, e201301004.

REECE, E. A., LEGUIZAMON, G. & WIZNITZER, A. 2009. Gestational diabetes: the need for a common ground. *Lancet*, 373, 1789-97.

ROBIN, X., TURCK, N., HAINARD, A., TIBERTI, N., LISACEK, F., SANCHEZ, J. C. & MULLER, M. 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12, 77.

SMART, K. F., AGGIO, R. B., VAN HOUTTE, J. R. & VILLAS-BOAS, S. G. 2010. Analytical platform for metabolome analysis of microbial cells using methyl chloroformate derivatization followed by gas chromatography-mass spectrometry. *Nat Protoc*, 5, 1709-29.

STOREY, J. D. & TIBSHIRANI, R. 2003. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A*, 100, 9440-5.

SULEK, K., HAN, T. L., VILLAS-BOAS, S. G., WISHART, D. S., SOH, S. E., KWEK, K., GLUCKMAN, P. D., CHONG, Y. S., KENNY, L. C. & BAKER, P. N. 2014. Hair metabolomics: identification of fetal compromise provides proof of concept for biomarker discovery. *Theranostics*, 4, 953-9.

WICKHAM, H. 2009. *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag New York.

WILLETT, W., MANSON, J. & LIU, S. 2002. Glycemic index, glycemic load, and risk of type 2 diabetes. *Am J Clin Nutr*, 76, 274S-80S.

WU, G., FANG, Y. Z., YANG, S., LUPTON, J. R. & TURNER, N. D. 2004. Glutathione metabolism and its

489   implications for health. *J Nutr*, 134, 489-92.

490   XIA, J. & WISHART, D. S. 2016. Using MetaboAnalyst 3.0 for Comprehensive Metabolomics Data

491   Analysis. *Curr Protoc Bioinformatics*, 55, 14.10.1-14.10.91.

492   YANG, Y. X., WANG, G. Y. & PAN, X. C. 2009. China Food Composition (Book 1) , Beijing, Peking

493   University Medical Press.

494   ZHU, W. W., YANG, H. X., WANG, C., SU, R. N., FENG, H. & KAPUR, A. 2017. High Prevalence of

495   Gestational Diabetes Mellitus in Beijing: Effect of Maternal Birth Weight and Other Risk

496   Factors. *Chin Med J (Engl)*, 130, 1019-1025.

497

Figure 1. Heat map of the significant metabolites detected in urine. The ratio of identified metabolite levels between GDM case and control groups. Red indicates metabolite abundance from women identified as GDM cases higher than controls (green), plotted using a log2 scale. Metabolites with FDR < 0.05 are labelled by ***, < 0.1 are labelled by **, and < 0.2 are labelled by *.

Figure 2. Heat map of the significant metabolites detected in serum. The ratio of identified metabolite levels between GDM case and control groups. Red indicates metabolite abundance from women identified as GDM cases higher than controls (green), plotted using a log2 scale. Metabolites with FDR < 0.05 are labeled by ***, < 0.1 are labeled by **, and < 0.2 are labeled by *.

Figure 3. Heat map of the significant metabolites detected in hair. The ratio of identified metabolite levels between GDM case and control groups. Red indicates metabolite abundance from women identified as GDM cases higher than controls (green), plotted using a log2 scale. Metabolites with FDR < 0.05 are labeled by ***, < 0.1 are labeled by **, and < 0.2 are labeled by *.

Table 1. Participant Characteristics

	GDM n=49	Normal n=44	P value
Age, years	32 (29, 34)	29 (27, 33)	0.0075 ^a
Pre-pregnant BMI, kg/m ²	21.6 (20.2, 24.6)	20.7 (19.0, 21.8)	0.0132 ^a
Parity			
Nulliparous	39 (79.6)	36 (81.8)	
Primiparous	9 (18.4)	8 (18.2)	
Multiparous	1 (2.0)	0 (0.0)	0.6336 ^c
Monthly household income	10000 (8000, 15000)	10000 (8000, 12000)	0.8620 ^a
Maternal education			
Junior middle school	5 (10.2)	1 (2.3)	
Senior middle school	4 (8.2)	5 (11.4)	
Bachelor degree	33 (67.3)	32 (72.7)	
Master degree and above	7 (14.3)	6 (13.6)	0.4559 ^c
Family history of diabetes			
First-degree relatives	11 (22.4)	6 (13.6)	
Secondary relatives	6 (12.2)	4 (9.1)	
None	32 (65.3)	34 (77.3)	0.4345 ^c
Previous GDM			
Yes	1 (2.0)	0 (0.0)	
No	48 (98.0)	44 (100.0)	1.0000 ^c
Passive smoking during pregnancy			
Yes	12 (24.5)	13 (29.5)	
No	37 (75.5)	31 (70.5)	0.7529 ^c
Systolic blood pressure, mmHg	112.0±11.3	113.2±11.8	0.5995 ^b
Diastolic blood pressure, mmHg	63.0 (59.0, 66.3)	66.0 (59.0, 70.3)	0.0693 ^a
OGTT_fasting, mmol/L	5.0±0.5	4.5±0.3	<0.0001 ^b
OGTT_1h, mmol/L	10.6±1.7	7.6±1.2	<0.0001 ^b
OGTT_2h, mmol/L	9.4 (8.7, 10.6)	6.9 (6.0, 7.6)	<0.0001 ^a
Gestational age, week	39.2±1.2	39.8±1.1	0.0263 ^b
Placental weight, g	577±108	558±67	0.3381 ^b
Infant birth weight, g	3299±450	3380±415	0.3669 ^b
Infant head circumference, cm	34.0 (33.9, 35.0)	34.8 (34.0, 35.0)	0.3165 ^a
Infant length, cm	49.6±2.0	49.9±1.6	0.3909 ^b
Infant sex			
Male	21 (42.9)	23 (52.3)	
Female	28 (57.1)	21 (47.7)	0.4839 ^c

511 Values are means ± SE, median (IQR) or n (%)

512 ^aP value from Mann-Whitney test ^bP value from Student T-test ^cP value from chi-square test